ECOLOGIC NICHE MODELING AND DIFFERENTIATION OF POPULATIONS OF *TRIATOMA BRASILIENSIS* NEIVA, 1911, THE MOST IMPORTANT CHAGAS' DISEASE VECTOR IN NORTHEASTERN BRAZIL (HEMIPTERA, REDUVIIDAE, TRIATOMINAE)

JANE COSTA, A. TOWNSEND PETERSON, AND C. BEN BEARD

Entomology Branch, Centers for Disease Control and Prevention, Atlanta, Georgia, and Entomology Department, Fundaçao Oswaldo Cruz, Rio de Janeiro, Brazil; Natural History Museum, University of Kansas, Lawrence, Kansas

Abstract: Ecologic niche modeling has allowed numerous advances in understanding the geographic ecology of species, including distributional predictions, distributional change and invasion, and assessment of ecologic differences. We used this tool to characterize ecologic differentiation of *Triatoma brasiliensis* populations, the most important Chagas' disease vector in northeastern Brazil. The species' ecologic niche was modeled based on data from the Fundação Nacional de Saúde of Brazil (1997–1999) with the Genetic Algorithm for Rule-Set Prediction (GARP). This method involves a machine-learning approach to detecting associations between occurrence points and ecologic characteristics of regions. Four independent "ecologic niche models" were developed and used to test for ecologic differences among *T. brasiliensis* populations. These models confirmed four ecologically distinct and differentiated populations, and allowed characterization of dimensions of niche differentiation. Patterns of ecologic similarity matched patterns of molecular differentiation, suggesting that *T. brasiliensis* is a complex of distinct populations at various points in the process of speciation.

INTRODUCTION

Successful control measures have led to great progress in combating *Triatoma infestans*, the main Chagas' disease vector in Brazil.[1] Currently, another species, *T. brasiliensis*, is the most important vector in semiarid areas of northeastern Brazil. *T. brasiliensis* is widely distributed geographically, occurring in 12 states and presenting variable natural infection rates.[2] In contrast with *T. infestans*, which is considered an introduced species in Brazil that occurs only in domiciliary ecotopes, *T. brasiliensis* is a native species also found in natural environments, being able to colonize and reinvade domiciles after successful control interventions.[3] A diversity of natural and artificial ecotopes has been recorded as suitable habitats for this species.[4,5]

T. brasiliensis presents at least four distinct populations that can be distinguished on the basis of patterns of coloration. These differences are also reflected in morphologic, ecologic, biologic, and genetic variation.[5–8] Nevertheless, considerable debate exists regarding the nature of these differences: Do the various *T. brasiliensis* populations simply represent morphs of one broadly distributed species, or do they represent distinct biologic entities that could have distinct roles in ecologic communities and in disease transmission cycles?

A relatively new tool in understanding the geographic ecology of species is ecologic niche modeling.[9] This approach centers on estimating the dimensions of species' ecologic niches—the ecologic space within which a species can maintain populations without immigration.[10,11] Using a customized genetic algorithm (a machine-learning approach to solving complex computational problems),[12–14] recent work has confirmed: (1) high predictive ability of the approach regarding species' distributions;[9,15–17] (2) the ability to predict species' potential distributions across scenarios of change, both on ecologic and evolutionary time scales;[18,19] (3) the ability to predict the course of species' invasions;[20] (4) the capacity to understand and predict the geographic outcomes of species'

interactions;[21] and (5) useful insight into various other aspects of species' distributional ecology.[9,22]

In this study, we apply these new tools of ecologic niche modeling to the question of whether *T. brasiliensis* populations present differentiated ecologic niches, a question critical to understanding the distribution, ecology, and significance of this species as a Chagas' disease vector.

