The global distribution and transmission limits of lymphatic filariasis: past and present

Jorge Cano1*, Maria P Rebollo2,3, Nick Golding4, Rachel L Pullan1, Thomas Crellen1, Anna Soler1, Louise A Kelly-Hope3, Steve W Lindsay5, Simon I Hay4, Moses J Bockarie3 and Simon J Brooker1

Abstract

Background: Lymphatic filariasis (LF) is one of the neglected tropical diseases targeted for global elimination by 2020 and to guide elimination efforts countries have, in recent years, conducted extensive mapping surveys. Documenting the past and present distribution of LF and its environmental limits is important for a number of reasons. Here, we present an initiative to develop a global atlas of LF and present a new global map of the limits of LF transmission.

Methods: We undertook a systematic search and assembly of prevalence data worldwide and used a suite of environmental and climatic data and boosted regression trees (BRT) modelling to map the transmission limits of LF.

Results: Data were identified for 66 of the 72 countries currently endemic and for a further 17 countries where LF is no longer endemic. Our map highlights a restricted and highly heterogeneous distribution in sub-Saharan Africa, with transmission more widespread in West Africa compared to east, central and southern Africa where pockets of transmission occur. Contemporary transmission occurs across much of south and South-east Asia and the Pacific. Interestingly, the risk map reflects environmental conditions suitable for LF transmission across Central and South America, including the southern States of America, although active transmission is only known in a few isolated foci. In countries that have eliminated LF, our predictions of environmental suitability are consistent with historical distribution.

Conclusions: The global distribution of LF is highly heterogeneous and geographically targeted and sustained control will be required to achieve elimination. This first global map can help evaluate the progress of interventions and guide surveillance activities.

Keywords: Lymphatic filariasis, Global distribution, Transmission limits, Boosted regression tree modelling

Background

Lymphatic filariasis (LF) is a mosquito-borne disease which in its advanced forms can manifest as severe lymphoedema, hydrocele and elephantiasis [1]. The majority of global cases are caused by *Wuchereria bancrofti*, with *Brugia malayi* and *B. timori* as important local causes of the disease in South-east Asia. These nematode parasites are transmitted by various species of mosquito vectors from the genera *Anopheles*, *Aedes*, *Culex*, *Mansonella* and *Ochlerotatus*. LF is one of nine infectious diseases targeted for global elimination [2]. The selection of LF for elimination was based on (i) the absence of animal reservoirs for *W. bancrofti* (the most common form of LF) and only a small animal reservoir for *B. malayi* (which occurs in restricted foci) [3,4], (ii) the existence of effective and practical interventions to interrupt transmission, and (iii) availability of an accurate diagnostic tool [5]. The main intervention strategy is mass drug administration (MDA) with albendazole in combination with diethylcarbamazine (DEC) (or ivermectin in countries where onchocerciasis is endemic) to entire communities in districts where the prevalence of LF is equal or more than 1% [6], supported by vector control to reduce exposure to mosquitoes and morbidity management to alleviate suffering and prevent disability of those affected by the disease [7,8]. During the last half century, several countries have successfully eliminated LF, including Japan, China, South Korea, the Solomon Islands, Egypt and Togo [9-13].

© 2014 Cano et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
A key component of national elimination programmes is a detailed understanding of the geographical distribution of LF so that all endemic areas can be targeted. Early attempts to map the distribution of LF included detailed literature reviews [14-20] or mapping at national or sub-national levels [21-26]. Although useful for estimating the global burden of LF [27], such national estimates belied the highly focal distribution of LF [28-32] and cannot be used to geographically target control. The mapping of LF has in recent years been greatly facilitated by the use of simple and rapid detection tests for *W. bancrofti* (antigen-based test) and *Brugia* spp (antibody-based test), based on the immuno-chromatographic test (ICT card test), which avoids the need to collect blood at night and the time-consuming preparation and examination of blood slides [33-38]. By the end of 2012, 59 out of the 72 endemic countries had completed national mapping surveys [39]. The results of these surveys highlight the marked within-country geographical heterogeneity [26,40-46]. LF mapping is ongoing in the remaining endemic countries except Eritrea, where it has not started yet [2].

To augment available field surveys, a number of recent studies have sought to predict the distribution of LF on the basis of climatic and environmental factors. This approach is predicated on the fact that *W. bancrofti* is inefficiently transmitted, requiring thousands of infective bites to establish a patent infection [47], and as such LF is only likely to occur where climatic conditions are suitable to support mosquito vector populations over extended time periods [14,48]. An early attempt to develop a risk map for LF in Africa was by Lindsay and Thomas in 2000, based on data from 32 studies using frequentist logistic regression and coarse-resolution environmental data [49]. More recently, Slater and Michael have used maximum entropy ecological niche modelling [50] and Bayesian model-based geostatistics [51] to predict the geographical occurrence and distribution of LF in Africa. Risk maps using environmental factors or spatial interpolation have also been developed at national or sub-national scales in West Africa [52], Egypt [53] and India [54].

