PREDICTING TRIATOMA DIMIDIATA ABUNDANCE AND INFECTION RATE: A RISK MAP FOR NATURAL TRANSMISSION OF CHAGAS DISEASE IN THE YUCATÁN PENINSULA OF MEXICO

ERIC DUMONTEIL AND SÉBASTIEN GOURBIÈRE

Laboratorio de Parasitología, Centro de Investigaciones Regionales Dr. Hideyo Noguchi, Universidad Autónoma de Yucatán, Mérida, Mexico; Laboratoire de Théorie des Systèmes, Université de Perpignan, Perpignan, France

Abstract. Chagas disease, a major public health problem in Latin America, is caused by the protozoan parasite Trypanosoma cruzi and transmitted by hematophagous insects from the Triatominae subfamily. Control of this disease is based on domestic vector control with insecticides and improvements in housing. As with other vector-borne diseases, the identification of areas of high risk of disease transmission is a major prerequisite for the planning and implementation of cost-effective control programs. In this study, we explored the relationship between Triatoma dimidiata geographic distribution and bioclimatic factors in the Yucatán peninsula in Mexico, using geographic information systems, and developed predictive models of T. dimidiata domestic abundance and of its infection rates by T. cruzi. These predictions were then used to build the first natural transmission risk map for Chagas disease in the Yucatán peninsula, a tool that should prove very valuable for the implementation of effective vector control programs in the region.

INTRODUCTION

Chagas disease, or American trypanosomiasis, is caused by the protozoan parasite Trypanosoma cruzi, which is transmitted by hematophagous insects from the Triatominae subfamily. This disease represents a major public health problem in Latin America, where the World Health Organization estimates that approximately 100 million people are at risk of infection, and between 16 million and 18 million are infected. In Mexico, Chagas disease is endemic in various regions, including the Yucatán peninsula, where seroprevalence rates of 11–18% and 5.6% have been reported in the general population and blood donors, respectively.

Due to the limited efficacy of drug treatments and the lack of a vaccine, control of Chagas disease remains based on domestic vector control with insecticides and improvements in housing. Such prevention programs in South America have been enormously successful at reducing house infestation by triatomines, and consequently, natural transmission of Chagas disease to humans. Thus, it is expected that natural transmission will be interrupted in the coming years in the southern cone countries (Argentina, Brazil, Bolivia, and Chile), and comparable programs have been initiated in Central America.

In the Yucatán peninsula, Triatoma dimidiata is the main Chagas disease vector. In a previous work, we reported the first data on T. dimidiata domestic distribution in the Yucatán peninsula, and its seasonal variations. We showed that domestic T. dimidiata was more abundant and its T. cruzi infection rate was higher during the hot and dry months of April–June, suggesting a higher risk of transmission of T. cruzi to humans during this period. In addition, this Chagas disease vector was more abundant in the northern part of the peninsula, possibly in relation with environmental factors. In fact, several other studies have shown that bioclimatic factors can greatly influence triatomine geographic distribution. As with other vector-borne diseases, the knowledge of vector geographic distribution, vector competence, and overall of areas of high risk of disease transmission can allow the development of sensitive tools for disease prediction, and the optimization of planning and implementation of effective control programs.

In recent years, analytical tools such as geographic information systems (GIS), remote sensing, and niche modeling have greatly enhanced our ability to evaluate the relationships between environmental factors and vector ecology or disease transmission. Indeed, such tools have been extensively used to define the geographic distribution of malaria, African trypanosomiasis, schistosomiasis, leishmaniasis, and Lyme disease among others, at a scale ranging from continental to county level. Most of these studies focused on predicting the presence/absence of the vector, but some further predict vector density, vector infection rate, or disease risk. These predictive maps have proven very useful for disease monitoring and control, and are increasingly used in public health.