MATERIALS AND METHODS

The *T. brasiliensis* occurrence data set was collected by the Fundação Nacional de Saúde of Brazil during 1997–1999 as part of the national control program. For this study, all municipalities in which *T. brasiliensis* was collected inside domiciles were considered, for a total of 111 unique locality samples. The four different *T. brasiliensis* populations were characterized according to their geographic distribution[5] and coloration patterns;[6,23] in all, sample sizes for the four populations were 43 for brasiliensis, 7 for melanica, 14 for juazeiro, and 4 for macromelasoma. Distributional points for *T. brasiliensis sensu lato* for which specimens have not been checked by one of the authors (JC) and identified as to morphotype were excluded from specific analyses and are depicted on maps with question marks.

Distributional modeling. Ecologic niches and potential geographic distributions were modeled using the *Genetic Algorithm for Rule-Set Prediction* (GARP).[12–14] In general, GARP focuses on modeling ecologic niches (the conjunction of ecologic conditions within which a species can maintain populations).[10] GARP relates ecologic characteristics of known occurrence points to those of points randomly sampled from the rest of the study region, seeking to develop a series of decision rules that best summarize those factors associated with the species' presence.[9] GARP includes several distinct algorithms for niche modeling in an artificial-intelligence-based approach. Occurrence points are divided evenly into training and test 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. Predictive accuracy is then evaluated based on 1,250 points resampled from the test data and 1,250 points sampled randomly from the study region as a whole. 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 1,000 iterations or until convergence.

All modeling in this study was carried out on a desktop implementation of GARP that is in a beta-testing stage for eventual public release. This implementation offers much-improved flexibility in choosing predictive environmental/ ecologic geographic information system (GIS) data coverage. In this case, we used 22 data layers: elevation, slope, and aspect from the U.S. Geological Survey's Hydro-1K data set (http://edcdaac.usgs.gov/gtopo30/hydro/); aspects of climate including annual mean (1960–1990) cloud cover, daily temperature range, minimum and maximum annual precipitation, vapor pressure, wet days, and wind speed, plus monthly data (January and July) for precipitation and the two temperature measures from the Intergovernmental Panel on Climate Change (http://www.ipcc.ch/); and tree cover and land use/ land cover from the Geological Survey's Global Land Cover Database (http://edcdaac.usgs.gov/glcc/glcc.html).

To optimize model performance, we developed 100 replicate models of the species' or population's ecologic niche based on random 50–50 splits of available occurrence points. Unlike previous applications, which either used single models to predict species' distributions[17] or summed multiple models to incorporate model-to-model variation,[20] we used a new procedure[21] for choosing best subsets of models. This procedure is based on the observations that: (1) models vary in quality, (2) variation among models involves an inverse relationship between errors of omission (leaving out true distributional areas) and commission (including areas not actually inhabited), and (3) best models (as judged by experts blind to error statistics) are clustered in a region of minimum omission of independent test 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. To choose best subsets of models, we: (1) eliminated all models but those that had no omission error based on the independent test points, (2) calculated the average area predicted present among these zero-omission models (% of the area analyzed), and (3) identified models that were within 1% of the overall average extent. The projection of the rule-sets for these models onto maps of North America provided distributional predictions.

