

# SandflyMap: leveraging spatial data on sand fly vector distribution for disease risk assessments

Desmond H. Foley, Richard C. Wilkerson, L. Lynnette Dornak, David B. Pecor, Arpad S. Nyari, Leopoldo M. Rueda, Lewis S. Long, Jason H. Richardson

*Entomology Branch, Walter Reed Army Institute of Research, 503 Robert Grant Avenue, Silver Spring, MD 20910, USA*

**Abstract.** We feature SandflyMap ([www.sandflymap.org](http://www.sandflymap.org)), a new map service within VectorMap ([www.vectormap.org](http://www.vectormap.org)) that allows free public online access to global sand fly, tick and mosquito collection records and habitat suitability models. Given the short home range of sand flies, combining remote sensing and collection point data give a powerful insight into the environmental determinants of sand fly distribution. SandflyMap is aimed at medical entomologists, vector disease control workers, public health officials and health planners. Data are checked for geographical and taxonomic errors, and are comprised of vouchered specimen information, and both published and unpublished observation data. SandflyMap uses Microsoft Silverlight and ESRI's ArcGIS Server 10 software platform to present disease vector data and relevant remote sensing layers in an online geographical information system format. Users can view the locations of past vector collections and the results of models that predict the geographic extent of individual species. Collection records are searchable and downloadable, and Excel collection forms with drop down lists, and Excel charts to country, are available for data contributors to map and quality control their data. SandflyMap makes accessible, and adds value to, the results of past sand fly collecting efforts. We detail the workflow for entering occurrence data from the literature to SandflyMap, using an example for sand flies from South America. We discuss the utility of SandflyMap as a focal point to increase collaboration and to explore the nexus between geography and vector-borne disease transmission.

**Keywords:** database, distribution, SandflyMap, VectorMap, leishmaniasis, georeferencing, geographical information system.

---

## Introduction

Vector distribution and demography are important but neglected components of disease transmission. In the past, online resources were not adequate to allow scientists to freely explore occurrence records and simultaneously access vector and disease models on a global scale. This was especially unfortunate for vectors and diseases that are focal, such as sand flies and leishmaniasis, where combining remote sensing and collection point data could give a powerful insight into the environmental determinants of distribution. Such online resources could help us understand the spatial relationship of vectors and disease for any area of interest, including arthropod-borne neglected tropical diseases (NTDs).

MosquitoMap ([www.mosquitomap.org](http://www.mosquitomap.org)) was established to address some of those needs. It is a free public online spatial database and geographical information system (GIS)-type viewing and analysis platform for global

occurrence records and habitat suitability models of many of the major mosquito vectors of disease (Foley et al., 2010). MosquitoMap was recently rebuilt and incorporated within a new umbrella application, VectorMap ([www.vectormap.org](http://www.vectormap.org)), that serves global distribution data for mosquitoes, ticks ([www.tickmap.org](http://www.tickmap.org)) and sand flies ([www.sandflymap.org](http://www.sandflymap.org)). These arthropods transmit some of the worst NTDs, such as leishmaniasis, dengue, Japanese encephalitis, yellow fever, Rift Valley fever, lymphatic filariasis, tick-borne relapsing fever and bartonellosis. This paper focuses on the map service within VectorMap called SandflyMap. To illustrate the published literature data capture process for SandflyMap, we describe the digitization and georeferencing of South American sand fly collection records from Martins et al. (1978). We explore the possibilities of this application to enhance the study and understanding of disease distribution world wide, but particularly of leishmaniasis in South America.