This type of analysis is only beginning to be applied to Chagas disease, where control programs have mostly relied on labor-intensive field monitoring of entomologic and epidemiologic indicators. Remote sensing and GIS have been recently used to successfully predict the presence or absence of Triatoma infestans in South America. This work showed that vegetation index and five climatic factors were sufficient to predict T. infestans distribution at a continental scale, and confirmed that this approach can greatly reduce the requirements for field studies. In another study, ecologic niche modeling was used to predict the geographic distribution of triatomines from the Protracta complex and their potential reservoir hosts in Mexico. This type of modeling was found to be very powerful for the analysis of vector/host relationships, and the identification of potential risk areas. Population distribution and structure of T. brasiliensis was similarly assessed by niche modeling.

In the present work, we explored the relationship between T. dimidiata and bioclimatic factors in the Yucatán peninsula in Mexico, and developed predictive models of T. dimidiata domestic abundance and of its infection rates by T. cruzi. These predictions were then used to build the first natural transmission risk map for Chagas disease in the Yucatán peninsula, a tool that should prove very valuable for the design
and implementation of effective vector control programs in the region.

MATERIALS AND METHODS

Data collection and processing. The Yucatán peninsula is located in southeastern Mexico at latitudes 17°–22°N and longitudes 86°–92°W. It includes the Mexican states of Campeche, Yucatán, and Quintana Roo. Entomologic field data from 23 villages in the Yucatán peninsula were obtained from a previous study. This included *T. dimidiata* abundance and infection rates by *T. cruzi* for each village. In addition, *T. dimidiata* was collected in 11 new villages during 2001 and analyzed for *T. cruzi* infection, following the same protocols as in our previous work. We thus used entomologic data from a total of 34 collection sites (Figure 1).

Annual mean temperature, minimum and maximum temperature, precipitation rate, relative humidity, precipitable water, pressure, and wind speed maps for the corresponding years of field collection were obtained from the National Oceanic and Atmospheric Administration–Cooperative Institute for Research in Environmental Sciences Climate Diagnostics Center (Boulder, CO) (http://www.cdc.noaa.gov/), which provides digital maps of modeled climate variables based on a combination of field, aircraft, and satellite observations. Annual cumulative precipitation, mean, minimum and maximum temperatures, and evaporation from a total of 140 geographic locations were obtained from the Comisión Nacional del Agua (Mérida, Yucatán, Mexico). Demographic (village total population, village number of houses) and administrative data were obtained from the Instituto Nacional de Estadística, Geografía e Informática (Colony San Juan Mixcoac, Mexico) 2000 national population census, as well as the georeference of each locality where insect collection had been performed. For vegetation type, we used a satellite image of March 2000 from the SeaWiFS sensor (National Aeronautics and Space Administration, Washington, DC) and performed a supervised classification into four main vegetation classes, corresponding to mangrove, agriculture and pastures, low secondary forest, and high primary forest using Multispec software (Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN). Kappa was greater than 0.9 for each vegetation type indicating unambiguous class identification. In addition, ground truthing of each collection site showed satisfactory agreement with the satellite image classification analysis.

Geographic location of seropositive patients reported during 2003 through the blood bank screening program were generously provided by Dr. R. Ojeda-Baranda, (Secretaria de Salud, Mérida, Yucatán, Mexico), and the geographic location of the residence of chronic chagasic cardiomyopathy patients were generously provided by Dr. V. Suárez-Solis (Universidad Autónoma de Yucatán, Mérida, Mexico).

All the data were imported into a raster-based GIS (MacGIS 3.0: University of Oregon, Eugene, OR) to build a comprehensive database of entomologic and bioclimatic factors for the Yucatán peninsula. Point data from field meteorologic stations were interpolated to provide continuous data for the whole study area. All maps layers were set at a pixel resolution of 13 × 13 km. Bioclimatic data matching each 34 *T. dimidiata* collection sites were then extracted from the data base for statistical analysis and modeling (data available on request).