Model quality was tested via the independent sets of points (50%) set aside before GARP modeling. A χ^2 test was used to compare observed success in predicting the distribution of test points with that expected under a random model. The predicted proportional area that is presented provides an estimate of occurrence points correctly predicted were the prediction to be random with respect to the distribution of the test points.) Ecologic differentiation was measured as the ability or inability of each population's model to predict the known points for each other population;[18] the resulting matrix of ecologic similarity or differentiation summarizes patterns of differentiation in ecologic space. Ecologic niches were visualized by combining (in ArcView 3.2) the summed best-subsets prediction with the ecologic/environmental coverage of particular interest and exporting the resulting attributes table.[17]
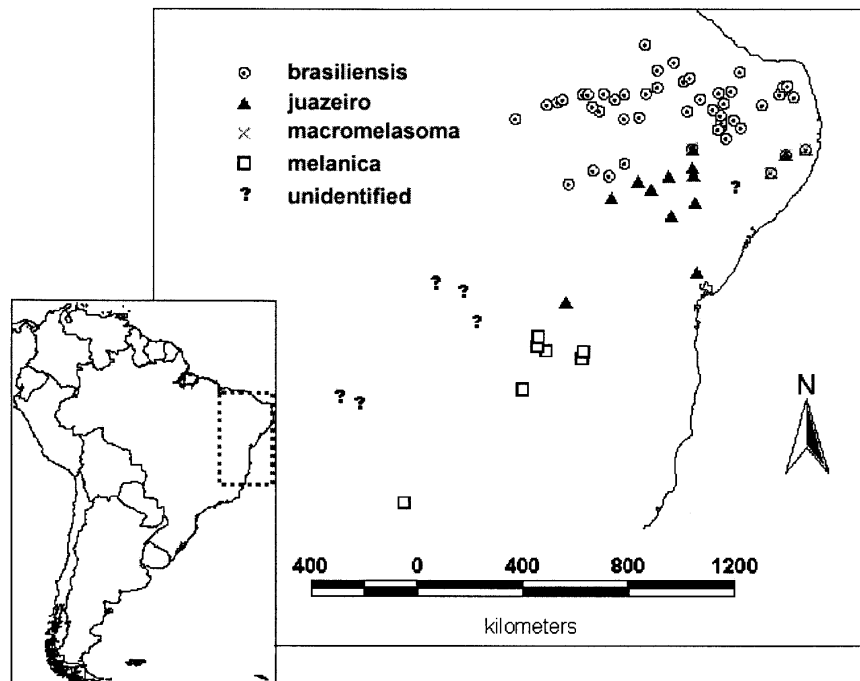

FIGURE 1. Known occurrences of four populations of *Triatoma brasiliensis*. Inset shows localization of study region within South America. Question marks indicate specimens of *T. brasiliensis* populations whose group membership has not been determined.

TABLE 1

Ecologic similarity matrix among four distinct populations of *Triatoma brasiliensis* (br-brasiliensis, me-melanica, ma-macromelasoma, ju-juazeiro) based on the ability of the model for one population to predict the distribution of another.

		Predictor			
		br	me	ma	ju
Predicted	br	0.98	0.00	0.44	0.51
	me	0.38	0.85	0.10	0.21
	ma	0.87	0.00	1.00	0.87
	ju	0.87	0.00	0.76	0.93

RESULTS

Distributional points of each *T. brasiliensis* population (brasiliensis, macromelasoma, juazeiro, melanica) (Figure 1) illustrate the largely allopatric and parapatric nature of the distribution of the populations.

The GARP models for each population were highly significant statistically. Based on the 50% subsets of points set aside before modeling, all the models were highly statistically significant. Indeed, compared with random models, the probability of such high predictivity was quite low (χ^2 tests, $10^{-50} < P < 10^{-20}$), indicating that the models were quite predictive and that our ecologic niche models indeed summarize the ecologic requirements of each population.

The ecologic similarity matrix among populations, based on the ability of the ecologic niche model for one population to predict the geographic distribution of each other population, shows that brasiliensis was best able to predict the distributions of other populations. Interestingly, ecologic similarity was not symmetric—brasiliensis predicts macromelasoma and juazeiro very well (0.87), yet the converse was not true (0.44 and 0.51). The distribution of melanica was poorly predicted by other populations (0.38, 0.10, and 0.21), and melanica was little able to predict other populations (Table 1).

To allow visualization of niche variation, we graphed ecologic niche models in a small subset of the modeled ecologic space, using the variables annual mean precipitation (in mm) versus annual mean temperature (in °C). Brasiliensis is modeled to occupy regions presenting 16.5–22.5°C and 10–35 mm of precipitation (Figure 2); juazeiro was similar but showed a narrower ecologic amplitude (Figure 2). Macromelasoma presented the highest amplitude, occupying two possibly disjunct ecologic zones in which annual mean temperature is 11.0–23.5°C and annual mean precipitation is 10–45 mm (Figure 2). Melanica appears restricted to a narrow ecologic zone of 18–19°C and 20–35 mm of precipitation (Figure 2).

