

Ecological niche modeling for visceral leishmaniasis in the state of Bahia, Brazil, using genetic algorithm for rule-set prediction and growing degree day-water budget analysis

Prixia Nieto¹, John B. Malone¹, Maria E. Bavia²

¹Department of Pathobiological Sciences, School of Veterinary Medicine, Louisiana State University, Baton Rouge, Louisiana; ²Preventive Medicine Department, Federal University of Bahia, Salvador, Bahia, Brazil

Abstract. Two predictive models were developed within a geographic information system using Genetic Algorithm Rule-Set Prediction (GARP) and the growing degree day (GDD)-water budget (WB) concept to predict the distribution and potential risk of visceral leishmaniasis (VL) in the State of Bahia, Brazil. The objective was to define the environmental suitability of the disease as well as to obtain a deeper understanding of the eco-epidemiology of VL by associating environmental and climatic variables with disease prevalence. Both the GARP model and the GDD-WB model, using different analysis approaches and with the same human prevalence database, predicted similar distribution and abundance patterns for the *Lutzomyia longipalpis*-*Leishmania chagasi* system in Bahia. High and moderate prevalence sites for VL were significantly related to areas of high and moderate risk prediction by: (i) the area predicted by the GARP model, depending on the number of pixels that overlapped among eleven annual model years, and (ii) the number of potential generations per year that could be completed by the *Lu. longipalpis*-*L. chagasi* system by GDD-WB analysis. When applied to the ecological zones of Bahia, both the GARP and the GDD-WB prediction models suggest that the highest VL risk is in the interior region of the state, characterized by a semi-arid and hot climate known as Caatinga, while the risk in the Bahia interior forest and the Cerrado ecological regions is lower. The Bahia coastal forest was predicted to be a low-risk area due to the unsuitable conditions for the vector and VL transmission.

Keywords: visceral leishmaniasis, ecological niche model, GARP, *Lutzomyia longipalpis*, *Leishmania chagasi*, geographical information systems, remote sensing.

Introduction

More than 90% of all cases of visceral leishmaniasis (VL), a complex multi-systemic disease with a case-fatality rate that approaches 90% if left untreated (Ahasan et al., 1996), in Latin America come from the northeastern part of Brazil. The State of Bahia has one of the highest incidence rates of VL

in Brazil (Grimaldi et al., 1989). *Leishmania chagasi*, an intracellular protozoan, is transmitted from one mammalian host to another by the bite of an infected female sandfly. *Lutzomyia longipalpis* (Lutz and Neiva, 1912) is the most important vector of VL in the New World. Quantitative comparisons of nDNA fragment patterns indicate that *L. chagasi* (the main etiologic agent of American VL) and *L. infantum* are very closely related and may actually represent the same species. There is increasing evidence that *L. chagasi* may represent strains of *L. infantum* that were introduced into the New World by earlier human migrants or their dogs from the Mediterranean region (Momen et al., 1987; Killick-Kendrick et al., 1980). Human beings are incidental

Corresponding author:
Prixia Nieto
School of Veterinary Medicine
Louisiana State University
South Stadium Road Baton Rouge, LA 70803, USA
Tel. +1 225 578 9671; Fax +1 225 578 9157
E-mail: pnieto1@lsu.edu

hosts, dogs are the main reservoir hosts among domestic animals, opossums (*Didelphis marsupialis*) and foxes (*Cerdocyon thous*) have been reported as the main wildlife reservoirs (Sherlock, 1996; Grimaldi et al., 1989).

VL was previously known as a rural disease, but large outbreaks and epidemics have recently been reported in large cities in both industrialized and developing countries in the Americas (WHO, 1995). In Brazil, VL is a re-emerging disease considered to be one of the most severe domestic public health problems (Arias et al., 1996). Climate change, deforestation, agrarian practices, a poorly nourished population and unsanitary housing make inhabitants of the State of Bahia easy prey for this disease (Sherlock, 1996).

Earlier studies have demonstrated that historical field and laboratory data regarding the ecology of vector-borne diseases can provide the basis for development of spatial-temporal environmental risk models by use of standardized analysis of Geographical Information Systems (GIS) and remote sensing (RS) data from earth-observing satellites (Malone, 2005; Kitron et al., 2006; Hay et al., 2001).

Temperature and moisture are fundamental determinants of the distribution and abundance of most species and are an essential part of any attempt to assess environmental risk factors of disease agents and their vectors. Each species has its own unique thermal-hydrological regime requirements that determine its biological or habitat preferences, limits of tolerance and optimum conditions for development. Growing degree day (GDD) – water budget (WB) analysis can be used to define the suitability gradient in the environment for propagation and transmission of the vector-parasite system. WB is a moisture accounting system used to indicate the presence of an adequate amount of surface water or soil moisture based on the water balance in the top layers of soil (Malone et al., 2001; Malone et al., 2003).

GIS and RS technologies are being used more frequently in community health studies to identify and classify risk areas as well as to predict the distribu-

tion of vector borne diseases (Thompson et al., 2002). In this study, we investigate the potential of (i) ecological niche modeling to predict the distribution of VL and its potential use as a risk assessment tool to guide control programs, and (ii) GDD and WB analysis to define a suitability gradient in the environment for propagation and transmission of the vector-parasite system.