Building on this previous work, we describe a new initiative to develop a global atlas of LF infection, which aims to collate all available survey data into a single, freely available resource and describe the historical and contemporary distribution of LF. Understanding these distributions has more than cartographic interest. First, changes in the epidemiology of LF over time can be quantified. Second, factors underlying such changes can be investigated in an effort to assess the degree to which changes are directly related to the scaling-up of interventions or other factors. Third, analysis of the historical risks of infection prior to large-scale intervention can be used to quantify the intervention needs required to reach programme goals [55], identify factors that contribute to the persistence of transmission, help define the intrinsic sensitivity (receptivity) of transmission to future changes in the intensity and frequency of control [56-59], and provide a basis to stratify surveillance activities. The specific aims of the present paper are (i) to detail the methods and approaches used to develop the database, (ii) to map historical and contemporary distributions of LF, and (iii) to delineate the global transmission limits of LF. The work is conducted within the context of the Global Atlas of Helminth Infections (www.thiswormyworld.org) [60] which aims to develop a suite of geographical resources and tools for neglected tropical diseases (NTDs).

**Methods**

**Approaches to the diagnosis and mapping of LF**

Night collection and examination of blood slides is considered the gold standard approach for detecting microfilariae. However, the sensitivity of this approach crucially relies on the volume of blood sampled [61]. Other parasitological methods, such as Knott’s concentration test and membrane filtration [62], increase the sensitivity of diagnosis but are prohibitively expensive for routine use. Alternatively, the presence of *W. bancrofti* antigenaemia can be detected using the ICT card test [33-35] and presence of specific IgG4 antibodies to *Brugia* spp can be detected using the Brugia Rapid™ test [36-38]. Surveys for the mapping of LF have been based on a variety of sampling designs, including the rapid geographical assessment of Bancroftian filariasis (RAGFIL) method [29,63], lot quality assurance sampling [64,65], population-based household surveys [66] and sentinel site surveys [43,67,68], with the choice of survey methodology dependent on available resources and the stage of the control programme [69].

**Identification of survey data**

Our approach to identify suitable data follows that developed originally by Norman Stoll in his seminal work *This Wormy World* in 1947 [70], and adopted by efforts to map malaria [71,72] as well as soil-transmitted helminth and schistosome infections [73-75]. Relevant data on the prevalence of LF were identified through a combination of (i) structured searches of electronic bibliographic databases, (ii) additional searches of the ‘grey’ literature, including unpublished surveys and government and international archives, and (iii) direct contact with researchers and control programme managers. The online databases PubMed, MEDLINE, EMBASE and SCOPUS were used to identify relevant studies for LF, using the following predefined Medical Subject Heading (MeSH) terms: lymphatic filariasis, bancroftian filariasis, *Wuchereria bancrofti*, *Brugia malayi*, *Brugia timori*, malayarian filariasis AND current and former country names. All permutations of MeSH terms were entered, with no restrictions on language or date of publication. The
abstracts of returned articles were reviewed and if they did not explicitly report prevalence data, they were discarded. When the abstract was not available, pre-selection was made according to the title. Authors were contacted when additional information was required or when data needed to be disaggregated. Studies were included if they provided: (i) the number of people surveyed, (ii) the number of LF positive cases, (iii) details about the methodology of diagnosis and (iv) details of the specific site where they were conducted, regardless of the administrative level. Surveys reporting only prevalence data without provision of the denominator were also included as these can be used to delineate the limits of transmission. Baseline data from clinical and diagnostic trials fulfilling the inclusion criteria were abstracted, whereas surveys carried out at hospitals, prisons, mental institutions or military facilities were excluded.

Data were extracted into a customized Microsoft Access (Microsoft 2007) database and linked to an identical SQL Server database. Abstracted data included three types of information: (i) epidemiological data on survey method, type of diagnostic method, dates, age range and gender of targeted people, time of survey (day or night), ongoing control activities and number of MDA rounds undertaken in the area at the time of survey, number of people sampled, number of positive individuals and prevalence, diagnostic method used, blood sampling volumes for detection of microfilaraemia and, where available, morbidity data based on hydrocele and/or lymphoedema; (ii) each record was assigned a unique identifier which linked the record to the source publication which was included in an Endnote library (Thomson Reuters 2010), with a pdf copy of each publication obtained; and (iii) all data were linked to first and second administrative units, based on the United Nations’ Second Administrative Level Boundaries (SALB) database [76] and the GADM version 2 database of Global Administrative Areas [77].

Geo-positioning of survey data
A decision-based algorithm (Figure 1) was applied to determine the longitude and latitude of survey locations, using a variety of gazetteers: Bing Maps [78], GeoNet Names Server [79], Fuzzy gazetteers [80] and the Open Street Map project [81]. Geographical coordinates provided

![Figure 1](http://www.parasitesandvectors.com/content/7/1/466)

**Figure 1** Decision-based algorithm for the geopositioning of community surveys. This algorithm was developed to ensure the maximum level of accuracy when geopositioning survey data. Briefly, when longitude and latitude of a survey site were provided by a publication, they were cross-checked against a range of cartographic resources, including NGA GEOnet Names server (http://earth-info.nga.mil/gns/html/), Bing maps (http://www.bing.com/maps/), Fuzzy Gazetteer/ISODP project (http://isodp.hof-university.de/fuzzyg/query/) and OpenStreetMap (http://www.openstreetmap.org/). The same resources were used to geoposition surveys for which coordinates were not provided.
by source publications were cross-checked against these resources. The reliability of geopositioning was established on a scale of 0 to 4, ‘0’ being no coordinates found, ‘1’ highly reliable, ‘2’ fairly reliable (spelling differences in 1 or 2 characters), ‘3’ less reliable (spelling differences in various characters) and ‘4’ highly unreliable when less than 60% of name similarity in gazetteers or when located in nearby sites. All geographic coordinates were standardized to decimal degrees in order to be displayed in the WGS84 geographic coordinate system. Ideally, surveys were located to a point location but in certain instances surveys were located to a wide-area polygon (10-25 km² area), and then the centroid of the polygon used.