Analyzing the four *T. brasiliensis* populations together, as a single ecologic unit, yields both a niche model and a predicted geographic distribution much broader than the individual niches or modeled distributions of individual populations (Figure 3). The breadth of this niche was significantly greater than that of any of the individual populations, and indeed it was significantly greater than the sum of the individual populations' niches, reflecting the ecologic differentiation among the populations currently lumped within *T. brasiliensis*. The differentiated phenotypes showed no clear relationship to any geographic gradient of climate variables; for example, the distribution of phenotypes along the ecologic dimensions' wet days and solar radiation showed no clear pattern (Figure 4).

DISCUSSION

Application of GARP modeling technology in understanding species' distributions in geographic and ecologic space is relatively new, and yet increasingly supported by rigorous empirical tests.[9,12–20,22,24,25] This article represents a first application in the field of epidemiology and disease transmission.

The ecologic matrix generated from the ecologic niche models corresponds very closely with similarity in molecular
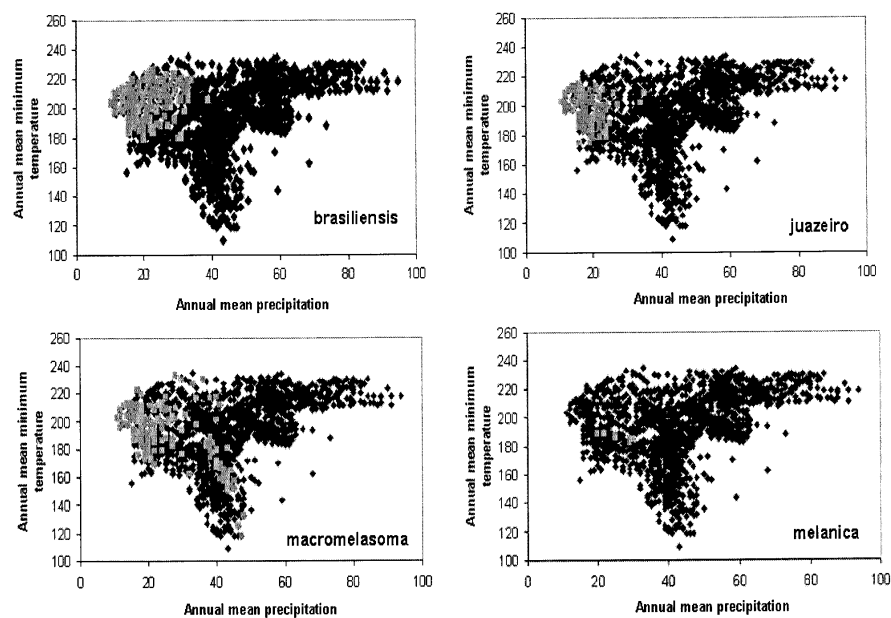