Materials and Methods

Study area

The area under study included the whole of Bahia, the largest state of northeastern Brazil located between 8° 00' and 18° 30' South Latitude and 36° 00' and 46° 00' West Longitude. Bahia has an area of 567,295 km² and is the largest state in Northeast Brazil. The climate is classified as dry-arid to arid-tropical with annual average temperatures between 19° and 27°C. Tropical forest is the dominant vegetation within coastal areas and mangroves occur in the littoral areas, while the Cerrado and Caatinga areas are inland (Bavia et al., 2001).

Parasitological data

The parasitological data used consisted of annual human prevalence data from 1990 to 2000 for each of the 415 municipalities of the State of Bahia collected by the Fundação Nacional de Saúde, Salvador - Bahia.

Environmental data

These models were developed from environmental data layers from WorldClim (<http://bioge.berkeley.edu/worldclim/methods.htm>) and a global digital elevation model GTOPO 30 program (1 Km pixel resolution) from the United States Geological Survey (<http://edcdaac.usgs.gov/gtopo30/gtopo30.asp>) (Table 1). The WorldClim layers were based on interpolation of recorded climate data

Table 1. List of bioclimatic variables, derived from the monthly temperature and rainfall values (<http://biogeo.berkeley.edu/worldclim/bioclim.htm>) analysed in the study.

BIO 1 = Annual Mean Temperature
BIO 2 = Mean Diurnal Range Mean of monthly (max temp - min temp)
BIO 3 = Isothermality (P2/P7) (* 100)
BIO 4 = Temperature Seasonality (standard deviation *100)
BIO 5 = Max Temperature of Warmest Month
BIO 6 = Min Temperature of Coldest Month
BIO 7 = Temperature Annual Range (P5 - P6)
BIO 8 = Mean Temperature of Wettest Quarter
BIO 9 = Mean Temperature of Driest Quarter
BIO10 = Mean Temperature of Warmest Quarter
BIO11 = Mean Temperature of Coldest Quarter
BIO12 = Annual Precipitation
BIO13 = Precipitation of Wettest Month
BIO14 = Precipitation of Driest Month
BIO15 = Precipitation Seasonality (Coefficient of Variation)
BIO16 = Precipitation of Wettest Quarter
BIO17 = Precipitation of Driest Quarter
BIO18 = Precipitation of Warmest Quarter
BIO19 = Precipitation of Coldest Quarter
ALT = Topography (mean elevation)

from different weather stations to produce average monthly climate data as a 1 Km² resolution grid.

Ecological niche modelling

Ecological niche models were developed using the Genetic Algorithm for Rules-set Prediction (GARP) software program (<http://www.lifemapper.org/desktopgarp>). The GARP approach models the ecologic niche of species based on relating point-occurrence data to the electronic maps of relevant ecological dimensions, producing a heterogeneous set of rules that describe the potential distribution of species (Peterson et al., 2002b). Previous tests of the predictive power of this modeling technique have been published elsewhere (Peterson et al., 2001, 2002a,b; Stockwell and Peterson 2002; Peterson and Shaw, 2003).

Within GARP, input data are further divided randomly and evenly into training and intrinsic testing data sets. GARP works in an iterative process of rule selection, evaluation, testing, and incorporation or rejection. A method is chosen from a set of possibilities (e.g. logistic regression, bioclimatic rules), applied to the training data and a rule is developed

or evolved (Stockwell, 1999). Rules may evolve by a number of means that mimic DNA evolution: point mutations, deletions, crossing over, etc. the change in predictive accuracy from one iteration to the next is used to evaluate whether a particular rule should be incorporated into the model, and the algorithm runs either 1000 iterations or until coverage (Peterson et al., 2004a). GARP relates ecologic characteristics of occurrence points to those of ecologic characteristics sampled randomly from the rest of the study region, developing a series of decision rules that best summarize factors associated with presence (Peterson et al., 2000b).

To generate the niche models, WorldClim data variables and elevation data were processed using Arc-View 3.3 Spatial Analyst and DIVA-GIS 4.0 (<http://www.diva-gis.org/>). GARP was run using these environmental variables with two different groups of human leishmaniasis prevalence data, i.e. (i) the places where no cases were reported, and (ii) the sites where prevalence ranged from 1 percent to the highest observed percentage. This procedure was carried out for each of the eleven years (1990-2000) of available human VL data (Fig. 1). To optimize model performance, 100 replicate models were developed based on random 50-50 splits of available occurrence points. The procedure for choosing best subsets models was based on the observation that (i) models vary in quality, (ii) variation among models involves an inverse relationship between errors of omission (leaving out true distributional areas) and commission (including areas not actually inhabited), and (iii) best models (as judged by experts blind to error statistics) are clustered in a region of minimum omission of independent points and moderate area predicted (an axis related directly to commission error). The position of the cloud of points relative to the two error axes provides an assessment of the relative accuracy of each model (Costa et al., 2002) and (Peterson et al., 2004a).