**Environment and demographic covariates**

Geolocated prevalence data were linked to a range of environmental and climatic variables which are known to affect the development and survival of LF parasites and its mosquito-vector species [82-86] (Table 1). These data included mean, minimum and maximum estimates of temperature and precipitation at 1 km² resolution obtained from WorldClim database [87]. Elevation data at 1 km resolution were derived from gridded digital elevation models (DEM) produced by the Shuttle Radar Topography Mission (SRTM). An aridity index, which is a generalized function of precipitation, temperature, and/or potential evapotranspiration, was obtained at 1 km resolution from CGIAR-CSI [88]. Estimates of averaged enhanced vegetation index for period 1981 to 2006 were derived from imagery obtained from the Advanced Very High Resolution Radiometer (AVHRR) instrument onboard the NOAA satellites series, available from the data library of the International Research Institute for Climate and Society at Columbia University (IRI) [89]. Land cover data were obtained from the GlobCover project [90] which comprises 22 land cover classes, which we aggregated into seven major classes potentially relevant to the eco-epidemiology of LF: agricultural lands, forest areas, shrubland, grasslands and woodlands, bare soil, urban areas, snow/ice and water areas.

Estimates of population density were obtained from the Gridded Population of the World (GPW) [91], which was used to classify areas as urban, peri-urban or rural areas, based on the assumption that urban extents (UE) have a population densities ≥1,000 persons/km², peri-urban >250 persons/km² within a 15 km distance from UE edge, and rural <250 persons/km² and/or >15 km from the UE edge [92]. A gridded map of urban accessibility 1 km resolution was obtained from the European Commission Joint Research Centre Global Environment Monitoring Unit (JRC) [93]. This dataset defined urban accessibility as the predicted time taken to travel from that grid cell to a city of ≥50,000 persons in the year 2000 using land- or water-based travel. Finally, a gridded map at 5 km resolution of the main geographic regions (Europe, Asia, Africa and America) was created. Western Pacific countries were grouped upon the Asian region whilst countries located in the Arabian Peninsula were included within the African region.

**Table 1 Description of environmental variables used to model the global distribution of lymphatic filariasis transmission**

<table>
<thead>
<tr>
<th>Environmental variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean minimum temperature</td>
<td>Mean of average monthly minimum temperature across all 12 months (period 1950-2000)</td>
</tr>
<tr>
<td>Annual mean temperature</td>
<td>Mean of average monthly mean temperature across all 12 months (period 1950-2000)</td>
</tr>
<tr>
<td>Precipitation in wettest quarter</td>
<td>Total precipitation in the wettest quarter</td>
</tr>
<tr>
<td>Precipitation in driest quarter</td>
<td>Total precipitation in the driest quarter</td>
</tr>
<tr>
<td>Annual precipitation</td>
<td>Mean of accumulative precipitation across all 12 months (period 1950-2000)</td>
</tr>
<tr>
<td>Aridity Index</td>
<td>Indicator of the degree of dryness of the climate at a given location, and result from dividing the mean annual precipitation by the mean annual potential evapo-transpiration</td>
</tr>
<tr>
<td>Elevation</td>
<td>Metres above the sea level obtained by radar imagery</td>
</tr>
<tr>
<td>EVI</td>
<td>Mean of 8-day composite MODIS images for the period 2001 to 2005. EVI raw values (scale from 0 to 1) have been multiplied by 1,000 for being storage as integers.</td>
</tr>
<tr>
<td>Global Land Cover 2006</td>
<td>A simplified version of seven major land cover classes obtained from the UN Global Land Cover (22 classes)</td>
</tr>
<tr>
<td>WHO regions</td>
<td>WHO regions: AFRO/EMRO (African and Eastern Mediterranean regions), AMRO (Americas region), SEARO (South-east Asia) and WPRO (western Pacific)</td>
</tr>
<tr>
<td>Accessibility</td>
<td>Cost (in time units) to access to large urban areas</td>
</tr>
</tbody>
</table>
To bring the spatial resolution of these covariate layers in line with the spatial accuracy of the survey data, covariates were resampled to a common 5 km resolution raster layout based on the WGS-1984 Web Mercator projection using ArcGIS 10.1. (ESRI Inc., Redlands CA, USA). Bilinear interpolation was applied to resample numeric (continuous) raster data sets, whereas nearest neighbour interpolation was used with ordinal raster layers. Possible collinearity between the covariates was explored using cross-correlations. All correlation coefficients were less than 0.7, indicating that covariates were largely orthogonal.