FIGURE 2.   Distribution of four populations of *Triatoma brasiliensis* in ecologic space (annual mean minimum temperature vs. annual mean precipitation, units multiplied by 10). The ecologic space for each studied population is in gray, and all environments are in black.
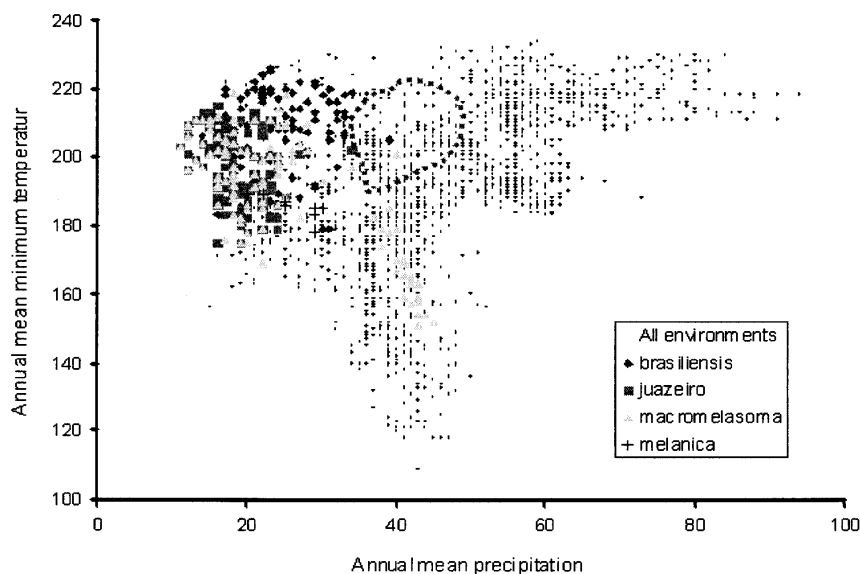
FIGURE 3. Distribution of four populations of *Triatoma brasiliensis* in ecologic space (annual mean minimum temperature vs. annual mean precipitation, units multiplied by 10) in relation to the availability of environmental combinations across the study region. The dotted area on the graph indicates the area over predicted when all four populations were analyzed as one species, showing the artificial nature of the four populations when analyzed as a single species.

character sets (e.g., isoenzymes, mitochondrial DNA sequences).[7,26] These parallel patterns of genetic and ecologic differentiation suggest that current processes of population differentiation and possible speciation in *T. brasiliensis* occur in several evolutionary dimensions, including ecologic dimensions. Melanica populations in particular appear to be qualitatively different from the other populations; however, the identity of the undetermined populations (Figure 1) needs to be checked because if they pertain to melanica, their distinct geographic situation may change the view of this population's ecologic amplitude. Nevertheless, these marginal (undeter-

mined) distributional points do not represent areas with high densities of this species group: Only 543 *T. brasiliensis* specimens were collected from 1993–1999, whereas in the same 6 years, a total of 422,965 specimens were collected in the rest of the distribution of *T. brasiliensis*.[2] This low density suggests that these areas represent marginal areas within the species' distribution.

The lack of correlation observed between climatic gradients and phenotypic differences of the four populations (Figure 4) supports the idea that the characteristic coloration patterns of each population are genetically based and not
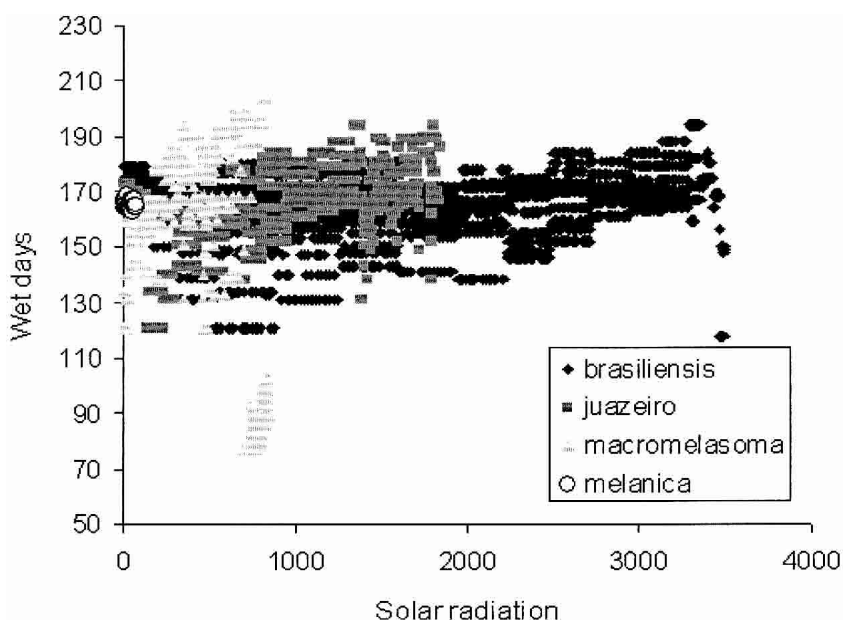

FIGURE 4. Distribution of four populations of *Triatoma brasiliensis* in ecologic space (wet days vs. solar radiation, units multiplied by 10), addressing the question of whether the four populations represent a cline.

influenced by environmental differences across the distribution of the complex. This point has already been emphasized: Populations, even when reared with the same environmental conditions and food sources, maintained their original color patterns.[6] Hence, the case against ecotypic variations producing the different *T. brasiliensis* population phenotypes is supported.