## Materials and methods

### *SandflyMap website*

Vector data were stored in an ESRI ArcGIS 10 Server platform Enterprise SDE/SQL Server 2005/2008

---

Corresponding author:  
Desmond H. Foley  
Walter Reed Army Institute of Research  
Smithsonian Institution MSC MRC534  
4210 Silver Hill Rd. Suitland, MD 20746, USA  
Tel. +1 301 238 1069; Fax +1 301 238 3168  
E-mail: [foleydes@si.edu](mailto:foleydes@si.edu)

Standard geodatabase. The map viewer used ArcGIS API for Microsoft® Silverlight, which provides mapping and geoprocessing capabilities from ArcGIS Server and Bing™ services. The application server is maintained at the University of Kansas (KU) through a collaboration between the Walter Reed Army Institute of Research (WRAIR), Walter Reed Biosystematics Unit (WRBU) and the KU Biodiversity Institute. The SandflyMap viewer requires installation of Silverlight, a free web-browser plug-in. Mosquito, tick and sand fly ArcMap documents were constructed as linked map services, as were services for disease and vector habitat suitability models. Vector habitat suitability models (see below) were grouped according to geographic range (global, Africa, Asia, Oceania, North America, South and Central America, Palearctic region). Web sites were constructed for vectors ([www.vectormap.org](http://www.vectormap.org)), ticks ([www.tickmap.org](http://www.tickmap.org)) and sand flies ([www.sandflymap.org](http://www.sandflymap.org)) that link to the one already available for mosquitoes ([www.mosquitomap.org](http://www.mosquitomap.org)), and whose construction has been described previously (Foley et al., 2010).

#### *SandflyMap case study*

Martins et al. (1978) provided extensive collection records for South and Central American sand flies gleaned from the literature and from the author's own collections. These records are also the basis for many of the records and maps in Young and Duncan (1994), however, location data were given as town and state, and maps cannot reliably be used to obtain georeferenced occurrence data. Therefore, we scanned a copy of Martins et al. (1978) and enabled optical character read (OCR) within the PDF, to allow copying and pasting of relevant text into Microsoft® Excel. Georeferences were obtained with the online gazetteer exploration tool, Biogeomancer Version 1.2.4 (<http://bg.berkeley.edu/latest/>). Search strings were arranged with as much geographic information as possible to minimize errors, i.e. precise location, and state and country. Where only one record for a populated place was obtained this was assumed to be the correct location. In such cases, the gazetteer details, the geocoordinates, and the estimate of spatial error were copied from Biogeomancer. Spatial error in Biogeomancer is estimated by the point radius method (see <http://manisnet.org/GeorefGuide.html>). The point radius method (Wieczorek et al., 2004) portrays error as a radius around a geocoordinate. Having an error estimate for a collection georeference allows the user to judge the spatial resolution of remotely sensed data

that are appropriate (Foley et al., 2009).

In other cases, no records were identified, only non-populated place names were identified, or more than one populated place was identified. Where no record was identified, this was often a result of errors in the spelling or the wrong current name for the state, for instance Pôrto Feliz rather than Porto Fleiz, or Mato Grosso rather than Mato Grosso do Sul. Internet searches were useful to resolve spelling and locations in many cases. Where more than one populated place was obtained these locations were either clustered or far apart. When possible locations were close to one another (e.g. less than 25 km apart), an attempt was made to give one location to encompass all possibilities. For example, two locations for Aracruz, Espirito Santo, Brazil were obtained that were 20 km apart. In that case the Global Administrative Areas (GADM) administrative category that they occurred in was chosen to represent the location, but in other cases one representative location was chosen and the estimate of spatial error manually increased to cover the extent of all locations. When more than one populated place was identified that were far apart (e.g. >25 km apart), these records were set aside for future georeferencing, when more information may become available. For example, "Esperanza, Belize" resulted in four population centres, one archeological site and one man-made feature. In all cases, information about georeferencing decisions was associated with the collection record in SandflyMap.

Species names were those from the WRBU Catalog of Subfamily Phlebotominae (<http://www.sandflycatalog.org/>), which incorporates the classification of Galati (2003).

