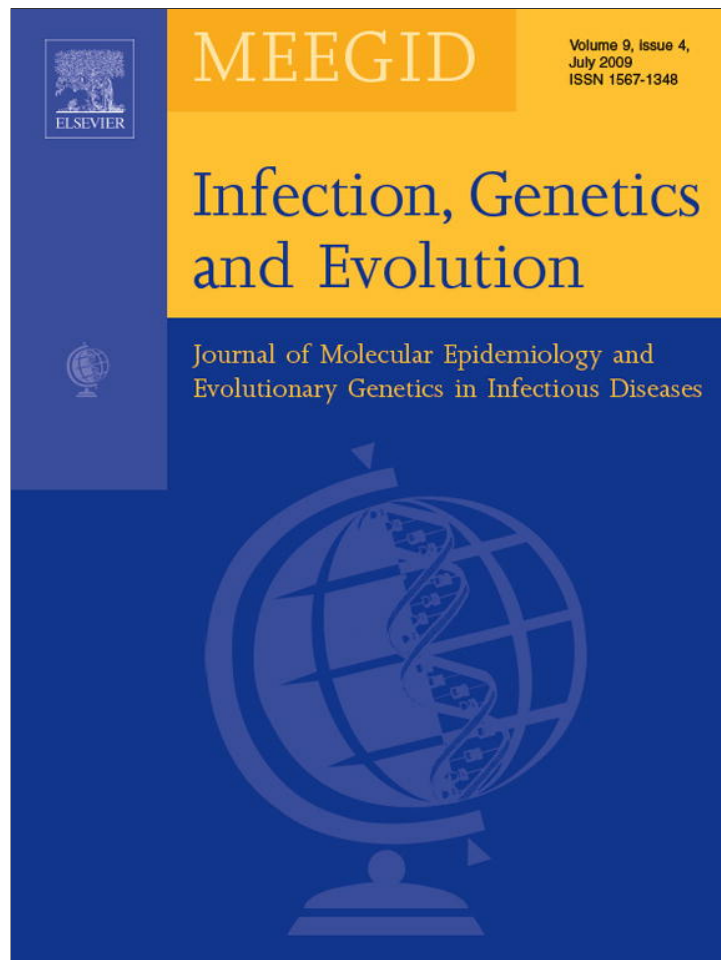


Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

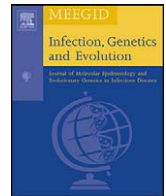
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

## Infection, Genetics and Evolution

journal homepage: [www.elsevier.com/locate/meegid](http://www.elsevier.com/locate/meegid)

## Development of a geographical distribution model of *Rhodnius pallescens* Barber, 1932 using environmental data recorded by remote sensing

S. Arboleda<sup>a</sup>, D.E. Gorla<sup>b</sup>, X. Porcasi<sup>b,c</sup>, A. Saldaña<sup>d</sup>, J. Calzada<sup>d</sup>, N. Jaramillo-O<sup>a,\*</sup>

<sup>a</sup> Grupo Chagas, Instituto de Biología, Universidad de Antioquia, Sede de Investigaciones Universitarias – SIU, Calle 62 # 52–59 Laboratory 620, P.O. Box: 1226, Medellín, Colombia

<sup>b</sup> Centro Regional de Investigación Científica y Transferencia Tecnológica (CRILAR), Anillaco, La Rioja, Argentina

<sup>c</sup> Instituto Gulich, Comisión Nacional de Actividades Espaciales (CONAE), Córdoba, Argentina

<sup>d</sup> Instituto Conmemorativo Gorgas de Estudios de la Salud, Ciudad de Panamá, Panamá

## ARTICLE INFO

## Article history:

Received 24 January 2008

Received in revised form 4 December 2008

Accepted 9 December 2008

Available online 24 December 2008

## Keywords:

Remote sensing

Ecology

GIS

Triatominae

*Rhodnius pallescens*

Chagas disease

## ABSTRACT

*Rhodnius pallescens*, main vector of *Trypanosoma cruzi* in Panama and secondary vector in Colombia, Costa Rica and Nicaragua, represents an important epidemiological risk in those countries. It occupies sylvatic ecotopes, and because of this its distribution and abundance could be conditioned by environmental factors. In this work, we integrated environmental variables recorded by remote sensing and data of *R. pallescens* presence in the countries mentioned above in order to know the environmental variables with better capacity to describe the insects' distribution, which will help to entomological surveillance and control. Standard discriminant analysis (DA) was used to determine if there is a significant difference in the environmental variation between the presence and the absence sites of *R. pallescens*. Forward stepwise discriminant analysis (fDA) was used to determine the environmental variables which better discriminated between presence and absence sites, and to construct a predictive map of geographical distribution. Univariate analyses were used to determine the discriminatory power of individual variables. The model derived from DA showed 89% of sensitivity and 92.8% of specificity. Multivariate and univariate analyses showed the vapor pressure deficit minimum as the principal variable among the nine most important to describe the distribution of the species, which is in agreement with the *R. pallescens* stenohydric status. Map shows insects' distribution predicted by environmental variables, and moreover includes the distribution of most species belonging to *Rhodnius* genus, except *R. domesticus*, *R. nasutus* and *R. neglectus*. We interpreted these results as a reflection of the common evolution of the most *Rhodnius* species, except for the last ones that probably evolved isolated due to particular environmental conditions. In conclusion, this study showed that a reduced number of environmental variables can predict the distribution of *R. pallescens* and related species. This methodology can be very useful to make critical decisions for vectorial surveillance and control of Chagas disease vectors.

© 2008 Elsevier B.V. All rights reserved.

### 1. Introduction

*Rhodnius pallescens* (Hemiptera: Reduviidae) is a triatominae vector of *Trypanosoma cruzi*, the etiological agent of Chagas disease. It has been reported in Belize, Nicaragua, Costa Rica, Panama and Colombia, where it inhabits sylvatic environment, and often visits human dwellings (Calzada et al., 2006; Moreno-Mejía et al., 1992a,b; Vasquez et al., 2004; Zeledón et al., 2006); although without following domestication processes, according to the criteria of Dujardin et al. (2000). However, the finding of nymphs and adults infected with trypanosomes and fed with

human blood (Calzada et al., 2006; Christensen and de Vasquez, 1981) suggests a vector-human contact that represents an important epidemiological risk. In addition, anthropic forest transformations, generalized in its distribution area, could induce changes in the vector ecology initiating a behavioral gradient leading to synanthropic (Abad-Franch and Monteiro, 2007).

*R. pallescens* contact with humans needs to be monitored and controlled, and thus strategies should be developed to attain this. Among strategies a geographical information system (GIS) could help to identify the environmental variables whose interaction defines, at least in a significant part, the characteristics of the ecosystem associated with the vector presence.