The Chagas' disease control program of the Fundação Nacional de Saúde of Brazil has shown that rapid ecologic shifts can occur in response to control measures, demanding constant entomologic surveillance.[3] In this sense, developing methods for evaluating the ecologic potential of vector species in a proactive manner to monitor Chagas' disease vectors is important. Recently, in Rio Grande do Sul, Brazil, after *T. infestans* control measures, *T. rubrovaria* became the most frequently found species inside human domiciles, though it was previously known only in sylvatic and peridomiciliary areas. It is important to mention that *T. rubrovaria* now presents remarkable invasive behavior, and in spite of it is not being found colonizing domiciles, the species is reported in higher population densities than those observed for *T. infestans* before the control program actions.[27]

Although encompassing only a few possible niche dimensions,[28] the ecologic/environmental data sets currently available cover many of the major physical variables that commonly influence species' macro-distributions. These ecologic niche models have clarified ecologic differences and similarities among populations of *T. brasiliensis*: Populations of this species group are clearly and distinctly differentiated in ecologic dimensions. Future analyses will approach the challenges of predicting dispersal potential for this species into new areas, changes to be expected as a part of climate change, and transitions between sylvatic and domestic distributional situations for Chagas' disease vectors.

Authors addresses: Jane Costa, C. Ben Beard, Entomology Branch, Centers for Disease Control and Prevention, 4770 Buford Highway, Chamblee, Atlanta, Georgia F22, 30341. A. Townsend Peterson, Department of Ecology & Evolutionary Biology, Natural History Museum, University of Kansas, Lawrence, Kansas 66045.

Reprint requests: Jane Costa, Entomology Branch, Centers for Disease Control and Prevention, 4770 Buford Highway, Atlanta, Georgia F22, 30341, Phone: 770-488-3609, Fax: 770-488-4258, E-mail: zgq6@cdc.gov.

REFERENCES

1. Moncayo A, 1999. Progress towards interruption of transmission of Chagas disease. *Mem Inst Oswaldo Cruz 94 (suppl I):* 401–404.

2. Costa J, Almeida CE, Dotson E, Lins A, Vinhaes M, Silveira AC, Beard CB, 2001. Domiciliary surveillance of *Triatoma brasiliensis* (Hemiptera-Reduviidae-Triatominae). Mem Inst Oswaldo Cruz (Submitted).

3. Silveira AC, Vinhaes MC, 1999. Elimination of vector-borne transmission of Chagas disease. *Mem Inst Oswaldo Cruz 94 (suppl I) 405:* 411.

4. Alencar JE, 1987. *História natural da doença de Chagas no Estado do Ceará.* Imprensa Universitária, UFC, Fortaleza, Ceará.

5. Costa J, Almeida JR, Britto C, Duarte R, Marchon-Silva V, Pacheco R, 1998. Ecotopes, natural infection, and trophic resources of *Triatoma brasiliensis* (Hemiptera, Reduviidae, Triatominae). *Mem Inst Oswaldo Cruz 93:* 7–13.

6. Costa J, Barth OM, Marchon-Silva V, Almeida CE, Freitas-Sibajev MG, Panzera F, 1997a. Morphological studies on the *Triatoma brasiliensis Neiva,* 1911 (Hemiptera, Reduviidae, Triatominae) genital structures and eggs of different chromatic forms. *Mem Inst Oswaldo Cruz 92:* 493–498.

7. Costa J, Freitas-Sibajev MG, Marchon-Silva V, Pires MQ, Pacheco R, 1997b. Isoenzymes detect variation in populations of *Triatoma brasiliensis* (Hemiptera-Reduviidae-Triatominae). *Mem Inst Oswaldo Cruz 92:* 459–464.