#### *Vector models*

Ecological niche models were generated using maximum entropy - Maxent 3.2.1 (Phillips et al., 2006) for species that had sufficient collection points (>10) and were medically important. These models are designed as placeholders to display the functionality of VectorMap until other models become available. Occurrence points were scrutinized in ESRI ArcMap and any points that were dubious (e.g. not falling in the correct administrative area) were trimmed from the dataset. A merged dataset was created in ArcMap by merging all points for a region. For modelling, a total of 16 layers were used: 10 BioClim variables (1, 4, 5, 6, 9, 12, 13, 14, 15, 19) (<http://www.worldclim.org/bioclim>), five variables from HYDRO1k ([http://eros.usgs.gov/#/Find\\_Data/Products\\_and\\_Data](http://eros.usgs.gov/#/Find_Data/Products_and_Data)

\_Available/gtopo30/hydro) (h\_aspect, h\_dem, h\_flowacc, h\_flowdir, h\_slope, h\_topoind), and one variable from the Harmonized World Soils Database (<http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/>). Continental Australia lacks three hydro1k layers (h\_topoind, h\_flowacc, h\_flowdir) for MaxEnt runs done in Australo\_Papua, Asia\_Pacific and Global extents. All 16 variables were clipped for geographic regions via a mask (Spatial Analyst Tools/Extraction/Extract by Mask). Raster versions of the clipped datasets were converted to ASCII format (Conversion Tools/Raster to ASCII). MaxEnt model parameters included: 1,000 iterations, 25% random testpoints, minimum training presence threshold (MTPT) applied and random seed. ASCII files were converted to raster grids (Conversion Tools/To Raster/ASCII to Raster) with floating output type for the base model and to grids with integer output type for the MTPT model. The float grid was multiplied by 1,000, converted to an integer and then multiplied by the MTPT model (Spatial Analyst/Raster Calculator). This process produced a model with continuous data restricted by the MTPT.

## Results

### SandflyMap website

The SandflyMap client side website is modeled after the MosquitoMap website described in Foley et al. (2010). Sand flies are available as a vector theme within the new Silverlight VectorMap viewer application. Many of the features that were described by Foley et al. (2010) are retained in the Silverlight version. Fig. 1 shows the layout of the opening screen and some of the features. The viewer gives the user access to a variety of base maps: ESRI Topo, ESRI Street, ESRI Aerial, Bing Streets, Bing Aerial, Bing Aerial with labels. A variety of layers are provided depending on the vector theme. For sand flies these include: type or holotype collection locality, spatial uncertainty, GADM administrative boundaries, populated places, streams, water bodies, built-up areas and human density (for details see under the Metadata tab within [www.mosquitomap.org](http://www.mosquitomap.org), [www.sandflymap.org](http://www.sandflymap.org), and [www.tickmap.org](http://www.tickmap.org)). Transparency slider bars allow overlapping layers to be visualized. The viewer