Data for GIS may be obtained from remote sensors (RS) onboard earth observation satellites. RS provides global information of a broad spectrum, spatial and temporal resolution, sometimes freely

\* Corresponding author. Tel.: +57 4 2106521; fax: +57 4 2106565.

E-mail addresses: [njaram@une.net.co](mailto:njaram@une.net.co), [nicolas.jaramillo@siu.udea.edu.co](mailto:nicolas.jaramillo@siu.udea.edu.co) (N. Jaramillo-O).

available in internet databases about climate and ground cover variation (Kitron, 2000; Thomson and Connor, 2000).

These are critical factors to determine the geographical distribution of arthropod vectors and the incidence of the parasites they transmit (Hay, 1997). The study of environmental conditions affecting biological properties of a vector species based on ground collected data is generally expensive and sometimes impossible to carry out, because of the geographical amplitude of the species distribution. But remotely sensed data integrated within a GIS is a low cost and a good approach to study the relationship between environmental variables and the distribution species (Beck et al., 2000).

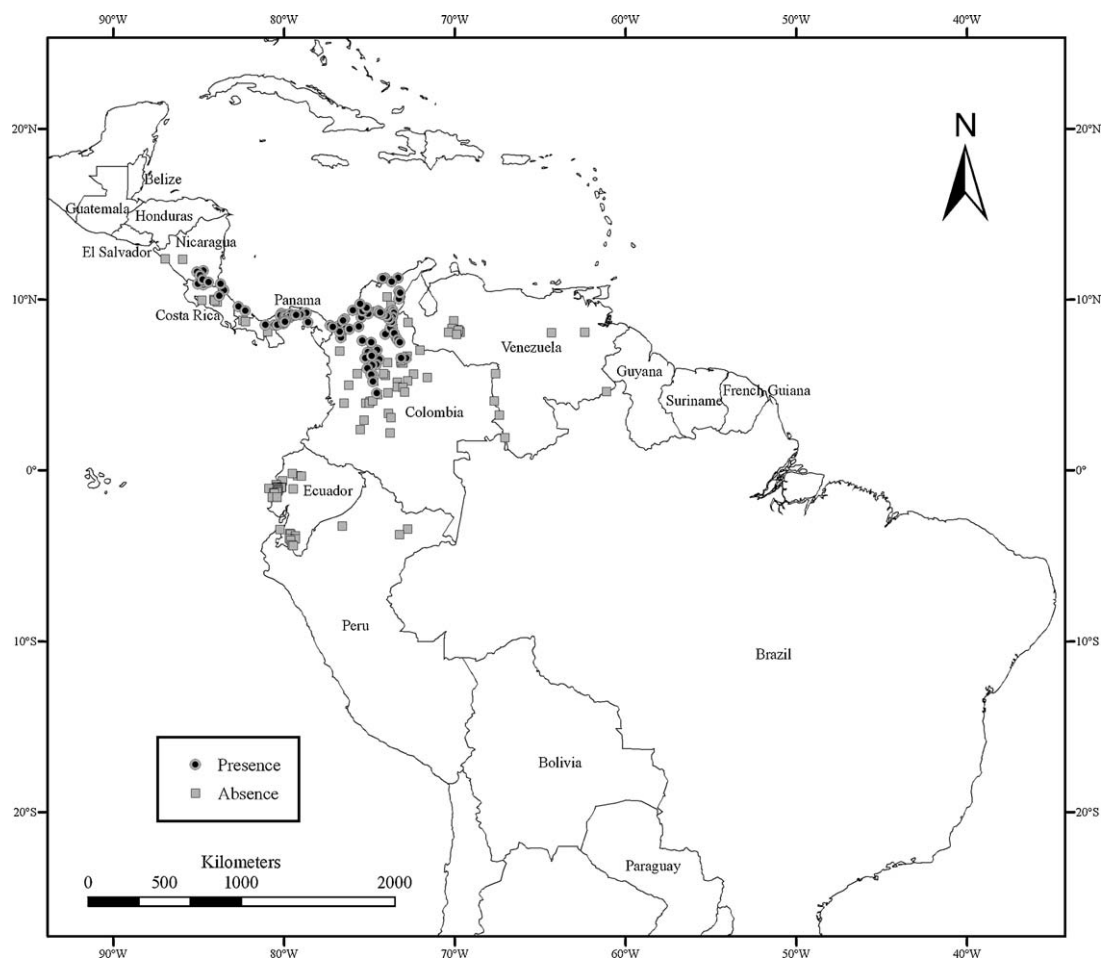
A few works have been done using RS data for study the potential distribution of Triatominae, mainly on *Triatoma infestans* and *T. dimidiata* (Dumonteil and Gourbiere, 2004; Gorla, 2002; Townsend et al., 2002). Dumonteil and Gourbiere (2004) exploring the relationship between *T. dimidiata*'s geographic distribution and bioclimatic factors in the Yucatán peninsula in Mexico developed predictive models of its domestic abundance and of its infection rates by *T. cruzi*. These predictions were used to build the first natural transmission risk map for Chagas disease in the Yucatán peninsula. Gorla (2002) found that only six environmental variables are enough to describe the *T. infestans* distribution in South America. Townsend et al. (2002) applied ecologic niche modeling to identify relationships between *Triatoma* species implicated in the transmission of Chagas disease and their mammal hosts, which allowed stratifying the risk factors according to overlapping of vectors and particular mammal species. In

this work we used a similar ecological approach to determine for the first time in the literature the environmental factors impacting *R. pallenscens* distribution, and then use this information to build a predictive map, which may help to planning vector-control and surveillance strategies.

## 2. Materials and methods

### 2.1. Study area and data set

Data were compiled from the literature, and also collected continuously during the field work by authors since 1989 (Appendix A). The geographical window analyzed includes the inter-tropical zone stretching from southern Mexico to northern Argentina (upper left 25°16'N, 93°46'W; lower right 27°10'S, 34°19'W). They were used geographical coordinates of 101 sites where *R. pallenscens* is present (60 at Colombia, 27 at Panama, 8 at Costa Rica, and 6 at Nicaragua) (Fig. 1). These sites are stretched between 4°31', 11°41'N and 85°5', 72°29'W. Belize was not included because the specific site of collection is not known (Lent and Wygodzinsky, 1979). Absence sites were selected by considering places where triatomines belonging mainly to *Rhodnius* genus were searched, but species different to *R. pallenscens* were found. Thus, we took in account 299 sites reported in the literature for presence of triatomines different to *R. pallenscens* (Guhl et al., 2007; Abad-Franch et al., 2005) or kindly informed by several investigation groups (Biodiversity Laboratory—Medical Entomology, Centro de Pesquisa Leônidas and Maria Deane,



**Fig. 1.** Presence and absence sites of *Rhodnius pallenscens*. Circles represent sites where the species has been collected and squares indicate sites where an intensive search of triatomines has been made but other species different to *R. pallenscens* have been found.