8. Costa J, Marchon-Silva V, 1998. Período de intermuda e resistência ao jejum de diferentes populações de *Triatoma brasiliensis* (Hemiptera, Reduviidae, Triatominae). *Entomologia & Vectores 5:* 23–34.

9. Peterson AT, Stockwell DRB, Kluza DA, In press. Distributional prediction based on ecological niche modeling of primary occurrence data. Scott JM, ed. *Predicting Species Occurrences: Issues of Scale and Accuracy.* Washington, D.C.: Island Press.

10. Grinnell J, 1917. Field tests of theories concerning distributional control. *Am Naturalist 51:* 115–128.

11. MacArthur R, 1972. *Geographical Ecology.* Princeton, N.J.: Princeton University Press.

12. Stockwell DRB, Noble IR, 1992. Induction of sets of rules from animal distribution data: A robust and informative method of analysis. *Math Comp Simul 33:* 385–390.

13. Stockwell DRB, 1999. Genetic algorithms II. Fielding, AH, ed. *Machine Learning Methods for Ecological Applications.* Boston: Kluwer Academic Publishers, 123–144.

14. Stockwell DRB, Peters DP, 1999. The GARP modelling system: Problems and solutions to automated spatial prediction. *Int J Geographic Information Syst 13:* 143–158.

15. Peterson AT, Cohoon KC, 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecol Mod 117:* 159–164.

16. Peterson AT, Ball LG, Cohoon KC, 2002. Predicting distributions of tropical birds. *Ibis 144:* 27–32.

17. Peterson AT, 2001. Predicting species' distributions based on ecological niche modeling. *Condor 103:* 599–605.

18. Peterson AT, Soberon J, Sánchez-Cordero V, 1999. Conservatism of ecological niches in evolutionary time. *Science 285:* 1265–1267.

19. Peterson AT, Sánchez-Cordero V, Soberón J, Bartley J, Buddemeier RW, Navarro-Sigüenza AG, 2001. Effects of global climate change on geographic distributions of Mexican Cracidae. *Ecol Mod 144:* 21–30.

20. Peterson AT, Vieglais DA, 2001. Predicting species invasions using ecological niche modeling. *BioScience 51:* 363–371.

21. Anderson RP, Lew D, Peterson AT, in preparation.

22. Peterson AT, Egbert SL, Sánchez-Cordero V, Price KP, 2000. Geographic analysis of conservation priorities using distributional modelling and complementarity: Endemic birds and mammals in Veracruz, Mexico. *Biol Conserv 93:* 85–94.

23. Galvão AB, 1956. *Triatoma brasiliensis macromelasoma* n.subsp. (Reduviidae, Hemiptera). *Rev Bras de Malariol Doencas Trop 7:* 455–457.

24. Joseph L, 1996. Preliminary climatic overview of migration patterns in South American austral migrant passerines. *Ecotropica 2:* 185–193.

25. Stockwell DRB, Peterson AT, 2001. Controlling bias in biodiversity data. Scott JM, ed. *Predicting Species Occurrences: Issues of Scale and Accuracy.* Washington, D.C.: Island Press.

26. Costa J, Monteiro F, Beard CB, 2001. *Triatoma brasiliensis* Neiva, 1911, the most important Chagas disease vector in Brazil—Phylogenetic and population approaches correlated to epidemiologic importance. *Am J Trop Med Hyg 65:* 280.

27. Almeida CE, Vinhaes MC, Almeida JR, Silveira AC, Costa J, 2000. Monitoring the domiciliary and peridomiciliary invasion process of *Triatoma rubrovaria* in the State of Rio Grande do Sul, Brazil. *Mem Inst Oswaldo Cruz 95:* 761–768.

28. Hutchinson GE, 1958. Concluding remarks. *Cold Spring Harbor Symposium in Quantitative Biology 22:* 415–427.