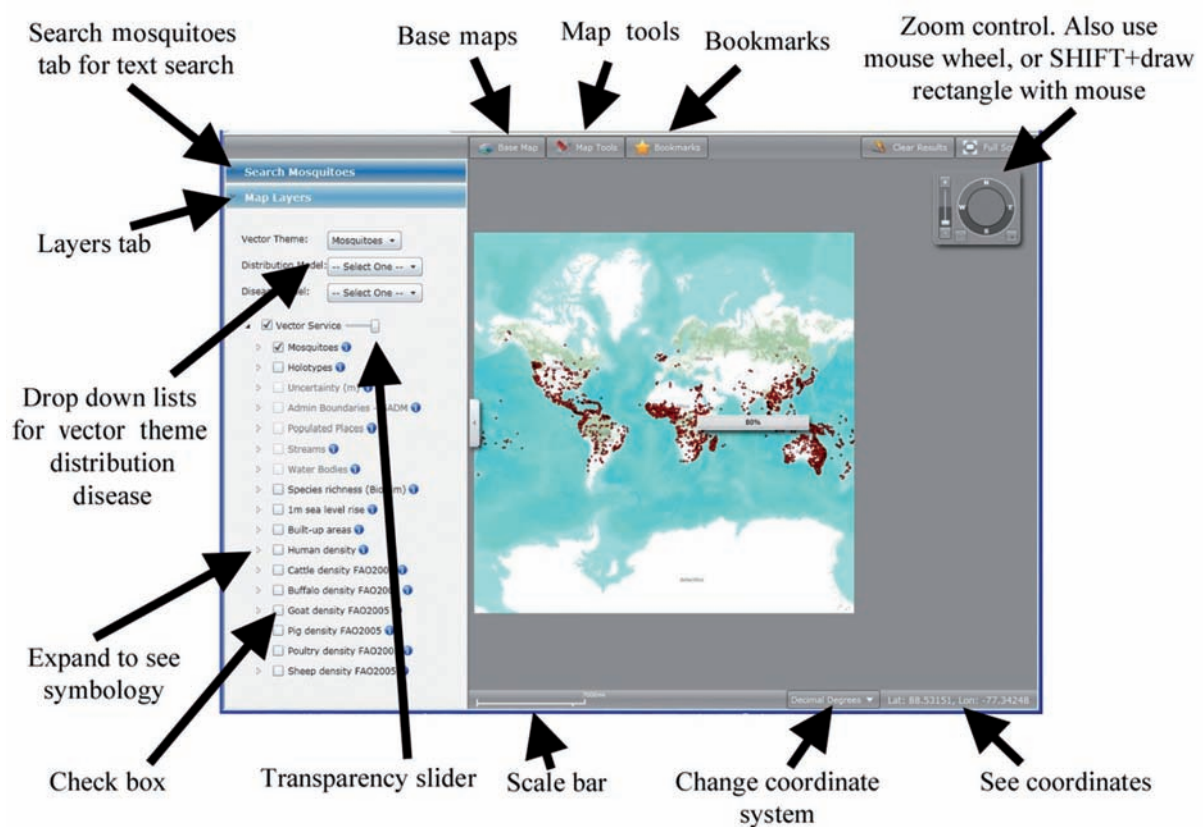


Fig. 1. The opening screen of the application viewer for VectorMap with various features arrowed.

gives access to mosquito, sand fly and tick collection records. Up to 10,000 records can be exported to Excel and to Google Earth as a KML file. As of December 2011, SandflyMap contains over 7,000 collection records, of approximately 280,000 within VectorMap.

#### *SandflyMap case study*

A total of 2,750 records (Fig. 2) for 209 species and 515 locations were georeferenced at the time of writing, with approximately 2,300 remaining to be georeferenced. Mean spatial accuracy was  $8,054 \pm 3,723$  m (range = 2,654-63,000 m). This suggests that ecological niche models (ENMs) constructed with these data, and remote sensing layers compared to these data should be on average not more than 8 km (10 arc minute) in resolution.

#### *Vector models*

An example of the ENMs available in SandflyMap is given in Fig. 3 for *Nyssomyia intermedia* (Lutz & Neiva). As of 12 December 2011, 113 vector habitat suitability models were available in MosquitoMap, six

models in SandflyMap, and 40 models in TickMap, although many more are ready to be added. The details for these models are available under the Metadata tab within the client websites. Models are given global unique identifiers (GUIDs) that contain information about the institution, the vector type, vector species, author and model number (e.g. USNM:MosquitoModel:Ae\_africanus\_Nyari\_1). Within the metadata for each model, files for the ESRI grid, MaxEnt output, and a data point file (.cvs) are available for download, to enable the interested user to repeat each model if so desired.

#### *Disease models*

Expert opinion maps for visceral, mucocutaneous, cutaneous (*Leishmania major*) and cutaneous (*L. tropica*) leishmaniasis (WHO, 2010) have been included in SandflyMap. The map reproduced in SandflyMap is a re-rendered version of those available on the WHO site and may contain errors of translation. Briefly, images were copied, georeferenced, then polygons created to encompass coloured areas of presence on the original map. These shapefiles were then converted to raster format at 0.05 degree resolution. The relation-



Fig. 2. Map of collection points for multiple species that were georeferenced from Martins et al. (1978) and now are available in SandflyMap.