Predicting the probability of LF transmission using boosted regression trees

Boosted regression trees (BRT) modelling was used for mapping the spatial limits of LF transmission. This approach has been shown to have higher predictive accuracy than other distribution models [94] and has been successfully applied to map dengue [95] and malaria vector mosquitoes [96]. BRT combines two machine learning approaches; regression trees (simple hierarchical models which allow non-linear effects of predictors) and boosting (fitting ensemble models by iterative improvements on the existing ensemble [97,98]). A first step in the BRT approach is the definition of occurrence and absence data. Records of disease occurrence were defined as surveys during which one or more cases of LF were detected, regardless of diagnostic method used. Absence records were defined as surveys conducted prior to large-scale control activities and from which no cases of LF were reported. Because relatively few absence records were available (prevalence surveys are typically carried out in areas where disease presence is expected), these data were supplemented with pseudo-absence records following a similar procedure to that used for mapping dengue and malaria vectors [93,94]. Five thousand pseudo-absence data points were generated at random in areas known not to be endemic for LF based on expert knowledge [14-20,99] or those areas considered as unsuitable habitats for mosquito breeding - areas of bare and hyper-arid land, as classified by the GlobCover [90] and Global Aridity Index datasets [88]. In order to maximise the ability of the model to discriminate between suitable and unsuitable areas, regression weights were used to down-weight pseudo-absence records, so that the summed weights of the absence and pseudo-absence records matched that of the presence records.

In the second stage, eight environmental variables along with altitude, population density, accessibility and macro-geographical regions were used to predict LF occurrence in a single BRT model, in order to explore the relative importance of each factor in explaining the global occurrence of LF (Table 1). Those factors that contributed little (relative contribution <2%) to the single BRT model were disregarded to build the final ensemble BRT model. Thus, six covariates (precipitation in the wettest quarter, annual minimum temperature, population density 1990-2015, elevation, enhanced vegetation index and regions) were subsequently selected and eventually used to build the final risk map.

In order to estimate uncertainty in the model and the resulting risk maps, we finally fitted an ensemble of 120 BRT submodels, each fitted to a random bootstrap of the full dataset. The predicted distributions of LF from each of these submodels were then averaged to generate the final risk map. Predictive performance of each submodel was evaluated using the following statistics: proportion correctly classified (PCC), sensitivity (proportion of absences correctly classified), specificity (proportion of presences correctly classified), Kappa (κ) and area under the receiver operator characteristic (AUC). The mean and confidence intervals for each statistic were used to evaluate the predictive performance of the ensemble BRT model. Marginal effect curves were plotted to visualise dependencies between the probability of LF occurrence and each of the covariates. These show the marginal effect of each covariate on the response after averaging the effects of all other covariates. Finally, the relative contribution of each covariate (the percentage of tree branches in each submodel that used the covariate) to the final BRT model was also quantified.

Defining limits of transmission

The resulting predictive map quantifies the environmental suitability for LF transmission. In order to convert this continuous metric into a binary map outlining the limits of transmission, a threshold value of suitability was determined, above which transmission was assumed to be possible. Based on the receiver operating characteristic curve, a threshold value of environmental suitability was chosen which maximised the trade-off between sensitivity, specificity and PCC. Whilst the resulting map delineates environmental suitability for LF transmission, it may include areas where transmission does not actually occur, due either to the disease never having been imported into the area or the consequence of control leading to local elimination. To reflect this, we masked the environmental distribution map to remove areas which are known to be currently non-endemic according to WHO [39,100,101] and other sources [20,40,41,44,99]. Non-endemicity was considered when no cases had been reported for the last 10 years and transmission assessment surveys confirmed the interruption of LF transmission.

Results

LF database

The search strategy identified 9,033 surveys, conducted between 1902 and 2013 in 85 countries, which were
eligible for inclusion. Summary characteristics of included surveys are reported in Table 2. An extended version of this table is additionally provided (see Additional file 1). Data were identified for 35 of 37 current endemic countries of sub-Saharan Africa and Eastern Mediterranean WHO-region (AFRO/EMRO); 9/9 from South East Asia region (SEARO); 18/22 from western Pacific region (WPRO); 4/4 from the Americas region (AMRO). Data were also available for a further 17 countries where LF is no longer endemic or where transmission has recently been declared as interrupted. Current endemic countries for which no data were available were Angola and Gabon, in the AFRO region, and Cambodia, Lao People’s Democratic Republic, Niue, and North Korea, in the SEARO and WPRO regions.

Of eligible surveys, 7,852 (87.1%) represented disaggregated, community-based surveys of which 7,420 (94.3%) could be geo-positioned: 6,442 to point locations and 978 to small areas, such as households with scattered distributions, small islands and small administrative areas. Data extracted from published sources accounted for 53.2% of included survey locations and were the main source of information in the SEARO and WPRO regions. Grey literature, which included country reports, GAELF reports and other unpublished articles, accounted for 46.8% of survey locations, and were the main source of data in the AFRO/EMRO region. The majority (82.9%) of data points were obtained through mapping or prevalence surveys, whereas 1,141 (15.4%) were sentinel site surveys and 16 transmission assessment surveys. Among mapping/prevalence surveys, 609 (9.9%) surveys were obtained by lot quality assurance sampling (LQAS), mostly conducted after 2000. Finally, 110 surveys derive from a countrywide clinical survey in Thailand, 1949-1950, where lymphoedema of lower