Modeling. We first divided our data set into two subsets, one for modeling, and one for testing the models. For this, data points were paired to ensure that a similar range of vector abundance and infection rates, as well as data from both years of field collections would be equally represented in both subsets. Modeling Chagas disease transmission risk was then performed in two steps. First, we modeled *T. dimidiata* domestic abundance as a function of bioclimatic factors. The relationship between bug abundance and bioclimatic variables was obtained through an analysis of covariance (ANCOVA) logistic regression with a Poisson error of the probability of occurrence with GLIM 3.77 software (Numerical Algorithm Group, Ltd., Oxford, United Kingdom), using the first data subset (17 data points). The model was developed in a step-wise manner by incorporating one variable at a time into the regression, and only retaining at each step the variable that best improved the fit of the model as determined by its deviance. Variables that did not significantly improve the fit were excluded from the model. Vector abundance was thus fitted according to the equation $A = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n)$, where $A$ is the abundance, $\beta_n$ is the respective coefficient of the regression, and $X_n$ is the respective variables retained in the model. The relationship between fitted and observed bug abundance was evaluated by correlation and linear regression analysis. To further evaluate the predictive value of our model, we applied it to the second subset of data (17 data points), and similarly compared the predicted and observed abundance of bugs by correlation and regression analysis.

Second, we modeled the number of *T. cruzi* infected bugs through an ANCOVA logistic regression with a binomial error of the probability of infection using the first data subset. Because we had data on *T. cruzi* infection for only 27 of the 34 field sites, we used 14 data points from a first data subset to build the model. The number of infected bugs was thus fitted according to the equation: $N_i = N_a(Y/1 + Y)$, where $N_i$ is the number of infected bugs, $N_a$ is the number of bugs...
analyzed, and \( Y = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n) \), where \( \beta_n \) is the respective coefficient of the regression and \( X_n \) are the respective variables retained in the model. As before, we evaluated the relationship between fitted and observed infection rates by correlation and regression analysis. Again, the predictive value of model was further analyzed using our second data subset (13 data points) to compare the predicted and observed number of \( T. cruzi \)-infected bugs.

Bioclimatic data for the entire peninsula were then imported from the GIS database into Excel® (Microsoft, Redmond, WA) spreadsheets, and the logistic regression models for abundance and infection rates were applied to the whole study area. Maps of the predicted \( T. dimidiata \) abundance and its predicted infection rates where then generated with the GIS. Finally, we defined the risk of \( T. cruzi \) natural transmission as being proportional to the product of \( T. dimidiata \) abundance and its \( T. cruzi \) infection rate (i.e., proportional to the abundance of infected bugs) for the generation of the transmission risk map. The risk was divided into three levels (low, medium, and high) corresponding to the lower, middle, and upper third of infected bug abundance, respectively. Accuracy of the risk map was further evaluated by comparing the location of known human seropositivity and chronic chagasic cardiomyopathy cases with the estimated natural transmission risk level.

RESULTS

**Modeling \( T. dimidiata \) abundance.** Using the first part of our dataset for building the model, only five of all the bioclimatic and demographic variables studied were retained during the stepwise elaboration of the logistic regression model. These were wind speed, precipitation rate, vegetation type, relative humidity, and maximum temperature, which accounted for a reduction in the deviance from 744.5 to 58.1 (degrees of freedom [df] = 10). There was a strong correlation between observed and fitted bug abundance for the 17 sites used to build the model, and the slope of the regression was close to 1 (\( r^2 = 0.83, P < 0.001 \); slope = 1.06 ± 0.12; Figure 2A). Conversely, residuals of the logistic regression showed no correlation with the observed abundance or any of the dependent variables nor the predicted abundance. We then tested the predictive value of the model using the second half of our dataset. Observed and predicted abundance were again strongly correlated, and the slope of the regression was also close to 1 (\( r^2 = 0.45, P = 0.003 \); slope = 0.83 ± 0.24; Figure 2A). Overall, the combined data sets showed a very good agreement between observed and predicted abundance of \( T. dimidiata \) (predicted abundance = 7.15 + 0.82 × observed abundance; \( r^2 = 0.59, P < 0.001 \)).