To choose the best subsets models in this study the following procedure was carried out:

- (i) eliminate all models but those that had no omission (intrinsic) error based on independent

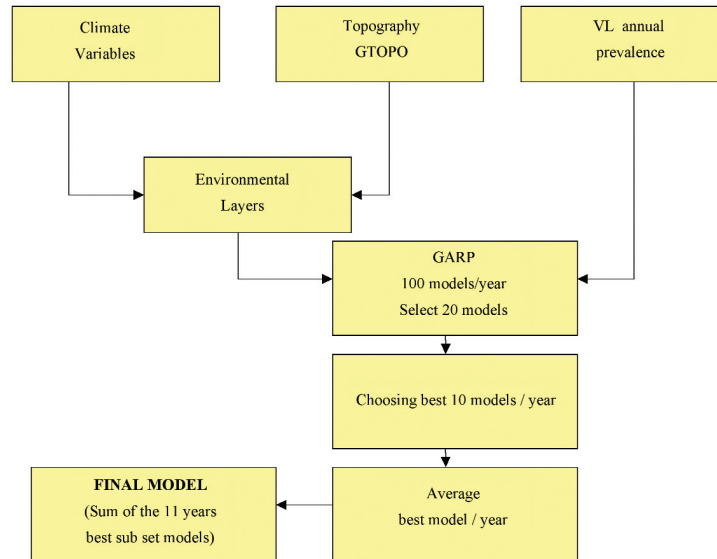


Fig. 1. Procedure to select the best subsets models in this study: (1) add climatic and environmental variables as well as VL data; (2) to optimize model performance, select 100 replicate models per year and choose 20 models with the least values in the omission (extrinsic) error; (3) identify the 10 models closest to the overall median area predicted; (4) average these 10 models to get a best subset annual model; (5) sum the final best subset annual models of the 11 years to get one final model.

tests points. Intrinsic omission is the percentage of the training points that are omitted from the prediction area; that is, those that are predicted to be absent but have presence records;

- (ii) select 20 models with the least values in the omission (extrinsic) error. Extrinsic omission is the percentage of the test points that are omitted from the prediction; that is, those that are predicted absent but have presence records;
- (iii) calculate the median area predicted present among these minimum omission points;
- (iv) identify the 10 models closest to the overall median area predicted;
- (v) average these 10 models to get a best subset annual model (Peterson et al., 2004);
- (vi) sum the 11 (year 1990-2000) final best subset annual models (Fig. 1).

Growing degree day-water budget model

GDD-WB analysis, a method widely used for crop

production models, can be used to define thermal-hydrological suitability gradients in the environment for propagation and transmission of vector-parasite systems. GDD can be used to indicate minimum temperature, optimum temperature, limiting high temperature thresholds and thermal energy required for development of generation times of vector-parasite systems (Malone, 2005). Water budget is a moisture accounting system used to indicate the presence of a suitable amount of soil moisture or surface water for development based on the water balance in the top layers of soil or ground surface (Malone et al., 2003).

Limited data is available on the environmental preferences and limits of tolerance of the *L. chagasi-Lu. longipalpis* system in relation to climate. Thus these factors were estimated by defining the climate regime associated with presence or absence of the parasite-vector within a GIS in combination with available published data. Resource data used to generate the GDD-WB model included a 30-year aver-

age monthly climate surface grid (18x18 Km) dataset of South America (Mud Springs Geographers, Temple, TX), a human VL prevalence database from the State of Bahia (Fundação Nacional de Saúde, Estado da Bahia) and development data obtained in different studies where *Lu. longipalpis* colonies were established and maintained in the laboratory (Killick-Kendrick et al., 1977). The objective was to define a suitability gradient in the environment for propagation and transmission of the vector-parasite system.

The 30-year average monthly climate surface grid of South America included data on maximum temperature, minimum temperature, precipitation and potential evapotranspiration (PET). These parameters were used to calculate mean monthly temperature $[(\text{max temp} - \text{min temp})/2]$ and water budget $[\text{PPE} = (\text{precipitation}/\text{PET})]$. Growing-degree-days (GDD) is defined as the number of the degrees over a base value (base temperature) below which no development of a species occurs, in this case the sand fly vector *Lu. longipalpis*. GDD can be accumulated for months in which conditions were within the suitability threshold range temperature and water budget threshold values to derive annual GDD values.

The water budget threshold value was defined by GIS analysis, as a condition where PPE, the ratio of rain/potential evapotranspiration (R/PET), was >0.7 , (i.e. the soil moisture content was 70% saturated). The PPE threshold was determined by extracting mean climate attribute data values from grid cells in which annual human prevalence data points were located for two infection prevalence ranges; moderate prevalence (1-5%) and high prevalence (5-14%) of VL. GIS map query functions were performed to confirm the PPE threshold of the presence of suitable amount of surface water or soil moisture; no positive sites were found if $\text{PPE} > 0.7$, suggesting a preference of the vector for drier environments.