Fiocruz, Manaus, Brazil; Instituto Conmemorativo Gorgas de Estudios de la Salud, Panama, Panama; LIPT, Universidad del Tolima, Ibagué, Colombia; CNRFV, Facultad de Ciencias de la Salud, Universidad de Carabobo, Núcleo Aragua, Maracay, Venezuela). From these 299 absence sites, 100 sites were chosen at random, so that sample size of absence data was balanced respect to presence data. Absence data are located between 5°36'S, 12°23'N and 87°1', 68°55'W (Fig. 1). Latitude and longitude data for each presence and absence site were imported into the integrated GIS and Image Processing Software IDRISI Andes (Eastman, 2006).

## 2.2. Satellite data

Satellite images were obtained from a temporal series of monthly images taken from 1982 to 2000 by the advanced very high-resolution radiometer (AVHRR) sensor onboard the meteorological satellites of the National Oceanic and Atmospheric Administration (NOAA). The sensor records the energy from 1.1 km × 1.1 km of ground cover and the information is stored in a digital format as a numeric data matrix that represents an image. The images correspond to monthly measures of the physical variables calculated with the maximum value composition method to diminish the contamination by clouds (Holben, 1986). Due to the computational process used to obtain the images, the final spatial resolution of each one is 8 km × 8 km per pixel. Images of air temperature (TAIR), land surface temperature (LST), middle infra-red (MIR), vegetation index (NDVI), and vapor pressure deficit (VPD) were calculated by equations described in Hay (2000) and the appropriate combination of the five bands of the AVHRR.

The temporal series were processed using Fourier decomposition (thanks to the TALA Research Group, Oxford University; Rogers and Robinson, 2004) in order to remove data redundancy and to produce a set of uncorrelated variables whilst retaining a description of seasonality. It is commonly applied to AVHRR data collected by more than one year producing orthogonal harmonic series, which exactly describe the temporal series of environmental variables, and are considered efficient descriptors of environment variability (Rogers, 2000; Rogers et al., 1996; Rogers and Robinson, 2004; Rogers and Williams, 1994). That decomposition produced a set of 11 descriptive statistics for each variable: average (A0), minimum (MN), maximum (MX), amplitudes and phases of the annual, bi-annual and tri-annual cycles (A1, A2, A3, P1, P2, P3), the percentage of the total variance of the three first Fourier components (DALL) and the variance of the complete series (VR) (Hay et al., 2006). In summary, the AVHRR images and additionally the land elevation model (DEM) obtained from Shuttle Radar Topography Mission (SRTM) gave 56 independent environmental variables set for each pixel location (11 descriptive statistics for each of the five environmental variables plus DEM). Originally DEM was of 90 m of resolution, but it was converted into images with 8 km of resolution to be able to process them together with images from the AVHRR which had such resolution.

## 2.3. Data analysis

### 2.3.1. Spatial autocorrelation and discriminant analysis (DA)

We calculated the spatial autocorrelation with the Moran index, and found no autocorrelation for presence data, and a partial autocorrelation for absence data, enabling to apply discriminant analysis on the environmental variables data set.

To test if there are significant environmental differences between the presence and absence sites, an standard DA (Fisher, 1936) was conducted using the dichotomic variable presence-absence as the independent one (201 sites) and the set of 56 environmental variables extracted over the same 201 geographical

locations as the dependent variables. Cross-classification analyses were conducted to verify the first DA outcome. For this each site was removed sequentially and used as external data for reclassification in subsequent DA analyses. Additionally, *post hoc* analyses were conducted to evaluate the stability of the model; that is, to test capacity of the new discriminant functions to reclassify when new presence and absence data are introduced. For this, five new matrix data sets were constructed with 75% of the data randomly selected from the original matrix, and new standard DAs were conducted over each matrix. The 25% of the remaining data were introduced as new, “unknown”, data to calculate the capacity of the discriminant functions for reclassify correctly the presence or absence sites.

To determine the variables better discriminate between the presence and the absence sites, a fDA was conducted with the data mentioned above. The model with higher discrimination and less number of variables was selected. The criteria for variables selection used only the statistically significant ones to discriminate between groups; non-significant variables were ignored. It was determined sensitivity and specificity of the model by the reclassification percentage calculated in the fDA. Sensitivity is the probability of classifying correctly a presence site, and the specificity is the probability of classifying correctly an absence site. Predictive accuracy of re-classification was determined by Kappa statistical,  $k$  (Cohen, 1960). Landis and Koch (1977) suggested that a value of  $k < 0.4$  is poor;  $0.4 < k < 0.75$  is good and  $k > 0.75$  is excellent.

A graphical representation of the presence and absence sites and of the selected variables in the fDA was done using a GH-Biplot (Gabriel, 1971), where the vector longitude represents the variability (or importance) of variables, and the relationship site-variable is deduced by projecting the presence or absence sites onto the vectors that represent the higher important variables.

The above analyses were performed using the computational programs JMP 6.0.3 (SAS, 2006) and PADwin (Dujardin, 2006).

### 2.3.2. Construction of a predictive map

The predictive map of *R. pallescens* distribution was constructed using the discriminant functions of the presence and absence sites derived from the fDA. The functions classified the recorded sites, and additional ones with environmental characteristics similar to those of presence or absence sites. The extension of the geographical predicted area is the same described in Section 2.1.

The image derived from the presence function was calculated using the equation:

$$f_{\text{presence}} = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k,$$

where  $k$  is the number of variables that has higher discrimination;  $b_0, b_1, b_2, \dots, b_k$  represent the function coefficients for the presence sites calculated in the fDA, and  $X_1, X_2, \dots, X_k$  represent the value that a variable took in the corresponding pixel.

The image for the absence function was calculated using a similar approach describe for presences sites.

These two images, representing the presence and absence functions resulting from the fDA, were compared pixel by pixel. Based on this comparison, each pixel was classified as absence or present site according to the function which had the higher value. The classification unit is an 8 km × 8 km pixel, and the total classified area is close to 130 millions of square kilometers. All analyses were performed using the IDRISI Andes software (Eastman, 2006).

### 2.3.3. Univariate analysis

*T*-tests were used to test each variable for differences between the presence and the absence sites. Afterward, probability density functions for both classes of sites were estimated for each variable,



which showed significant differences in the previous analyses. For this the total range of conditions of each variable was divided into 10 categories, and presence or absence pixels were assigned to their corresponding category. Thus, the ratio between the number of pixels corresponding to presence (or absence) sites (according to their appropriate categories) and the total number of presence (or absence) pixels were calculated. This produced probability density functions of presence or absence per unit of measurement on the x-axis (Fig. 5). For each environmental variable an optimal threshold for to predict presence or absence sites was calculated as the maximum difference between the accumulated probability density function of both types of sites (Robinson et al., 1997).