Table 2 Characteristics of surveys included in the lymphatic filariasis database

<table>
<thead>
<tr>
<th>WHO - regions</th>
<th>AFRO/EMRO</th>
<th>SEARO</th>
<th>WPRO</th>
<th>AMRO</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current endemic countries</td>
<td>37 (N)</td>
<td>9 (N)</td>
<td>22 (N)</td>
<td>4 (N)</td>
<td>72 (N)</td>
</tr>
<tr>
<td>Countries with data</td>
<td>35 (N)</td>
<td>9 (N)</td>
<td>25 (N)</td>
<td>16 (N)</td>
<td>85 (N)</td>
</tr>
<tr>
<td>Number of surveys identified</td>
<td>5,468 (N)</td>
<td>1,764 (N)</td>
<td>1,322 (N)</td>
<td>479 (N)</td>
<td>9,033 (N)</td>
</tr>
<tr>
<td>Non-community data</td>
<td>673 (12.3%)</td>
<td>246 (13.9%)</td>
<td>201 (15.2%)</td>
<td>43 (9.0%)</td>
<td>1,163 (12.9%)</td>
</tr>
<tr>
<td>Geopositioned</td>
<td>4,624 (84.6%)</td>
<td>1,320 (74.8%)</td>
<td>1,058 (80.0%)</td>
<td>418 (87.3%)</td>
<td>7,420 (82.1%)</td>
</tr>
<tr>
<td>Not geopositioned</td>
<td>171 (3.1%)</td>
<td>198 (11.2%)</td>
<td>63 (4.8%)</td>
<td>18 (3.8%)</td>
<td>432 (5.0%)</td>
</tr>
</tbody>
</table>

Indicators concerning to community-level geopositioned surveys

Surveys by period

<table>
<thead>
<tr>
<th>Period</th>
<th>AFRO/EMRO</th>
<th>SEARO</th>
<th>WPRO</th>
<th>AMRO</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>before 1990</td>
<td>1,157 (25.0%)</td>
<td>688 (52.1%)</td>
<td>774 (73.2%)</td>
<td>316 (75.6%)</td>
<td>2,935 (39.6%)</td>
</tr>
<tr>
<td>1990-2000</td>
<td>441 (9.5%)</td>
<td>146 (11.1%)</td>
<td>116 (11.0%)</td>
<td>70 (16.7%)</td>
<td>773 (10.4%)</td>
</tr>
<tr>
<td>2000-2010</td>
<td>2,040 (44.1%)</td>
<td>486 (36.8%)</td>
<td>159 (15.0%)</td>
<td>32 (7.7%)</td>
<td>2,717 (36.6%)</td>
</tr>
<tr>
<td>2010-now</td>
<td>937 (20.3%)</td>
<td>-</td>
<td>9 (0.9%)</td>
<td>-</td>
<td>946 (12.7%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>49 (1.1%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>49 (0.7%)</td>
</tr>
</tbody>
</table>

Type of survey

<table>
<thead>
<tr>
<th>Survey Type</th>
<th>AFRO/EMRO</th>
<th>SEARO</th>
<th>WPRO</th>
<th>AMRO</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping/prevalence</td>
<td>3,875 (83.8%)</td>
<td>924 (70.0%)</td>
<td>954 (90.2%)</td>
<td>400 (95.7%)</td>
<td>6,153 (82.9%)</td>
</tr>
<tr>
<td>SS/Spot check</td>
<td>749 (16.2%)</td>
<td>280 (21.2%)</td>
<td>94 (8.9%)</td>
<td>18 (4.3%)</td>
<td>1,141 (15.4%)</td>
</tr>
<tr>
<td>TAS</td>
<td>-</td>
<td>6 (0.5%)</td>
<td>10 (0.9%)</td>
<td>-</td>
<td>16 (0.2%)</td>
</tr>
<tr>
<td>Passive recording</td>
<td>-</td>
<td>110 (8.3%)</td>
<td>-</td>
<td>-</td>
<td>110 (1.5%)</td>
</tr>
</tbody>
</table>

Diagnostic method

<table>
<thead>
<tr>
<th>Method</th>
<th>AFRO/EMRO</th>
<th>SEARO</th>
<th>WPRO</th>
<th>AMRO</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical</td>
<td>24 (0.5%)</td>
<td>119 (9.0%)</td>
<td>5 (0.5%)</td>
<td>1 (0.2%)</td>
<td>149 (2.0%)</td>
</tr>
<tr>
<td>Parasitological</td>
<td>1,939 (41.9%)</td>
<td>1,086 (82.3%)</td>
<td>892 (84.3%)</td>
<td>370 (88.5%)</td>
<td>4,287 (57.8%)</td>
</tr>
<tr>
<td>Serological</td>
<td>2,423 (52.4%)</td>
<td>69 (5.2%)</td>
<td>114 (10.8%)</td>
<td>30 (7.2%)</td>
<td>2,636 (35.5%)</td>
</tr>
<tr>
<td>Other</td>
<td>238 (5.1%)</td>
<td>46 (3.5%)</td>
<td>47 (4.4%)</td>
<td>17 (4.1%)</td>
<td>348 (4.7%)</td>
</tr>
</tbody>
</table>

MDA Implemented

<table>
<thead>
<tr>
<th>Status</th>
<th>AFRO/EMRO</th>
<th>SEARO</th>
<th>WPRO</th>
<th>AMRO</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>19 (0.6%)</td>
<td>36 (2.7%)</td>
<td>-</td>
<td>18 (4.3%)</td>
<td>84 (1.1%)</td>
</tr>
<tr>
<td>Pre-intervention</td>
<td>3,918 (84.7%)</td>
<td>827 (62.7%)</td>
<td>558 (52.7%)</td>
<td>312 (74.6%)</td>
<td>5,615 (75.7%)</td>
</tr>
<tr>
<td>Post-intervention</td>
<td>676 (14.6%)</td>
<td>457 (34.6%)</td>
<td>500 (47.3%)</td>
<td>88 (21.1%)</td>
<td>1,721 (23.2%)</td>
</tr>
</tbody>
</table>

*AFRO - African Regional Office, AMRO - Americas Regional Office, EMRO - Eastern Mediterranean Regional Office, SEARO - South East Asian Regional Office, WPRO - Western Pacific Regional Office.
limbs were recorded upon systematic population screening by headmen of cantons [102] - this survey was included as it provided the best nationwide data on occurrence for Thailand.