The base temperature for development of *Lu. longipalpis* has not been experimentally determined. Thus a best estimate of the base temperature value was made by GIS query analysis using known endem-

ic site data in combination with data from the following literature reports on temperature requirements for reproduction and development of *Lu. longipalpis*:

- (i) a temperature of 25°C is the optimal temperature for reproduction and development in the laboratory of *Lu. longipalpis* (Killick-Kendrick et al., 1977; Mody and Tesh 1983; Rangel et al., 1996);
- (ii) the potential generations that can be completed in one year at 25°C for *Lu. longipalpis* is 7.7 generations per year (Killick-Kendrick et al., 1977);
- (iii) at 25°C, 40 days are required from engorgement of *Lu. longipalpis* female to the first adults emergence of the next generation (Killick-Kendrick et al., 1977);
- (iv) the average annual mean minimum temperature recorded in the State of Bahia for sites of $>1\%$ prevalence was 14.7°C, based on point-polygon extractions and GIS map queries on the 30-year-average monthly climate surface grid.

The base temperature of *Lu. longipalpis* was estimated using the reported laboratory development requirements (40 days for one generation, 7.7 generations per year at 25°C, 80% humidity) and the average annual mean minimum temperature value ($>14.7^\circ\text{C}$), in a two-step calculation:

- (i) the number of GDD that must be accumulated to complete one generation was estimated to be 412. For this estimate the following parameters were used: the optimum mean temperature in the laboratory (25°C) minus the minimum mean temperature registered at positive sites for *L. chagasi* in the State of Bahia (14.7°C) times days required from engorgement to first emergence of adults (40 days): $[(25^\circ\text{C} - 14.7^\circ\text{C}) \times 40 \text{ days}] = 412$;
- (ii) the number of GDD to complete one generation (412) and the potential generations that can be completed in one year (7.7) was used to solve for the base temperature (X): $[(25^\circ\text{C} - \text{X}) * 365]/412 = 7.7$; $\text{X} = 16^\circ\text{C}$.

A base temperature value of 16°C was accepted as the best estimate, pending more detailed laboratory +/-or field studies, of the minimum temperature in which no development of the sand fly vector *Lu. longipalpis* proceeds in the State of Bahia. Derivative climate variables were calculated, using 16°C as the base temperature, for each cell of the 18x18 Km climate grid, including monthly and annual values for:

- (i) Monthly GDD. This parameter was calculated as the mean temperature minus 16°C times days of the month (e.g. Jan = 31, Feb = 28, June = 30); the annual GDD is the sum of monthly GDD values;
- (ii) GDD times water budget (*PPE*), if $PPE < 0.7$. Using the formula $(GDD * (PPE < 0.7))$, a thermal hydrological gradient value was calculated that considered *both* thermal regime (GDD) and the influence of variable moisture regime for each month. The moisture threshold value < 0.7 was estimated by a GIS map query of the PPE value in grid cells that included positive *L. chagasi* points. This procedure was performed

for each month and multiplied by the number of days in the month to calculate the monthly value for the 'gradient model'.

Statistical analysis

The 11-year prevalence average data for VL in Bahia were analyzed with environmental data by stepwise logistic regression and by multiple regressions for the GDD-WB climate based model to establish associations between environmental variables and the prevalence of VL. The analysis was done using SAS software (Cory, NC). GARP is a statistical analysis software program.

Results

GARP ecological niche model

GARP analysis, based on 19 WorldClim bioclimatic data parameters and a topography data parameter for each of the 11 years in which preva-

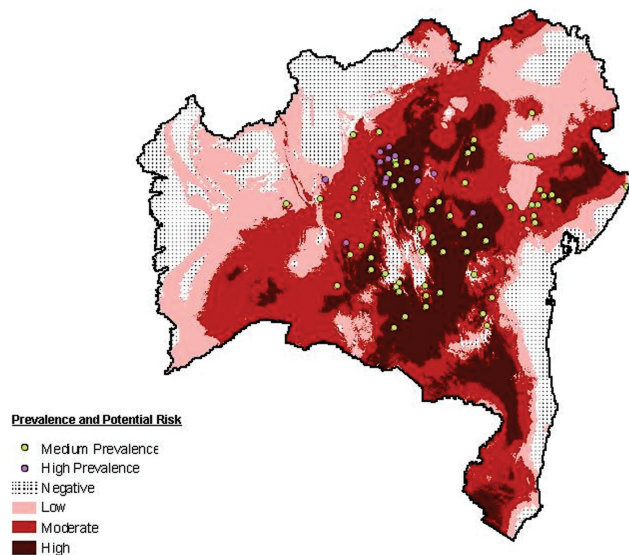


Fig. 2. Ecological Niche Model. The 11 best annual models were summed and ranked by the criteria of how many times each model predicted the same pixel within the endemic area, classified as: High, 11 times; Moderate, 6-10; Low, 1-5 and Negative, where none of the final models predicted the disease.

lence data was available is presented in Table 1. Data were further processed using ArcView 3.3 software to create and classify an eleven-year composite risk map (Fig. 2). The 11 best annual models were summed and ranked by the criteria of how many times each model predicted the same pixel within the endemic area, classified as high (11 times), moderate (6-10 times), low (1-5 times) and negative when none of the final models predicted the disease (Fig. 2). All of the high prevalence points fell into the predicted high risk area and all the middle prevalence points fell into the high or moderate risk area, while 32% of the negative prevalence points were predicted within the low risk areas. The rest of negative points fell in the area predicted as negative risk. The final model predicts as negative an ecological region known as the Bahia coastal forest.