**Modeling \( T. dimidiata \) infection by \( T. cruzi \).** Of the 27 sites for which we had \( T. cruzi \) infection data, we used 14 sites to fit the number of infected bugs with the different bioclimatic and demographic variables. Relative humidity, vegetation type, cumulative precipitation, and minimum temperature and maximum temperature range accounted for a reduction in deviance from 55.1 to 35.9 (df = 7). There was a good correlation between observed and fitted infected bug number for the 14 sites used to build the model (\( r^2 = 0.68, P < 0.001 \); slope = 0.81 ± 0.16; Figure 2B). Again, the residuals showed no correlation with infection rates or any of the dependent variables. When we tested the predictive value of the model using the remaining 13 sites of our dataset, some correlation was observed between observed and predicted infected bug numbers (\( r^2 = 0.26, P = 0.07 \); slope = 0.73 ± 0.36; Figure 2B). Overall, the combined data sets showed a very good agreement between observed and predicted infection numbers (Predicted infection = 2.54 + 0.73 × observed infection, \( r^2 = 0.45, P < 0.001 \)).

**Natural transmission risk map.** We then applied both models to the entire study area to obtain maps of \( T. dimidiata \) domestic abundance and \( T. cruzi \) infection in the Yucatán peninsula (Figure 3A and B, respectively). As expected from our previous work, these maps showed a very heterogeneous geographic distribution of \( T. dimidiata \), which was predicted to be more abundant in the northwestern part of the peninsula, and rare in the southern part. Similarly, infection rates varied importantly, with infection predicted to be lower in the northwestern region and higher in the Southeast. We then defined natural transmission risk as directly proportional to the abundance of infected bugs, which was generated by combining maps of \( T. dimidiata \) abundance and its infection rates by \( T. cruzi \) (Figure 4). Natural transmission risk appeared to be very high in most of the state of Yucatán, as well as in the northern parts of the states of Campeche and Quintana Roo. Conversely, Southern Campeche and Quintana Roo appeared to have a very low natural transmission risk. To further evaluate the accuracy of this risk map, we compared the location of \( T. cruzi \)-seropositive and chronic chagasic patients with the predicted transmission risk. As shown in Figures 4 and 5, most seropositive patients came from villages with a high predicted transmission risk, and only seven patients were from a low predicted transmission risk area. Similarly, nine of
10 chronic chagasic patients came from high risk areas, one of 10 from a medium risk area, and none from low risk areas, confirming the usefulness of the developed risk map.

**DISCUSSION**

Like other vector-borne diseases, Chagas disease is strongly influenced by bioclimatic variables, and numerous studies have investigated the effect of climatic factors on several aspects of triatomine life cycle. However, very few studies have attempted to establish integrated relationships between environmental variables and Chagas disease vectors to predict the entomologic risk of transmission. Here we present the first natural transmission risk map of Chagas disease in the Yucatán peninsula of Mexico. This map should be of major usefulness for further epidemiologic and entomological research because it would allow focusing these studies in high risk areas. Most importantly, it could also serve as a decision-making tool for the identification of priority areas for vector control and epidemiologic surveillance programs in the region, allowing one to optimize the cost-effectiveness of these public health programs.

Our models allow the prediction of *T. dimidiata* abundance in the domiciles and infection by *T. cruzi*, and we found an overall good agreement between the predicted and observed data subset used for either building the models or testing them. Prediction of vector abundance was very accurate, with more than 90% of the deviance explained by the model. Of the variables retained in this model, the influence of temperature and precipitation/humidity on triatomine biology and life cycle have been well established for a number of species, and *T. dimidiata* appeared more abundant in warmer and drier climate in the Yucatán peninsula. This seems to contrast with *T. dimidiata* populations from Oaxaca, Mexico, which are restricted to milder temperatures. This may suggest important differences in the ecologic niches between these distinct populations, confirming the major genetic differences observed between them. Indeed, *T. dimidiata* populations from Yucatán have been suggested to be involved in a strong differentiation process, which may result in major biologic and ecologic differences with other populations from Mexico and Central America. Interestingly, we found that wind speed was also influencing *T. dimidiata* domestic abundance, which may be interpreted as reflecting dispersal events. In agreement with this interpretation, we suggested in a previous work that house infestation by *T. dimidiata* in the region is mostly due to the seasonal dispersal of sylvatic/peridomestic
adult bugs, again contrasting with other T. dimidiata populations, which show stronger domiciliation. Inclusion of seasonal variations in future risk mapping studies may thus help identify potential bioclimatic factors related to the seasonal infestation of houses in the Yucatán peninsula.