Figures 2 and 3 show the geographical distribution of survey locations by time period and by diagnostic method, respectively. The date of surveys varied between regions. The majority (71.9%) of data gathered for the AFRO/EMRO region were collected post-2000, whilst for other regions much of the available data were collected pre-2000. Parasitological-based diagnosis accounted for 92.04% of surveys before 2000, whereas 63.8% of the 4,065 surveys undertaken after 2000 used serological tests. Only 348 (4.7%) surveys used two or more diagnostic methods.

Figure 2 The global distribution of data points included in the Global Atlas of Lymphatic Filariasis database by period of time. Red = surveys undertaken post-2000 when GFLEP was launched, and blue = before 2000. Current endemic countries are displayed in white, non-endemic countries in grey and hatching depicts countries where endemicity is uncertain.
Factors associated with LF transmission
A subset of 6,562 surveys (4,933 occurrences and 1,629 absences) available in the assembled LF database along with the generated pseudo-absence data (5,000) were used to model the global distribution of LF. Figure 4 shows the marginal effect of each variable on the response, averaging across the effects of all other variables, and its relative contribution to the final BRT model. High suitability for LF is positively associated with precipitation in the wettest quarter (reaching a plateau at rainfall greater than 1000 mm), increased vegetation cover, population density and minimum temperature (increasing from a minimum value of 10°C), and negatively associated with increasing elevation (Figure 4).

Global LF transmission map
Figure 5A presents the global map of environmental suitability for LF transmission and suggests that this...
Figure 4 Marginal effect curves for each covariate used in the ensemble of 120 boosted regression tree (BRT) models. Black lines represent the mean marginal effect over all 120 BRT ensembles and grey envelopes the 95% bootstrap confidence interval. The y-axis is the untransformed logit response and x-axis is the full range of covariates values. The percentage values in parenthesis show the mean relative contribution of the covariate over all 120 sub-models of the ensemble.
suitability occurs primarily in tropical and sub-tropical regions, with the highest suitability in parts of Central and Latin America, West Africa, coastal east and southern Africa, India, Southeast Asia, Indonesia, Papua New Guinea and western Pacific. Validation statistics indicated high predictive performance of the BRT ensemble model with area under the receiver operating characteristic (AUC) of 0.81 (95% CI: 0.78 – 0.83; sd: 0.01). An environmental suitability threshold of 0.36 provided the best discrimination between presence and absence records (Figure 6) and this threshold value was used to classify the environmental suitability map into a binary map of the environmental limits of transmission. This map and the map of the current transmission limits (excluding areas known to be non-endemic) are shown in Figure 5.
LF transmission by region

Figures 7, 8 and 9 present the observed occurrence and absence of LF and environmental suitability for LF transmission for each region. In Africa, LF transmission is predicted to occur across much of coastal and savannah West Africa (Figure 7) but is restricted mainly to the coastal areas of east and southern Africa. The predicted distribution in central and southern Africa is uneven, with large foci in northeast of South Sudan (Upper Nile and Jonglei), Uganda, eastern Democratic Republic of Congo (Bas Congo, Bandundu and Equateur provinces), southeast Zambia and southern Malawi.

In the Americas region, LF transmission is predicted to occur throughout north and north-east regions of South-America, Central America, major islands in the Caribbean region (Haiti and the Dominican Republic) and marginally in coastal areas of southern United States (Figure 8). LF has been eliminated from 20 countries in the Americas region, and known current endemicity is restricted to Brazil, Guyana and the Hispaniola (Dominican Republic and Haiti) [99].

In Asia and western Pacific, LF transmission is predicted to occur in the east of India, Sri Lanka, much of Southeast Asia and southeast China, Papua New Guinea, the northern coast of Australia and southern Japan (Figure 9). LF has been eliminated in China (2007), Japan (1980s) and South Korea (2008), but the predicted environmental suitability corresponds well with the known historical, pre-control distribution [11,17,20] (Figure 10).

Discussion

Here we present a first global map of the distribution and transmission limits of LF. This work is opportune as it provides a basis for tracking and interpreting progress in control over time and can define the pre-control transmission limits, which, in turn, can inform the intensity and duration of control [2,103]. Our work additionally provides insight to post-MDA surveillance by identifying areas of highest transmission which may be more prone to the resurgence of transmission following cessation of interventions.

In the current analysis we identify the environmental limits of potential transmission. We demonstrate that the probability of LF transmission increases with increasing precipitation, temperature and certain vegetation types but decreases with increasing altitude (Figure 4). These findings are consistent with previous analyses of environmental correlates at continental [49,50] and country [52,104,105] scales and are, undoubtedly, linked to temperature-related variation in vector survival and parasite development within the vector [82,84,85,106].
occurs throughout the forest and savannah regions of West Africa, coastal east Africa and Madagascar and restricted foci in central and southern Africa. Suitable environmental conditions also occur across tropical areas of south and South-east Asia and the Pacific as well as large areas of Central and South America, including the southern States of America. Interestingly, however, active transmission in Central and South America is restricted to isolated foci; the possible reasons for this discrepancy are discussed below. Our predictions of environmental suitability in temperate regions are consistent with documented historical distributions prior to large scale intervention and local elimination in Japan, South Korea, and southern China [15-19] (Figure 10), the north coast of Australia [107] and southeastern coast of United States [99].