The GDD-WB model

The predicted potential generation per year that can occur in various locations in Bahia ranged from

0 to 9 generations per year. The high and medium prevalence points were all found in regions predicted to have more than 5 generations per year (Fig. 3). The Caatinga ecological region was characterized as having more than 5 generations. It was the region with highest number of predicted potential generations per year. The Caatinga is characterized as a hot and semi-arid region suggesting that it would be suitable for the *Lu. longipalpis* vector. The Cerrado ecological region was predicted to have 2-4 potential generations per year and the Bahia coastal region 0-5 generations per year. The number of potential generations decreased at sites of greater proximity to the east and southeast. The GIS query analysis showed the maximum annual mean temperature for the development of *Lu. longipalpis* to be 28°C and the minimum annual mean temperature 16°C. All variables analyzed by stepwise logistic regression were significant for the occurrence of the disease (Table 2).

Multiple regression analysis was done to determine which environmental factors had the highest impact on the distribution and abundance of *Lu.*

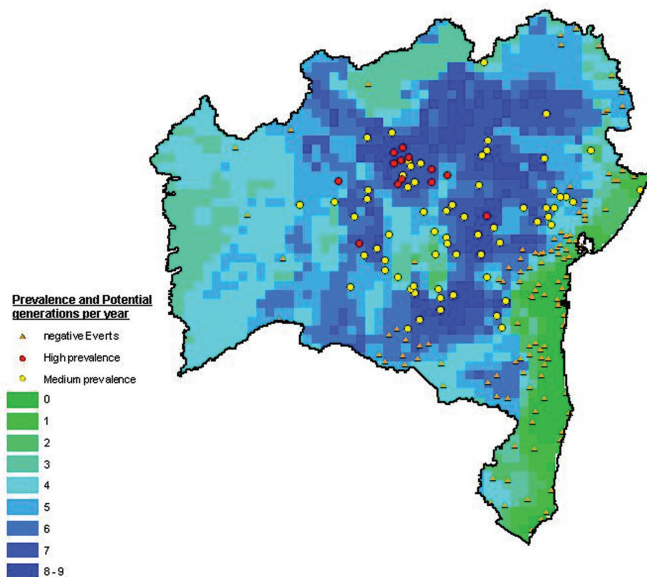


Fig. 3. GDD-Water Budget model. The predicted potential generation per year that can occur in various locations in Bahia ranged from 0 to 9 generations per year. The high and medium prevalence points were all found in regions predicted to have more than 5 generations per year

Table 2. Results of logistic regression analysis for the GDD-WB model. PPE, potential evapotranspiration; APG, annual potential generations; OR, odd ratios; CI confidence intervals.

Variables	OR	95% CI	P
PPE	0.997	0.996-0.998	<0.0001
APG	0.812	0.760-0.868	<0.0001

Table 3. Results of multiple regression analysis for the GDD-WB model. PPE, potential evapotranspiration; APG, annual potential generations.

Variables	Error	Type II	F value	Pr>F
PPE	0.37586	13.39	4.87	0.0277
APG	0.00084	32.68	11.87	0.0006
Ecoregions	0.04266	14.92	5.42	0.0202

longipalpis. Environmental factors that showed a significant influence in the presence and the distribution of *Lu. longipalpis* were water budget (rain/potential evapotranspiration = PPE); annual potential generations (APG); and ecosystems (Table 3).

Ecological zones

Comparison of the GARP ecological risk model

prediction and the GDD-WB model prediction of the distribution of VL in the State of Bahia to the ecological zone map of Bahia, suggests that the Caatinga ecological zone, a zone characterized by a hot and semi-arid environment, is the highest risk area and that the coastal region and the Cerrado ecological regions are lower risk areas (Fig. 4). The Northeast Brazil Dry Forest and the Bahia Interior Forest Zones may be intermediate transitional

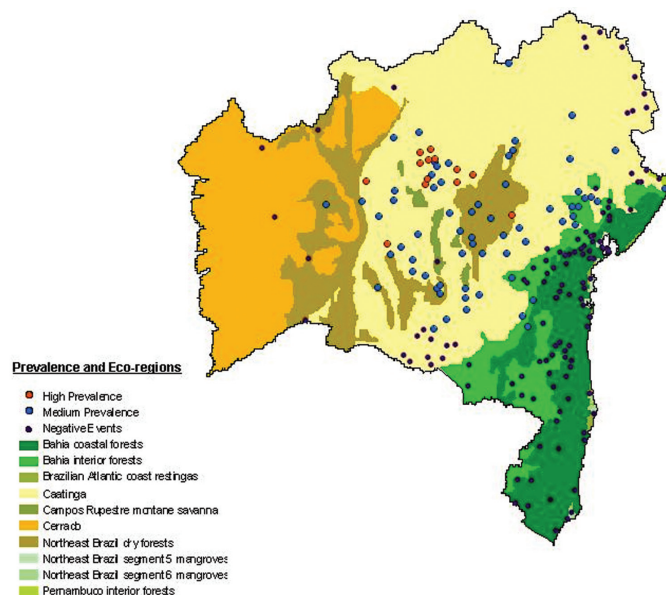


Fig. 4. Ecological Regions Bahia State. The final models GDD-WB and Ecological risk model predicts as negative an ecological region known as Bahia coastal forest. The high risk area corresponds mostly to the interior region of the state, matching with the ecological region known as Caatinga.

zones. Both the final GARP model and the GDD-WB generations per year model predicted major areas in the ecological region known as Bahia Coastal Forest as negative.