The predictive accuracy of the model for T. dimidiata infection by T. cruzi was lower, with only 35% of the deviance explained by the model. Nonetheless, there was a sufficient agreement between observed and predicted infected T. dimidiata as assessed by regression analysis. The lower fit of the model may be due in part to the lower number of data points available, but also, and most likely, because major factors influencing T. dimidiata infection by T. cruzi were not included in our study. Variables related to temperature and humidity have previously been shown to directly modulate T. cruzi development in T. infestans, and were indeed retained in the model. Higher infection was associated with lower temperature and higher humidity/precipitation. However, it is likely that infection is also critically dependent on several additional factors such as the host species on which bugs are feeding and their respective infection rates by T. cruzi, but such data are not yet available. This limited knowledge represents a major limitation for the development of more reliable vector infection maps.

Vegetation type also appeared as an important predictor of both abundance and infection of T. dimidiata. It is indeed a general proxy for environmental characterization because it integrates a wide number of climatic, geologic, zoologic, and anthropologic factors. It is indeed one of the variable most often associated with vector geographic distribution. This raises the possibility that the use of more detailed vegetation data such as remotely sensed NDVI data may allow the elaboration of simpler models, but with potentially increased spatial resolution. Higher bug abundance appeared associated with perturbed vegetation (agriculture and pasture), suggesting that deforestation and habitat degradation are important factors contributing to the domiciliation of T. dimidiata. Conversely, higher bug infection coincided with high forests, in agreement with the zoonotic origin of T. cruzi.

Our natural transmission risk map also appeared of high predictive value for human cases. Indeed, the large majority of patients detected in recent years were from high risk areas, while only seven were from predicted low risk areas. The abundance of infected bugs is the most straightforward indicator of natural transmission risk, but additional factors, such as the presence and nature of domestic animals may modulate transmission risk to humans. Our map predicts that approximately 1.6–1.7 million people are exposed to a medium-high risk of transmission of T. cruzi in the Yucatán peninsula, with more than 90% of these localized in the Yucatán state, while 1.5–1.6 million people live in low risk areas. This risk map represent a first decision-making tool for the identification of priority vector control areas by public health authorities, and further developments of this type of analysis should allow the elaboration of improved maps of higher spatial resolution and predictive accuracy. A generalized use of such tools for Chagas disease vector control and surveillance would allow their optimization and a better distribution of scarce resources.

Acknowledgments: We thank R. Ojeda-Baranda and V. Suarez-Solis for providing the geographic location of clinical cases, and J. Brownstein and J.A. Farfán-Ale for helpful discussions and comments on an earlier version of this work.

Financial support: This work was supported by grant no. 20020404 from SISIERRA/CONACYT to Eric Dumonteil.

Authors’ addresses: Eric Dumonteil, Laboratorio de Parasitología, Centro de Investigaciones Regionales Dr. Hideyo Noguchi, Universidad Autónoma de Yucatán, Ave. Itzaes #490 x 59, 97000 Mérida, Yucatán, México, Telephone: 52-999-924-5910, Fax: 52-999-923-6120, E-mail: oliver@tunku.uady.mx. Sébastien Gourbière, Laboratoire de Théorie des Systèmes, Université de Perpignan, 52, Avenue de Vileneuve, 66800 Perpignan Cedex, France, Telephone and Fax: 33-4-68-66-17-60.

Reprint requests: Eric Dumonteil, Laboratorio de Parasitología, Centro de Investigaciones Regionales Dr. Hideyo Noguchi, Universidad Autónoma de Yucatán, Ave. Itzaes #490 x 59, 97000, Mérida, Yucatán, México.

REFERENCES


Received October 29, 2003. Accepted for publication December 26, 2003.