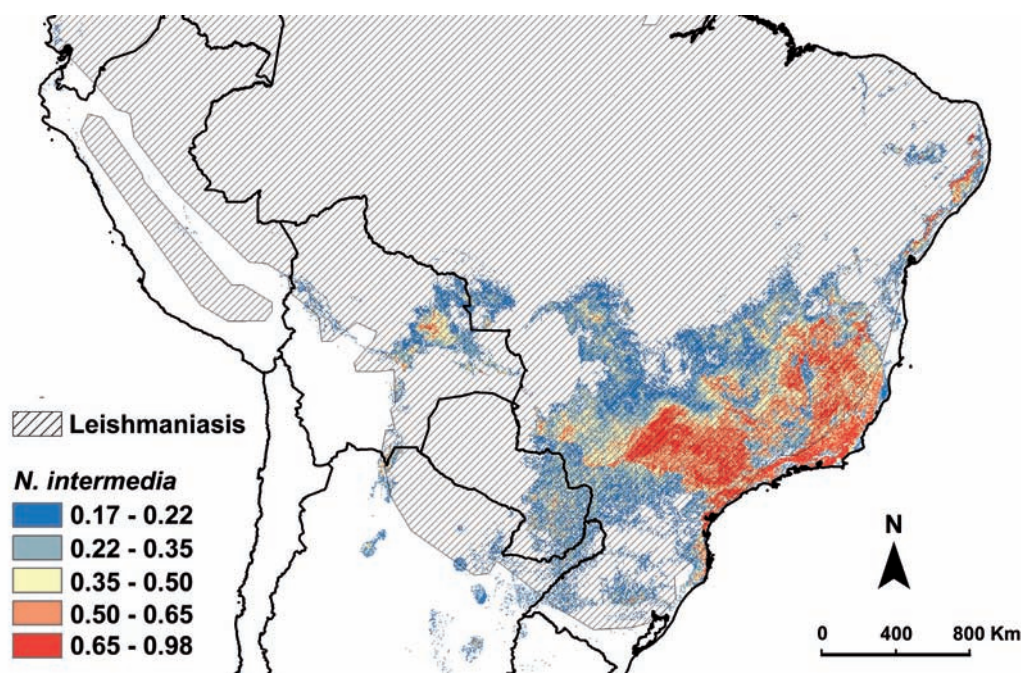


Fig. 3. Ecological niche model of the *Nyssomyia intermedia* (Lutz & Neiva) overlaid on the extent of cutaneous and mucocutaneous leishmaniasis, also available in the application SandflyMap.

ship of cutaneous and mucocutaneous leishmaniasis distribution with that predicted for *N. intermedia* is given in Fig. 3.

## Discussion

SandflyMap provides a home for, and makes freely accessible, collection data of all potential vector species within a vector group, not just those species that are currently recognized as vectors. We regard information on all vector species (primary and secondary vectors) in an area as important for understanding the sustainability of disease, the emergence of new vector/disease associations, and the challenges for vector control. SandflyMap is designed to provide distribution information to medical entomologists, vector disease control workers, public health officials and health planners. Vouchered collection data from museum specimens and both published and unpublished observation data, are checked for geographic and taxonomic errors, and made available as an interactive map for viewing in SandflyMap.

A few of the features of SandflyMap are:

- (i) datasets: high-quality historical and recent georeferenced collection data on vectors, pathogens, and their hosts that include fields such as geographic coordinate error; detailed, up-to-date taxonomy of vector and host species; full report-

ing of georeferencing procedures; GUIDs for digital tracking of collection instances; and a number of new vector-specific and epidemiologically significant data fields developed for SandflyMap;

- (ii) dataset histories;
- (iii) extensive metadata and downloadable GIS-ready files to the models and layers;
- (iv) records: access to published and unpublished records, as well as records from specimens;
- (v) online map viewer: MosquitoMap, TickMap, SandflyMap;
- (vi) flexible GIS environment to explore information;
- (vii) models: vector habitat suitability, disease occurrence. These will improve as new models are developed and included in VectorMap; and
- (viii) ability to see many vector- and disease-relevant GIS layers in one place.