In order to evaluate the selected variables in the fDA by their individual capacity of discrimination, sensitivity and specificity of each variable were calculated as TP/(TP + FN) and TN/(TN + FP) respectively, where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives (Fernández and Díaz, 2003). Correct classification for each variable was calculated dividing the sum of sensibility plus specificity by two (Fernández and Díaz, 2003).

### 3. Results

#### 3.1. Discriminant analysis

The standard DA, including all variables, showed that 89% of the presence sites and 92.8% of the absence sites were well classified. Fig. 2 shows a significant discrimination between the presence and the absence sites in the multivariate space represented by a canonical factor (Wilks lambda = 0.3587,  $F_{56,144} = 4.5981$ ,  $p < 0.0001$ ). The cross classification showed 86% of presence sites and 76% of absence sites correctly classified. The *post hoc* prediction for the five matrix data sets, in which the 25% of removed data were re-introduced, showed a good re-classification. The presence data were re-classified correctly between 62% and 85.2%, and the absence data between 68.2% and 81.8% (Table 1).

In the fDA, nine of 56 variables allowed to classify the localities in presence or absence sites (Wilks lambda = 0.551511,  $F_{9,191} = 17.2579$ ,  $p < 0.0001$ ). Three of them correspond to land surface temperature: mean, bi-annual amplitude and variance (LSTA0, LSTA2 and LSTVR); four to air temperature: tri-annual amplitude, percentage of total variance in the annual to tri-annual

**Table 1**

*Post hoc* classification between presence and absence sites.

	Model 1	Model 2	Model 3	Model 4	Model 5
Presence	85.2	74.1	62	68.9	80
Absence	70.8	79.2	81.8	68.2	72

Correct re-classification (%) of presence and absence sites for each of the five matrix data sets as indicated in the text. Models 1 and 5 showed better re-classification for presence sites, and model 3 showed better re-classification for absence sites.

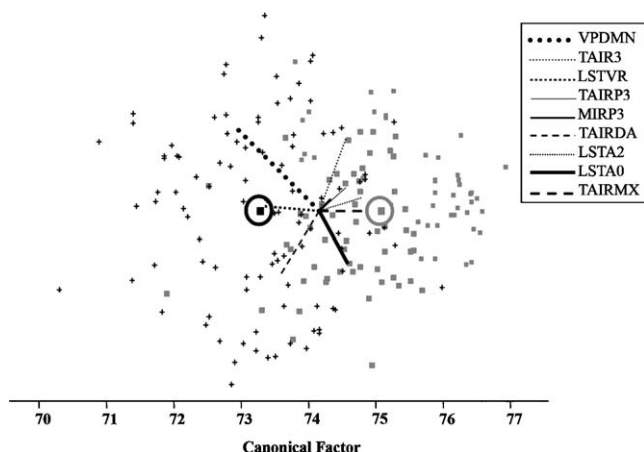
cycles, maximum and tri-annual phase (TAIRA3, TAIRDA, TAIRMX and TAIRP3); one to vapor pressure deficit: minimum (VPDMN); and one to middle infra-red: tri-annual phase (MIRP3). The function estimated for fDA classified correctly the 81.6% of localities: 81.2% of presence sites and 82% of absence sites. Kappa statistical had a value of 0.63 indicating a good concordance between the categorical variable presence-absence. Table 2 shows the standardized coefficients of functions, which indicate the most discriminant variable as the VPDMN. Variables with the lowest discriminatory power are the tri-annual phase of air temperature (TAIRP3), the middle infra-red (MIRP3), and the tri-annual amplitude of air temperature (TAIRA3).

GH-Biplot (Fig. 2) shows that the most discriminant variable (VPDMN) presents the highest variability, and together with LSTVR and TAIRDA shows high values in absence sites and low values in presence sites (projection of point-sites onto vector-variables). The TAIRMX and LSTA2 variables present high values in presence sites and low values in absence sites. These variables are correlated (angle less than 90° between vectors), which indicate redundancy in the information.

It is important to mention that VPDMN, LSTVR, TAIRDA, TAIRMX and LSTA2 allow greater discrimination; while LSTA0, MIRP3, TAIRA3 and TAIRP3, gave the worst discrimination.

#### 3.2. A geographical distribution model of *R. pallescens*

Of the 16,727,806.75 km<sup>2</sup> from the window geographical analyzed, 2,499,358.75 km<sup>2</sup> were classified as presence sites for *R. pallescens*, and 14,228,448 km<sup>2</sup> were classified as absence sites. Fig. 3 shows the potential geographical distribution map that resulted from the re-classification of the presence and absence sites.



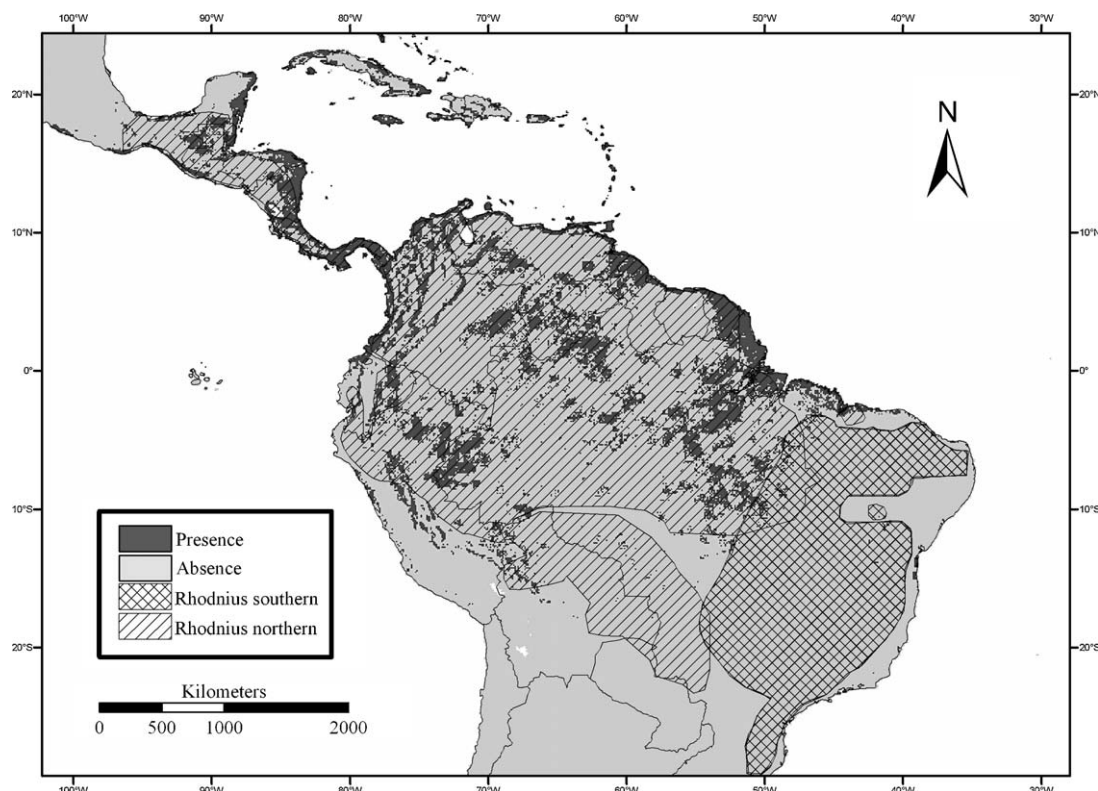
**Fig. 2.** Factorial map derived from standard discriminant analysis. Presence (gray squares) and absence sites (black crosses) are distributed in a one-dimensional canonical space separating both groups. Centroid (multivariate mean) for each group is showed as fill squares; circles surrounding them are determining 95% confidence intervals within which it is possible to find the centroid. Variables in the canonical space are showed as vectors emerging from the “grand media”. Vector longitude indicates variable importance.