The probability of LF transmission is additionally associated with population density. Such an association probably reflects differences in the distribution of different vector species and their habitat preference and susceptibility to LF [108-110]. In rural areas of Africa, LF is transmitted by *Anopheles* species, whereas in urban settings in east Africa and the Nile Delta transmission is by *Culex quinquefasciatus* [111,112]. In West Africa, *Culex* mosquitoes, although widely distributed [113], are considered refractory to infection [114-116], although some studies have demonstrated compatibility between West African strains of *W. bancrofti* and *Culex* and *Mansonia* mosquitoes [117-120]. *Culex* transmission also occurs in Asia [121-126] and *Culex quinquefasciatus* is the only vector known in the Americas [127,128]. The true extent of LF transmission in urban settings, especially in sub-Saharan Africa, remains poorly documented [30,129,130], and further work is warranted.

Our work provides interesting insight into the regional distributions of LF. In sub-Saharan Africa, LF transmission is highly heterogeneous (Figure 7), with the highest potential risk in the forest and savannah regions of West Africa and coastal areas of eastern Africa and Madagascar. Scattered and relatively small areas of high-moderate risk are predicted in central Africa. This distribution of LF across Africa corresponds well with the known historical distribution of LF on the continent [17,18] and predicts transmission in countries which have yet to be extensively mapped, such as Gabon [131].
or Angola [132]. Our predictions indicate a more geographically restricted distribution of LF (Figure 7) than earlier spatial predictions using environmental factors [49,50]. Our analysis included some 4,624 surveys across most endemic countries in Africa, whereas previous analysis [49,50] included fewer than 700 surveys which were mainly concentrated in West Africa, Egypt, Sudan, Kenya, Tanzania and Madagascar; studies which were typically conducted in known areas of transmission. Such paucity and biased clustering of data coupled with the use of regression-based modelling will result in the smoothing of prevalence across large areas and overestimate prevalence in unsurveyed areas [133,134]. Our use of BRT modelling overcomes the geographical bias of data by the use of pseudo absences [135], which are randomly generated from areas known to be unsuitable for mosquito breeding or non-endemic for LF.

Our risk map predicts widespread environmental suitability of LF in Central and South America, whereas active transmission is known to occur only in isolated foci [136]. We suggest two possible reasons for this discrepancy. First, historically, LF occurred across 20 countries and territories in the region [19,99], including islands of the Caribbean and coastal areas of southeast United States, where indigenous cases were reported as far north as Philadelphia, until 1930s [99]. A combination of intensive vector control and improvements in public sanitary works resulted in the local elimination in these settings. Second, the transmission of LF in the Americas is strongly influenced by historical socioeconomic and demographic factors, rather than environmental factors. Transmission of LF was introduced to the Americas by slaves transported from West Africa to work in sugarcane plantations during the 17th and 18th centuries [137]. Initially, large numbers of slaves were brought to Barbados and Brazil, and subsequently sent to other islands of the Caribbean, to the north American colonies (South Carolina and Virginia) and northern countries of South America (Venezuela, Guyana, French Guiana, and Suriname) [17,19,99]. Upon introduction, *W. bancrofti* readily adapted from transmission by African *Anopheles* mosquitoes to transmission by *Culex* mosquitoes commonly found in the overcrowded and insanitary towns and cities of the Americas. However, it appears that

![Figure 8 Reported and predicted distribution of lymphatic filariasis in the Americas. (A) Observed occurrence and absence of LF and (B) environmental suitability for lymphatic filariasis transmission, as predicted by the final boosted regression trees model, in the Americas.](http://www.parasitesandvectors.com/content/7/1/466)
transmission has remained restricted to those areas where the disease was firstly introduced. In settings where socioeconomic and sanitary conditions have improved and interventions have been implemented in recent decades, the disease has gradually disappeared, for example in the Caribbean region [138]. In those countries with active transmission today - Brazil, the Dominican Republic, Guyana and Haiti [39,139], infection occurs mainly in urban settings and is strongly associated with poor socioeconomic conditions [140-142]. Future work will explore the interplay of environmental, socioeconomic factors and coverage of interventions on the distribution and prevalence of LF.

The distribution of LF in Asia exhibits marked regional trends. In India, high transmission occurs across the northern Indian (Gangetic) plain which borders Nepal and Bangladesh, eastern and south western coastal areas, areas in the central Deccan region in the south and on the Andaman and Nicobar Islands (Figure 9). Moderate-low transmission occurs in inland areas of southwestern India. Such a distribution is consistent with previous district-level mapping [143] (Figure 10D) and a previous national-level environmental risk model [104]. The predicted environmental limits for China, Japan and South Korea correspond well with the historical distribution prior to large control and local elimination [20,144,145] (Figure 10A-B). Although LF has been declared eliminated in China, the vectors remain and occasional imported cases have been reported [146], so that long-term surveillance is required to ensure that recrudescence does not occur. In contrast to the predictions for India and east Asia, our environmental-based map overestimates risk in the southeast Asian countries of Vietnam, Cambodia, Thailand and PDR Laos, which are considered to have limited transmission [147]. In Thailand, B. malayi transmission occurs in the south of the country [102] and W. bancrofti is endemic in the western provinces bordering Myanmar [148], but control measures implemented since the 1960s have dramatically reduced transmission [102].