Using the GDD-WB model, the Caatinga ecological region was characterized as having more than 5 potential generations per year - it was the region with highest number of predicted potential generations per year. The Cerrado ecological region was predicted to have 2-4 potential generations per year and the Bahia coastal region 0-5 generations per year.

The number of potential generations per year decreased at sites of greater proximity to the east and southeast.

Discussion

Two climate-based risk models were developed for VL in Bahia, one using GARP statistical software and one using GDD-WB analysis. Both models, using different analysis approaches, predicted similar distribution and abundance patterns for the *Lu. longipalpis-L. chagasi* system in Bahia. Temperature and moisture are fundamental determinants of the distribution and abundance of a species and are an essential part of any attempt to assess environmental risk factors of disease agents and their vectors. Each species has unique thermal-hydrological regime requirements that determine its biological or habitat preferences, limits of tolerance and optimum conditions for development that determine where it is found (Pavlovsky, 1966). Where these requirements are unknown it is possible, within a GIS, to use site records of known positive and/or negative occurrence to define biological niche requirements with automated statistical packages (e.g. GARP).

Alternatively, it is possible to infer niche requirements by iterative step-by-step 'range-finding' of individual environmental feature layers extracted from known sites in a GIS. This would then be followed by extrapolation to areas with similar environmental features, or combinations of features, by

GIS query analysis of layers of statistically important environmental features (Malone, 2005). This approach was used in current studies to define the *Lu. longipalpis-L. chagasi* system as to its thermal-hydrological niche in Bahia using GDD-WB climate analysis and GIS query analysis based on known VL prevalence point records. The latter is a unique capability made possible by geospatial analysis tools available to biologists using GIS methodologies.

The GARP model - It is quite clear that this approach works well since the resulting 11-year-composite map of high, moderate, low or no risk of VL corresponded with actual disease prevalence records. Importantly, all of the high prevalence points fell into the predicted high risk area, all of the middle prevalence points fell into the high or moderate risk area, and 32% of the negative prevalence points were predicted within the low risk areas. Prevalence points with less than 1% were not considered in this model, since sites with <1% may represent potential migratory or sporadic cases.

The variation in area covered by annual GARP risk prediction maps over the 11-year study period as compared to reported cases of VL suggests that annual variation in geospatial risk may occur in individual years in endemic areas due to annual climate variation. Further study in Bahia and other VL endemic areas is warranted to confirm and extend this observation for use in development of annual risk prediction methods. The risk map shown in Fig. 2 is compatible with fundamental principles described for the distribution and abundance of species where there is a zone of more or less stable presence, outside of which there is a transitional zone of patchy presence or absence in marginally suitable conditions surrounded by an unsuitable zone where the species is absent (Andrewartha and Birch, 1954).

The GDD-WB model - Development of climate risk models using GDD-WB analysis depends on availability of data on the environmental preferences and limits of tolerance of the *L. chagasi-Lu. longipalpis* system in relation to climate. Only incomplete published data from the field or from laboratory studies are available for *L. chagasi-Lu.*

Table 4. Developmental times, temperature, adults longevity and productivity of *Lu. longipalpis* colony as observed by different authors.

Stages, longevity and productivity	Developmental times (in days)
From blood meal to the first adult emerge	- 35 ³ -40 ¹ , at room temperature 25 °C ¹ , relative humidity 80% ³ ; - 48-54 ¹ and 25-42 ² , breed in less constant conditions
Hatching	- 4-9 ¹ and 6-9 ²
Larvae	- 14-19 ² , for the four instars
First instar	- 3-5 (first week) ¹
Second instar	- 2-4 (second week) ¹
Third instar	- 1-5 (second week) ¹
Fourth instar	- 3-9 (third week) ¹
Pupae	- 10 ¹ , most of the adults, and 8 -9 ² ; few males emerges as early as 7 ¹
Adults longevity	- In general, adults are robust in rough conditions for 2 weeks-one month; - In blood fed females longevity is determined by oviposition; few survive 24 h after oviposition. - Most live for more than one month ¹
Productivity	- 23 generations in 36 months ¹
Cultures temperature of promastigotes and amastigotes	- Promastigotes at 22 °C, amastigotes at 34 °C, reversion from mastigotes to promastigotes at 28 °C ⁴

¹Killick-Kendrick et al., 1977; ²Rangel et al., 1986; ³Modi and Tesh, 1983; ⁴Teixeira et al., 2002.

longipalpis (Table 4). Thus these factors had to be estimated from available published data in combination with GIS query analysis to define the thermal-hydrological regime associated with presence or absence of VL. The GDD base temperature, below which there is no development, was estimated to be 16°C through review of literature reports and records of GIS condition queries at positive sites. No cases were found where the recorded average annual monthly water budget values were below 0.7 or where the average annual mean temperature exceeded 28°C.