**Table 2**

Significant variables selected by the forward stepwise discriminant analysis using 100% of the available data for presence and absence.

Variable	Standardized coefficients	Non-standardized coefficients	Absence function	Presence function
MIRP3	-0.177	-0.019	-3.666	-3.633
LSTVR	0.836	0.086	-2.516	-2.670
TAIRA3	-0.396	-0.067	-0.056	0.064
LSTA0	-0.426	-0.010	3.731	3.750
VPDMN	1.287	0.002	-0.340	-0.344
TAIRP3	-0.384	-0.039	-0.608	-0.538
LSTA2	-0.616	-0.069	-3.715	-3.590
TAIRDA	0.576	0.039	-4.107	-4.177
TAIRMX	-0.680	-0.015	3.353	3.379
Constant		74.164	-10,452.297	-10,585.371

Tri-annual phase of middle infra-red (MIRP3), variance of land surface temperature (LSTVR), tri-annual amplitude of air temperature (TAIRA3), mean of land surface temperature (LSTA0), minimum of vapor pressure deficit (VPDMN), tri-annual phase of air temperature (TAIRP3), bi-annual amplitude of land surface temperature (LSTA2), percentage of total variance in the annual to tri-annual cycles of air temperature (TAIRDA), maximum of air temperature (TAIRMX). For each variable, the standardized and non-standardized coefficients, and the presence and absence functions are showed. The most important variable in the model was the VPDMN.



**Fig. 3.** Predicted distribution map of *Rhodnius pallescens*. Dark gray areas are sites environmentally similar to sites where the species occur. Upward diagonal are indicating the northern radiation that gave rise to *R. robustus*, *R. prolixus*, *R. ecuadoriensis*, *R. colombiensis*, *R. pallescens*, *R. pictipes*, *R. stali*, *R. brethesi*, *R. dalessandroi*, *R. neivai* and *R. paraensis*, and the diagonal crosses are indicating the southern radiation that gave rise to *R. neglectus*, *R. nasutus* and *R. domesticus* (re-drawn after Carcavallo et al., 1999).

### 3.3. Univariate analysis

Individual variables with significant differences between the presence and the absence sites, among the nine selected by the fDA, were LSTVR, TAIRA3, MIRP3, TAIRP3 and VPDMN (Fig. 4). Probability density functions showing the range of values associated with the presence or absence sites are described for these variables (Fig. 5). Some of them showed an unimodal distribution and others a bimodal distribution (Fig. 5, Table 3), indicating these last more than one optimum range for the species occurrence.

In general, the variables in individual way did not showed a good classification capacity. They showed greater sensibility than specificity (Table 3). TAIRDA showed the greatest ability to describe the species distribution with a correct classification of 71.6% (77.2% of sensibility and 66% of specificity). Variables with better sensibility were TAIRP3 and TAIRMX (86.1%), and the variable with better specificity was TAIRA3 (73%).

Although in the fDA the best discriminant variable was VPDMN, in the univariate analysis this was the second variable to reach the best re-classification, but its sensibility and specificity was moderate (76.2% and 62%).