We have sought to conduct an exhaustive search assembly of historical and contemporary data on LF occurrence and prevalence and have used rigorous methodology to
predict the occurrence of LF transmission, but recognise a number of limitations. The BRT model presented here is driven by environmental parameters and spatial configuration of habitats that allow persistence of species in landscapes. This modelling approach has been applied successfully to map the distribution of mosquito-borne diseases, such as dengue and malaria, which are transmitted by one mosquito genera with limited species diversity within geographical regions [96,149]. LF is unique among mosquito-borne diseases because it is transmitted by mosquito species belonging to five genera: *Aedes*, *Anopheles*, *Culex*, *Mansonia* and *Ochlerotatus*. In Africa, where LF is transmitted principally by *Anopheles* species, the disease distribution corresponds well with the known historical distribution patterns across the continent. This is not surprising because the environmental factors determining the abundance and distribution of *Anopheles* mosquitoes across Africa have not changed much in the peri-domestic environment in the past 20 years. However, our model performs less well in areas where LF is transmitted by *Culex* mosquitoes, especially *Culex quinquefasciatus*, which is predominately an urban mosquito with breeding habits that are influenced more by human activity than environmental factors [150,151]. It follows, therefore, that environmental factors impact on transmission differently for different mosquito species. For example, precipitation, an important determinant in our environmental model, is likely to affect *Anopheles* and *Culex* species differently. While frequent rainfall generally tends to increase the densities of adult mosquitoes by producing breeding sites, their densities may be reduced by the flushing of sites when precipitation is high. It may be that anopheline and culicine mosquitoes differ in their response to heavy rainfall since they breed in different habitats. In areas suitable for transmission by *Culex quinquefasciatus*, human factors, such as sanitary and housing conditions, may play an important role in transmission. Our model may have therefore overestimated the current limits for LF in *Culex*-transmitted areas.

We additionally recognise the limitations of the current empirical evidence base, especially in regard to unpublished surveys that found an absence of transmission. We did not perform age correction of prevalence data since our current focus was on the occurrence of transmission.
For similar reasons we did not adjust for differences in sensitivity of various diagnostic methods [152]. Future work will seek to model the prevalence of LF infection and will take into account such differences in age patterns and diagnostic method. Finally, our study focused on the environmental limits of LF transmission and, as discussed above, we recognise our model does not capture the socioeconomic and intervention-related dynamics of transmission. This is the subject of ongoing work investigating the spatiotemporal distribution of LF and the degree to which changes are related to the scaling-up of interventions or other factors.

Conclusions
Despite the limitations and caveats acknowledged above, the assembled database represents a unique resource and the global map provides the best currently available indication of the global distribution of LF, past and present. Consistent with the open access approach of the Global Atlas of Helminth Infection, the assembled data and developed maps are publicly available (www.thiswormyworld.org). As the global LF community moves towards elimination, the assembled data maps and model predictions will help track progress and increase the cost-effectiveness of surveillance activities post-control.

Endnotes
4AFRO/EMRO region: Seychelles and Mayotte; AMRO region: Antigua & Barbuda, Costa Rica, Cuba, French Guiana, Martinique, Puerto Rico, Saint Lucia, Saint Vincent and the Grenadines, Suriname, Trinidad & Tobago, Venezuela and Virgin Islands, United States of America; and WPRO region: Cook Islands, Palau and Solomon Islands.

Additional file
Additional file 1: Characteristics of surveys included in the lymphatic filariasis database (extended table).

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
JC and MPR have equally contributed to develop this work. JC, RLP, MJB and SJBJ conceived the idea for the paper. JC, MPR, TCL, AS and SWL collected the epidemiological data and developed the global database. NG and JC carried out the spatial and statistical analysis. JC and MPR wrote and SJBJ revised the first draft of the manuscript and all authors contributed to subsequent draft, and agreed to the final version.

Authors’ information
Jorge Cano and Maria P Rebollo joint first authors.

Acknowledgements
We thank all the authors who have provided access to survey data – for a full list of acknowledgements see: www.thiswormyworld.org. We would also like to thank David Addiss and Mark Bradley for their comments at early stages of the work. This work was funded by grants from the Bill & Melinda Gates Foundation (OPP1053338: JC, NG, SWL, SJBJ, OPP1033751: JC, RLP, SJBJ), GlaxoSmithKline (JC, SJBJ, MJB) and the United Kingdom Department for International Development (MPR, LKH and MJB) and Senior Research Fellowships from the Wellcome Trust (095066: SJBJ, 098046: SJBJ). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author details
1Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom. 2NTD Support Center, Task Force for Global Health, Emory University, Atlanta, United States of America. 3Department of Parasitology, Centre for Neglected Tropical Diseases, Liverpool School of Tropical Medicine, Liverpool, United Kingdom. 4Department of Zoology, Spatial Ecology and Epidemiology Group, Tinbergen Building, University of Oxford, South Parks Road, Oxford, UK. 5School of Biological and Biomedical Sciences, Durham University, Durham, United Kingdom.

Received: 1 July 2014 Accepted: 29 September 2014
Published online: 11 October 2014

References