The GDD-WB model indicates that the highest prevalence of the disease corresponds to areas with the highest predicted number of potential generations per year (Fig. 3). The GDD-WB predictive model was based on data acquired from GIS analysis and literature review of parameters of the reproduction in the laboratory of the VL vector *Lu. longipalpis*. Further studies on the temperature requirements of *Lu. longi-*

palpis in natural environments and in the laboratory will be important to better define the distribution of the vector and potential risk of VL transmission based on the concept of GDD-WB potential generations per year.

Ecological zones - Comparison of the GARP ecological risk model prediction and the GDD-WB model prediction of the distribution of VL in the state of Bahia to the ecological zone map of Bahia, suggests that the Caatinga ecological zone, a zone characterized by a hot and semi-arid environment, is the highest risk area and that the coastal region and the Cerrado ecological regions are lower risk areas (Fig. 4). The Northeast Brazil Dry Forest and the Bahia Interior Forest Zones may be intermediate transitional zones. Literature supports our findings that *Lu. longipalpis* is found most often in brush land and not in open savannah, grasslands or heavily forested areas with broadleaf forest or on the Atlantic border with tropical littoral forests (Sherlock, 1996).

Both the final GARP model and the GDD-WB generations per year model predicted high or moderate risk in major areas in the ecological region known as the Caatinga. The western part of the Bahian Cerrado had low number of generations. The Bahia coastal forest that is part of the Bahian Mata Atlantica shows a lower number of generations per year indicating low development of the vector and consequently low risk for VL. Sherlock (1996), describes the distribution of the disease in the state of Bahia is limited to the central Plateau, where the vegetation is predominantly xerophilous. Although the Caatinga has been predicted as the most suitable ecological region for the diseases to occur, Sherlock (1996) concludes that the major modification of sylvatic ecosystems caused by deforestation, making different animals (wild and domestic) encroach on places near human dwellings, changes the distribution of the disease as well as the distribution of the vector. Moreover, Equatorial-semiarid areas of poor soils would predispose children to malnutrition, increasing the risk of becoming infected with VL (Thompson et al., 2002).

In the studies reported here, ecological niche models were developed within a GIS using GARP analysis and the GDD-WB generations-per-year concept to predict the distribution and potential risk of VL in the state of Bahia, Brazil based on thermal-hydrological climate regime. *Lu. longipalpis-L. chagasi* system in Bahia, and disease prevalence records were shown to be related to major ecological zone maps. At macrohabitat scales, both the GARP model and the GDD-WB generations-per-year model thus showed that climate and thermal-hydrological regime are key determinants of VL disease risk potential in Bahia. The different prevalence classifications (high, middle and low) coincided with the predicted areas of the GDD-WB model and the GARP ecological niche model. Additional studies to validate and apply GARP and GDD-WB prediction models of VL risk in Bahia and elsewhere in Brazil as control program management tools are warranted. Further development of biology-based disease risk models for VL will be conditional on availability more detailed data from laboratory and field studies on the biological requirements and

distribution at both macrohabitat and microhabitat scales of the *Lu. longipalpis-L. chagasi* vector-parasite system.

Acknowledgements

Fundação Nacional de Saúde, Estado da Bahia (FUNASA) for the eleven year VL data of the state of Bahia. Jorge Velásquez for the assistance with the GARP procedures. Dr. Emanuele Brianti and Dulce Maria Bustamante for collaboration during the study.

References

- Andrewartha HG, Birch LC, 1954. The distribution and abundance of animals. Chicago. University of Chicago Press, 728 pp.
- Ahasan HA, Chowdhury MA, Azhar MA, 1996. Deaths in visceral leishmaniasis (kala-azar) during treatment. *Med J Malaysia* 51, 29-32.
- Arias J R, Monteiro PS, Zicker F, 1996. The Reemergence of Visceral Leishmaniasis in Brazil. *Emerg Infect Dis* 2, 145-146.
- Bavia ME, Malone JB, Hale L, Dantas A, Marroni L, Reis R, 2001. Use of thermal and vegetation index data from earth observing satellites to evaluate the risk of schistosomiasis in Bahia, Brazil. *Acta Trop* 79, 79-85.
- Costa J, Peterson T, Beard CB, 2002. Ecologic niche modeling and differentiation of populations of *Triatoma brasiliensis* Neiva, 1991, the most important Chagas' disease vector in Northeastern Brazil (*Hemiptera, Reduviidae, Triatominae*). *Am J Trop Med Hyg* 67, 516-520.
- Grimaldi G, Tesh RB, McMahon-Pratt D, 1989. A review of the geographical distribution and epidemiology of Leishmaniasis in the New World. *Am J Trop Med Hyg* 40, 687-725.
- Hay SI, Rogers DJ, Shanks GD, Myers MF, Snow RW, 2001. Malaria early warning in Kenya. *Parasitol Today* 17, 95-98.
- Killick-Kendrick R, Leaney AJ, Ready PD, 1977. The establishment, maintenance and productivity of a laboratory colony of *Lutzomyia longipalpis* (Diptera: Psychodidae). *J Med Entomol* 13, 429-440.
- Killick-Kendrick R, Molyneux DH, Rioux JA, Lanotte G 1980. Possible origins of *Leishmania chagasi*. *Ann Trop*