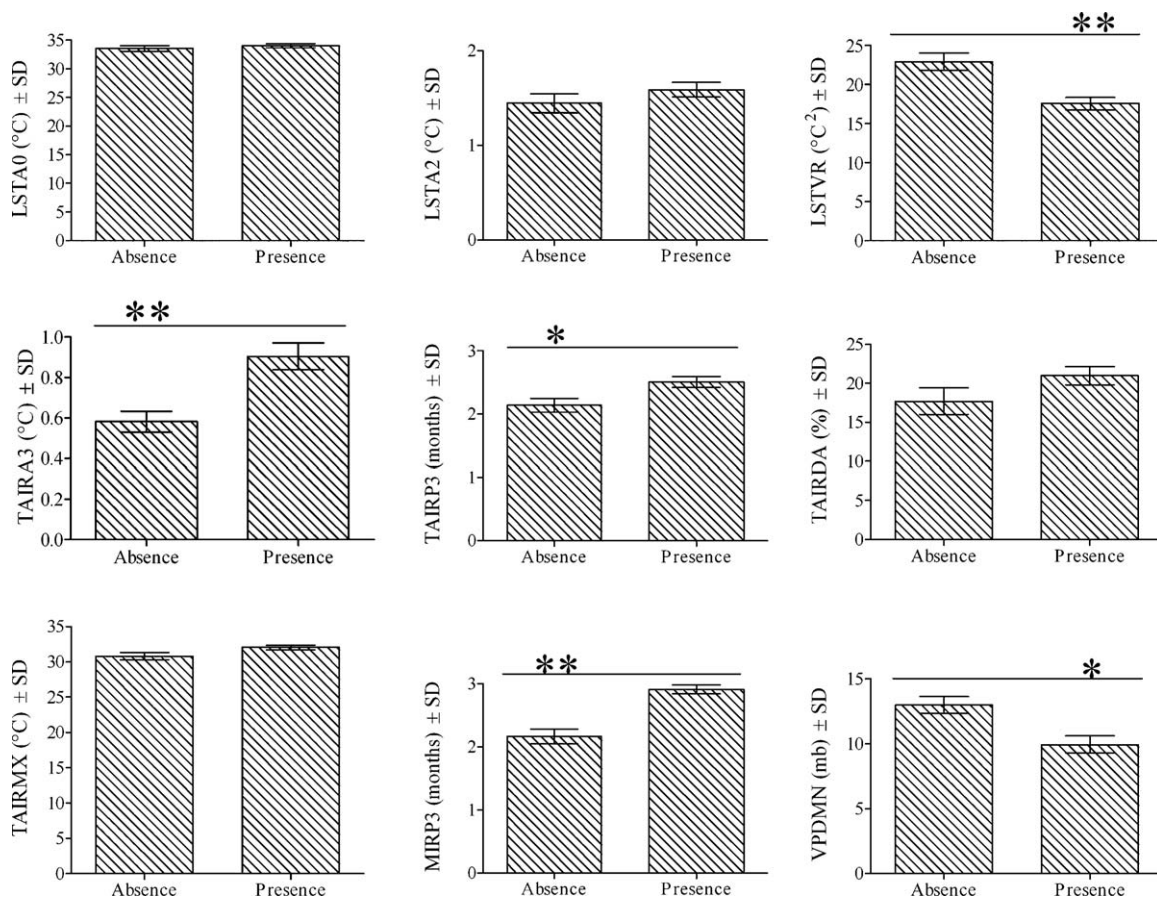
### 4. Discussion

*R. pallescens* inhabits tropical humid, tropical dry and tropical very dry forests (Pizarro and Romaña, 1998). This amplitude of life zones (Holdridge, 1947) suggests that a high number of variables besides the macro-climatic ones could condition the geographical distribution of this vector. *Attalea butyracea* palms, for example, seem to be very important indicators of the presence of this species (Romaña et al., 1999). However, only nine macro-climatic variables associated with land surface temperature, air temperature, medium infra-red, and vapor pressure deficit (LSTA0, LSTA2, LSTVR, TAIRA3, TAIRP3, TAIRDA, TAIRMX, MIRP3, VPDMN) were enough to predict the *R. pallescens* distribution. Multivariate

**Table 3**  
Capacity of individual variables to describe the *Rhodnius pallescens* distribution.

	Correct classification (%)	Sensibility (%)	Specificity (%)	Presence range	Units
MIRP3	66.7	83.2	50	>2.4	months
LSTVR	67.2	74.3	60	<17, 22–27	°C <sup>2</sup>
TAIRA3	62.7	52.5	73	>0.7	°C
LSTA0	67.2	83.2	51	29.2–34.6, 37.7–41.6	°C
VPDMN	69.2	76.2	62	1.2–11.3, 17.5–21.9	millibars
TAIRP3	64.2	86.1	42	1.2–2.1, >2.5	months
LSTA2	62.7	62.4	63	0.9–2	°C
TAIRDA	71.6	77.2	66	9–37	%
TAIRMX	66.7	86.1	47	28.2–39.1	°C

Outcomes from univariate analyses show the percentage of correct classification, the sensibility and the specificity, and the range values of species occurrence.



**Fig. 4.** Histograms showing the average and standard deviation for each of the most important variables in the stepwise discriminant analysis. Variables with significant differences between presence and absence sites are indicated with asterisks on the boxes (\* $p < 0.01$ , \*\* $p < 0.001$ ).

analysis showed that those nine variables were able of correctly identifying 89% of the presence sites and 92.8% of the absence sites.

The variable with the most discriminatory power in fDA was the VPDMN. VPD measures the difference (deficit) between the amount of moisture in the air and how much moisture the air can hold when it is saturated (<http://en.wikipedia.org/wiki/>). Higher VPD values indicate a stronger drying tendency. In the same way, higher VPDMN values also indicate higher potential desiccation rate. In this work, VPDMN showed the highest values in the absence sites and the lowest in the presence ones (Fig. 2). The importance that this variable had in the analysis reflects the stenohydric status known for *R. pallescens* (Zeledón et al., 2006; Jurberg and Rangel, 1984), and suggests that this species could preferentially occupy places with minimum VPD values. The variance of land surface temperature (LSTVR) and the total variance of the three first Fourier components of the air temperature (TAIRDA) are higher in the absence sites than the presence sites, too. A large variation in land surface temperature and air temperature could diminish the habitat stability; in consequence, they could limit the occurrence of *R. pallescens* in a particular site.

Amplitude of the bi-annual cycle of land surface temperature (LSTA2) and maximum of air temperature (TAIRMX) have the highest values in presence sites and the lowest values in absence sites. Both land surface temperature and air temperature play an important role in the humidity of a place, especially for species with stenohydric status as *R. pallescens*.

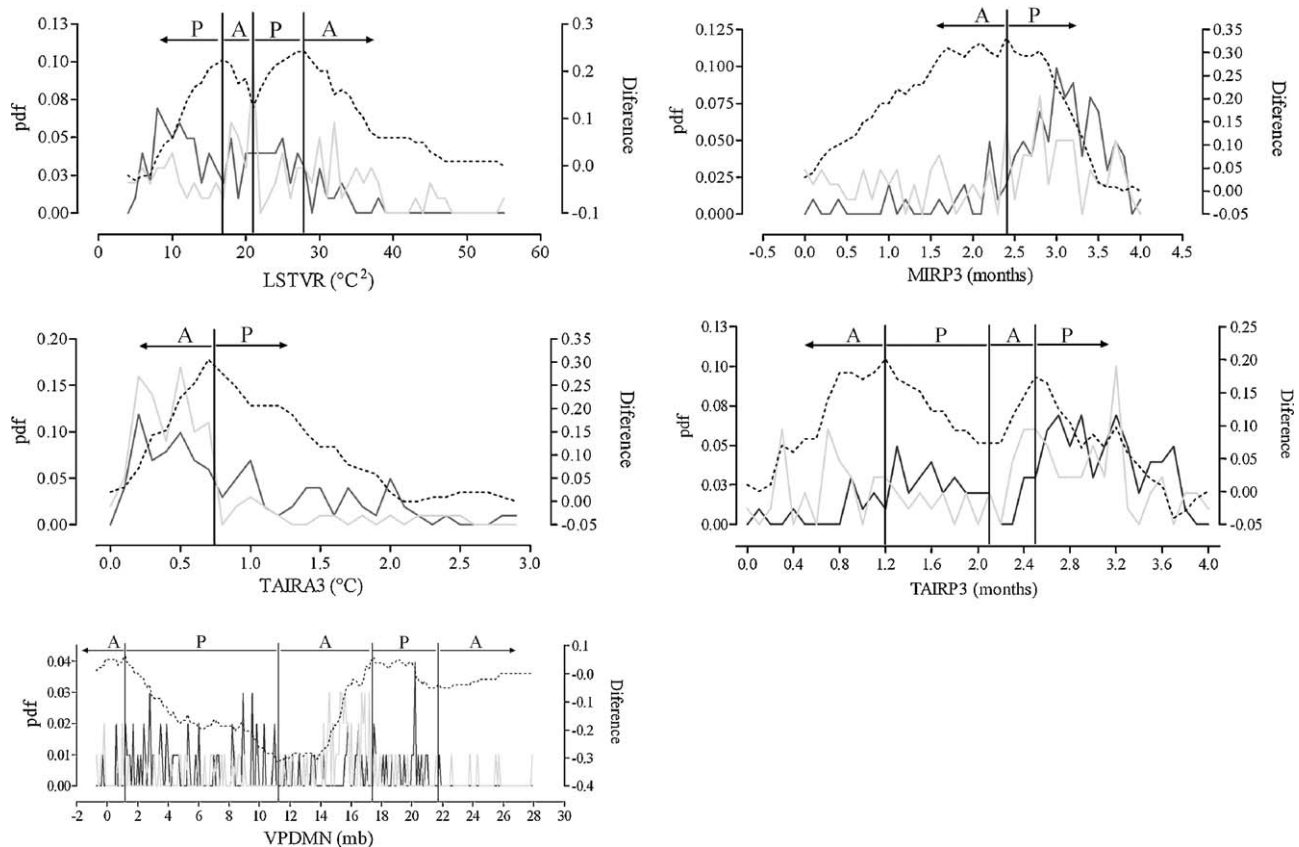
Stability of the geographical distribution model was tested for *post hoc* prediction by using standard DA over five new matrix data sets, each constructed with 75% of the data randomly selected from