Collection records are searchable and downloadable, and Excel charts to country, are available for data contributors to map and quality control their data (Foley, 2011). A downloadable Excel collection form for sand flies is available at <http://www.sandfly-map.org/contribute.htm>. This form contains drop-down lists, controlled vocabulary terms, a converter for degrees-minutes-seconds to decimal degrees format, and a mapper to visualize collection records on a map. This form is designed to assist data contributors with organizing their information and to minimize

error, not to replace their collection system. Original (vocabulary) information should be preserved, and this is important for assessing georeferencing uncertainty. The proper use of a global positioning system (GPS) device and a detailed text description of the locality are encouraged to minimize this uncertainty (Foley et al., 2009). The calculation of spatial uncertainty, for example for Martins et al. (1978), allows data to be matched to appropriate resolution remote sensing data, for modeling or other spatial analyses.

Tools under development include:

- (i) Mal-Area Calculator (MAC), a novel hazard/disease risk assessment tool. It is a raster calculator that focuses on where the vector hazard or disease risk is more likely to occur rather than on predicting when disease is more likely to occur. It takes into account spatial variation in the relative roles of different vectors and hosts, and quantifies these for an area of interest;
- (ii) VectorSurv, which allows the display of temporal data; and
- (iii) VectorHostMap, which records pathogen and ectoparasite data from vertebrates, such as those resulting from rodent surveys.

We are committed to high standards of data quality; free, open-access to data (to the extent permitted by data authors); and proper attribution of the sources of data. Submission of point data and vector and disease models to SandflyMap is encouraged, and data can be sent in any format to the corresponding author, or via [mosquitomap@si.edu](mailto:mosquitomap@si.edu). Point data will also be deposited with the Global Biodiversity Information Facility (GBIF). A search of GBIF for family Psychodidae on 10 December 2011 revealed 7,991 records with georeferences but most of these species are not relevant to disease, and very few records are present for Latin America. We invite readers to change this situation by using SandflyMap and by contributing their collection data and habitat suitability models.

### Acknowledgements

Funding for this work was provided by the Global Emerging

Infections Surveillance and Response System, a Division of the Armed Forces Health Surveillance Center, and from the Global Biodiversity Information Facility. This research was performed under a Memorandum of Understanding between the Walter Reed Army Institute of Research and the Smithsonian Institution, with institutional support provided by both organizations. The opinions and assertions contained herein are those of the authors and are not to be construed as official or reflecting the views of the Department of the Army or the Department of Defense.

### References

- Foley DH, 2011. A spreadsheet mapping approach for error checking and sharing collection point data. *Biodiversity Informatics* 7, 137-142.
- Foley DH, Wilkerson RC, Birney I, Harrison S, Christensen J, Rueda LM, 2010. MosquitoMap and the Mal-area calculator: new web tools to relate mosquito species distribution with vector borne disease. *Int J Health Geogr* 18, 9-11.
- Foley DH, Wilkerson RC, Rueda LM, 2009. Importance of the “what”, “when”, and “where” of mosquito collection events. *J Med Entomol* 46, 717-722.
- Galati EAB, 2003. *Classificação de Phlebotominae Flebotomíneos do Brasil* (EF Rangel, and R Lainson (ed.). Fiocruz, Rio de Janeiro, Brazil.
- Martins AV, Williams P, Falcao AL, 1978. American sand flies (Diptera: Psychodidae, Phlebotominae). *Academia Brasileira de Ciencias*, Rio de Janeiro.
- Phillips SJ, Anderson RP, Schapire RE, 2006. Maximum entropy modeling of species geographic distributions. *Ecol Model* 190, 231-259.
- WHO, 2010. WHO/NTD/IDM HIV/AIDS, Tuberculosis and Malaria (HTM) World Health Organization, October 2010. ([http://www.who.int/leishmaniasis/leishmaniasis\\_maps/en/index.html](http://www.who.int/leishmaniasis/leishmaniasis_maps/en/index.html)).
- Wieczorek J, Guo Q, Hijmans R, 2004. The point radius method for georeferencing locality descriptions and calculating associated uncertainty. *Int J Geogr Inf Sci* 18, 745-767.
- Young DG, Duncan MA, 1994. Guide to the identification and geographic distribution of *Lutzomyia* sand flies in Mexico, the West Indies, Central and South America (Diptera: Psychodidae). *Mem Am Entomol Inst* 54, 1-881 pp.