- Med Parasitol 74, 563-565.
- Kitron U, Clennon JA, Cecere MC, Gurtler RE, King CH, Vazquez-Prokopec G, 2006. Upscale and downscale: applications of fine scale remotely sensed data to Chagas disease in Argentina and schistosomiasis in Kenya. *Geospatial Health*, present issue.
- Lutz A, Neiva A, 1912. Contribuição para o conhecimento das espécies do gênero *Phlebotomus* existentes no Brasil. *Mem Inst Oswaldo Cruz* 4, 84-95.
- Malone JB, 2005. Biology-based mapping of vector – borne parasites by Geographic information systems and remote sensing. *Parassitologia* 47, 27-50.
- Malone JB, Bergquist NR, Huh OK, Bavia ME, Bernardi M, El Bahy MM, Fuentes MV, Kristensen TK, McCarroll JC, Yilma JM, Zhou XN, 2001. A global network for the control of snail-borne disease using satellite surveillance and geographic information systems. *Acta Trop* 79, 7-12.
- Malone JB, Poggi E, Iguarada F, Sintasath D, Ghebremeesk T, Corbett J, Mccarroll J, Chinnici P, Shililiu J, McNally K, Downer R, Perich M, Ford R, 2003. Malaria environmental risk assesment in Eritrea. *Proceedings IGARRS, Toulouse* 2, 1000-1003.
- Modi GB, Tesh RB, 1983. A simple technique for mass rearing *Lutzomyia longipalpis* and *Phlebotomus papatasi* (Diptera: Psychodidae) in the laboratory. *J Med Entomol* 20, 568-569.
- Momen H, Grimaldi Jr G, Dean LM, 1987. *Leishmania infantum* the aetiological agent of American visceral leishmaniasis? *Mem Inst Oswaldo Cruz* 82, 447-448.
- Pavlovsky EN, 1966. Natural nidality of transmissible diseases. Urbana: University of Illinois Press, 261pp.
- Peterson AT, Ball LG, Cohoon KP, 2002a. Predicting distributions of Mexican birds using ecological niche modeling methods. *Ibis* 144, E27-E32.
- Peterson AT, Bauer JT, Mills JN, 2004a. Ecologic and geographic distribution of filovirus disease. *Emerg Infect Dis* 10, 40-46.
- Peterson AT, Pereira RS, Camargo Neves VF, 2004b. Using epidemiological survey data to infer geographic distributions of leishmaniasis vector species. *Rev Soc Bras Med Trop* 37, 10-14.
- Peterson AT, Shaw J, 2003. *Lutzomyia* vectors for cutaneous leishmaniasis in Southern Brazil: ecological niche models, predicted geographic distributions, and climate change effects. *Int J Parasitol* 33, 919-931.
- Peterson AT, Stockwell DRB, Kluza DA, 2002b. Distributional prediction based on ecological niche modeling of primary occurrence data. In: Scott JM, editor. *Predicting species occurrences: issues of scale and accuracy*. Washington: Island Press, 617-623.
- Peterson AT, Vieglais DA, 2001. Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. *Bioscience*, 51, 363-371.
- Peterson AT, Vieglais D, Andreassen J, 2003. Migratory birds modeled as critical transport agents for West Nile Virus in North America. *Vector Borne Zoonotic Dis* 3, 27-37.
- Rangel EF, Lainson R, Souza AA, Ready PD, Azevedo CR, 1996. Variation between geographical populations of *Lutzomyia (Nyssomyia) whitmani* (Antunes & Coutinho, 1939) *sensu lato* (Diptera: Psychodidae, Phlebotominae). *Mem Inst Oswaldo Cruz* 91, 43-50.
- Sherlock IA, 1996. Ecological interactions of visceral leishmaniasis in the state of Bahia, Brazil. *Mem Inst Oswaldo Cruz* 91, 671-683.
- Stockwell DRB, 1999. Genetic algorithms II. In: AH fielding machine learning methods for ecological applications. Kluwer Academic publishers, Boston, 123-144.
- Stockwell DRB, Peterson AT, 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148, 1-13.
- Teixeira M, Santos RJ, Sampaio R, Pontes-de-Carvalho L, Dos-Santos W, 2002. A simple and reproducible method to obtain large numbers of axenic amastigotes of different *Leishmania* species. *Parasitol Res* 18, 963-968.
- Thompson AR, Lima De Oliveira J W, Maguire JH, Braud DH, Scholl DT, 2002. Climatic and demographic determinants of American Visceral Leishmaniasis in northeastern Brazil using remote sensing technology for environmental categorization of rain and region influences on Leishmaniasis. *Am J Trop Med Hyg* 67, 648-655.
- World Health Organization, 1995. Report of the second WHO meeting on emerging Infectious diseases. Document WHO/CDS/BVI95.2. Geneva, Switzerland: World Health Organization January 1995.