the original matrix and the remaining data introduced as new, “unknown” data. Sensitivity and specificity of this *post hoc* analysis was lower than that with 100% of data (Table 3); however it was good enough to reveal the accuracy of the model. This results show the utility of discriminant functions for differentiate presence and absence sites.

According the biogeographical regions proposed by Morrone (2002), the *R. pallescens* distribution corresponds to the neotropical region located within the holotropical kingdom (Morrone, 2004). The predicted zone of potential distribution of *R. pallescens* (Fig. 3) covers additional areas to which this species occur. Generally species are not occupying all the areas environmentally appropriate for them; indeed, they occupy only a subset of these because their evolutionary history, their internal restrictions for dispersion and the effect of others variables (e.g. *A. butyracea* distribution), which were not took into account in this analysis. Thus, multiple ecological factors interact to explain why the species occupies some regions and not others; but it is notorious here that only nine climatic variables are good predictors of *R. pallescens* distribution. On the other hand, it is important to know the sites with appropriate characteristics for *R. pallescens* establishment, because its passive transport by humans and its posterior development in non-endemic zones is a reality confirmed in Triatominae with serious consequences for public health (Dujardin et al., 1998).

In the geographical distribution model, the predicted areas of *R. pallescens* presence not only match 89% the observed records of this species, but also coincided with the distribution of the most species of the genus *Rhodnius* except *R. domesticus* Neiva and Pinto, 1932, *R. nasutus* Stål, 1859 and *R. neglectus* Lent, 1954, according to Atlas of Chagas diseases vectors (Carcavallo et al., 1999; Fig. 3). This





**Fig. 5.** Probability density functions for presence (pdfP) and absence (pdfA) sites. The pdfP is showed by dark gray lines and pdfA by light gray lines. Dotted lines are indicating the difference between cumulative density functions of presence and absence. The threshold separates with higher probability the presence and absence sites.

suggests that distribution of other species of this genus is similarly associated with the same environmental variables, which may in turn be an additional indication of their common ancestry (Schofield and Dujardin, 1999).

Univariate analysis of the nine more important variables according to the fDA showed good sensibility (62–85.2%) and specificity (70.8–81.8%), but did not have much discriminatory power as when they were analyzed together (89% of sensibility and 92.8% of specificity). This indicates that individual variables are not good as predictors of *R. pallescens* presence, but it is necessary to study them in a multivariate way to determine their relative weight in the species ecology.

### Acknowledgments

The authors thank the financial support contributed by Colciencias, project no. 1115-05-11485, by the European Community Specific Support Activity—SSA (Project Trypanosomiasis Update, contract no: INCO-SA 515942), by the “Dirección General de Cooperación para el Desarrollo, Presidencia de Gobierno de la Generalitat Valenciana”, Valencia, España (Expediente 2000/3042) and by the Committee for the Development of the Investigation of the Universidad de Antioquia, Medellín, Colombia (CODI, project: CPT-0604). This work was benefited from the international cooperation through the European Community-Latin American Network for Research on the Biology and Control of Triatominae (ECLAT). We acknowledge TALA Research Group (University of Oxford Department of Zoology, Oxford) the provision of the images processed with Fourier analyses from where the environmental variables were extracted. We thank to anonymous reviewer who help to improve this manuscript.

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.meegid.2008.12.006.

### References

- Abad-Franch, F., Monteiro, F.A., 2007. Biogeography and evolution of Amazonian triatomines (Heteroptera: Reduviidae): implications for Chagas disease surveillance in humid forest ecoregions. *Mem. Inst. Oswaldo Cruz* 102, 57–69.
- Abad-Franch, F., Palomeque, F.S., Aguilar, V.H.M., Miles, M.A., 2005. Field ecology of sylvatic *Rhodnius* populations (Heteroptera, Triatominae): risk factors for palm tree infestation in western Ecuador. *Trop. Med. Int. Health* 10, 1258–1266.
- Beck, L.R., Lobitz, B.M., Wood, B.L., 2000. Remote sensing and human health: new sensors and new opportunities. *Emerg. Infect. Dis.* 6, 217–227.
- Calzada, J.E., Pineda, V., Montalvo, E., Alvarez, D., Santamaria, A.M., Samudio, F., Bayard, V., Caceres, L., Saldaña, A., 2006. Human trypanosome infection and the presence of intradomicile *Rhodnius pallescens* in the western border of the Panama Canal, Panama. *Am. J. Trop. Med. Hyg.* 74, 762–765.
- Carcavallo, R., Curto de Casas, S., Sherlock, I., Galíndez, I., Jurberg, J., Galvao, C., Mena, C., Noireau, F., 1999. Geographical distribution and alti-latitude dispersion. In: Carcavallo, R.U., Galíndez, I., Jurberg, J., Lent, H. (Eds.), *Atlas of Chagas' Disease Vectors in the Americas*, 3. Editora Fiocruz, Rio de Janeiro, pp. 747–792.
- Christensen, H.A., de Vasquez, A.M., 1981. Host feeding profiles of *Rhodnius pallescens* (Hemiptera: Reduviidae) in rural villages of Central Panama. *Am. J. Trop. Med. Hyg.* 30, 278–283.
- Cohen, J., 1960. Coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46.
- Dujardin, J.P., 2006. PADwin version 82, Institut de Recherches pour le Développement (IRD), France. Available in: <http://www.mpl.ird.fr/morphometrics/>.
- Dujardin, J., Schofield, C., Panzera, F., 2000. In: Académie Royale des Sciences d'Outre-Mer (Ed.), *Les vecteurs de la maladie de Chagas. Recherches taxonomiques, biologiques et génétiques*. Bruxelles, Belgique, pp. 14–15.
- Dujardin, J.P., Muñoz, M., Chavez, T., Ponce, C., Moreno, J., Schofield, C.J., 1998. The origin of *Rhodnius prolixus* in Central America. *Med. Vet. Entomol.* 12, 113–115.



- Dumonteil, E., Gourbiere, S., 2004. Predicting *Triatoma dimidiata* abundance and infection rate: a risk map for natural transmission of Chagas disease in the Yucatan peninsula of Mexico. *Am. J. Trop. Med. Hyg.* 70, 514–519.
- Eastman, R., 2006. IDRISI Andes version 15. Clark Labs, Clark University, Worcester, USA. URL: <http://www.clarklabs.org/>.
- Fernández, P., Díaz, P., 2003. Pruebas diagnósticas. *Cad. Aten. Primaria* 10, 120–124.
- Fisher, R., 1936. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* 7, 179–188.
- Gabriel, K., 1971. The Biplot-graphic display of matrices with applications to principal component analysis. *Biometrika* 58, 453–467.
- Gorla, D., 2002. Variables ambientales registradas por sensores remotos como indicadores de la distribución geográfica de *Triatoma infestans* (Heteroptera: Reduviidae). *Ecol. Austral* 12, 117–127.
- Guhl, F., Aguilera, G., Pinto, N., Vergara, D., 2007. Actualización de la distribución geográfica y ecoepidemiología de la fauna de triatominos (Reduviidae: Triatominae) en Colombia. *Biomédica* 27 (Suppl. 1), 143–162.
- Hay, S.I., 1997. Remote sensing and disease control: past, present and future. *Trans. R. Soc. Trop. Med. Hyg.* 91, 105–106.
- Hay, S.I., 2000. An overview of remote sensing and geodesy for epidemiology and public health application. *Adv. Parasitol.* 47, 2–36.
- Hay, S.I., Tatem, A.J., Graham, A.J., Goetz, S.J., Rogers, D.J., 2006. Global environmental data for mapping infectious disease distribution. *Adv. Parasitol.* 62, 36–75.
- Holben, B., 1986. Characteristics of the maximum value composite images from temporal AVHRR data. *Int. J. Rem. Sen.* 7, 1417–1434.
- Holdridge, L., 1947. Determination of world plant formations from simple climatic data. *Science* 105, 367–368.
- Jurberg, J., Rangel, E., 1984. Ciclo biológico de *Rhodnius pallescens* Barber, 1932 (Hemiptera, Reduviidae, Triatominae). *Rev. Bras. Biol.* 40, 569–577.
- Kitron, U., 2000. Risk maps: transmission and burden of vector-borne diseases. *Parasitol. Today* 16, 324–325.
- Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
- Lent, H., Wygodzinsky, P., 1979. Revision of the Triatominae (Hemiptera: Reduviidae), and their significance as vectors of Chagas' disease. *Bull. Am. Mus. Nat. Hist.* 163, 123–520.
- Moreno-Mejía, J., Jaramillo, N., López Silva, E., Ramírez, L., 1992a. Biological study of *Trypanosoma cruzi* and *T. rangeli* in the Magdalena valley and Antioquia-Colombia. *Mem. Inst. Oswaldo Cruz* 87 (Suppl. 1), 216.
- Moreno-Mejía, J., Jaramillo, N., López Silva, E., Ramírez, L., 1992b. Study of triatomines of the Magdalena valley and Northeast in the state of Antioquia Colombia. *Mem. Inst. Oswaldo Cruz* 87 (Suppl. 1), 216.
- Morrone, J., 2002. Biogeographic regions under track and cladistic scrutiny. *J. Biogeogr.* 29, 149–152.
- Morrone, J., 2004. La zona de transición sudamericana: caracterización y relevancia evolutiva. *Acta Entomol. Chilena* 28, 41–50.
- Pizarro, J., Romaña, C., 1998. Variación estacional de una población silvestre de *Rhodnius pallescens* Barber 1932 (Heteroptera: Triatominae) en la costa caribe colombiana. *Bull. Inst. Fr. Études Andines* 27, 309–325.
- Robinson, T., Rogers, D., Williams, B., 1997. Univariate analysis of tsetse habitat in the common fly belt of southern Africa using climate and remotely sensed vegetation data. *Med. Vet. Entomol.* 11, 223–234.
- Rogers, D.J., 2000. Satellites, space, time and the African trypanosomiasis. *Adv. Parasitol.* 47, 129–171.
- Rogers, D.J., Robinson, T.P., 2004. Tsetse distribution. In: Maudlin, I., Holmes, P.H., Miles, M.A. (Eds.), *The Trypanosomiasis*. CABI International, Wallingford, UK, pp. 139–179.
- Rogers, D.J., Williams, B.G., 1994. Tsetse distribution in Africa: seeing the wood and the trees. In: Edwards, P.J., May, R.M., Webb, N. (Eds.), *Large-Scale Ecology and Conservation Biology*. Blackwell Scientific Publications, Oxford, England, pp. 247–271.
- Rogers, D.J., Hay, S.I., Packer, M.J., 1996. Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data. *Ann. Trop. Med. Parasitol.* 90, 225–241.
- Romaña, C.A., Pizarro, J.C., Rodas, E., Guilbert, E., 1999. Palm trees as ecological indicators of risk areas for Chagas disease. *Trans. R. Soc. Trop. Med. Hyg.* 93, 594–595.
- SAS Institute Inc., 2006. JMP version 6.0.3. SAS Institute Inc., Cary, NC, USA. URL: <http://www.jmp.com/>.
- Schofield, C., Dujardin, J., 1999. Theories on the evolution of *Rhodnius*. *Actual. Biol.* 21, 183–197.
- Thomson, M.C., Connor, S.J., 2000. Environmental information systems for the control of arthropod vectors of disease. *Med. Vet. Entomol.* 14, 227–244.
- Townsend, P., Sánchez-Cordero, V., Beard, B., Ramsey, J., 2002. Ecologic niche modeling and potential reservoirs for Chagas disease. *Mexico. Emerg. Inf. Dis.* 8, 662–667.
- Vasquez, A.M., Samudio, F.E., Saldaña, A., Paz, H.M., Calzada, J.E., 2004. Eco-epidemiological aspects of *Trypanosoma cruzi*, *Trypanosoma rangeli* and their vector (*Rhodnius pallescens*) in Panama. *Rev. Inst. Med. Trop. Sao Paulo* 46, 217–222.
- Zeledón, R., Marin, F., Calvo, N., Lugo, E., Valle, S., 2006. Distribution and ecological aspects of *Rhodnius pallescens* in Costa Rica and Nicaragua and their epidemiological implications. *Mem. Inst. Oswaldo Cruz* 101, 75–79.