UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ECOLOGICAL NICHE MODELING: INTEGRATING PATTERNS OF SPECIES RICHNESS, FACTORS SHAPING GEOGRAPHICAL RANGE LIMITS, AND CONSERVATION PLANNING

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

GABRIEL C. COSTA
Norman, Oklahoma
2009

ECOLOGICAL NICHE MODELING: INTEGRATING PATTERNS OF SPECIES RICHNESS, FACTORS SHAPING GEOGRAPHICAL RANGE LIMITS, AND CONSERVATION PLANNING

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF ZOOLOGY

BY

Dr. Laurie Vitt (Chair)

Dr. Janalee Caldwell

Dr. Michael Kaspari

Dr. Jeffrey Kelly

Dr. May Yuan

Acknowledgments

I thank members of my advisory committee Laurie Vitt, Janalee Caldwell, Michael Kaspari, Jeffrey Kelly, Tarek Rashed, and May Yuan for their guidance throughout the development of my dissertation work. I am especially indebted to my advisor Laurie Vitt and to Janalee Caldwell for the opportunity to study in the United States. I thank Rosemary Knapp for her guidance and help with all the procedures and requirements of graduate school. I am indebted to the staff and faculty of the Department of Zoology and Sam Noble Oklahoma Museum of Natural History. My research was partially funded by the following grants and/or institutions: M. Blanche Adams research and summer scholarships (University of Oklahoma, Department of Zoology), Graduate Student Senate grant (University of Oklahoma), Robberson Research Grant (Graduate College), Conservation International – Brazil, and a National Science Foundation grant to Laurie Vitt and Janalee Caldwell (DEB-0415430). A fellowship from CAPES/Fulbright (15053155–2018/04-7) defrayed costs involved in tuition and fees and living expenses in the US for four years. I thank my friends and collaborators in Brazil: Guarino Colli, Ricardo Machado, and Cristiano Nogueira for their support during my field trips in Brazil, for lending field equipment, and providing feedback on several parts of my research. Santos Balbino, Renan Bosque, Renato Recoder, and Paula Valdujo provided invaluable help during fieldwork. I thank all my great friends in Brazil, whether herpetologists or not: Adrian Garda, Alex Nascimento, Ana Herminia Soares, Daniel Mesquita, Fernanda Werneck, Gustavo Vieira, Henry Novion, Igor Carvalho, Julio Zolini, Rafael Chaves, and Reinaldo Moreira. My special thanks go to Don Shepard for

discussing papers and reading many versions of my proposals, paying special attention to my poorly elaborated English. Throughout all phases of my graduate work, I had unconditional support from my family, José Mario, Lúcia Helena, Eduardo, and Diogo Costa. I especially thank my wife Rachel Costa. Her companionship and friendship always inspired me, and gave me strength to keep working hard even in the most difficult moments.

Table of Contents

I. Squamate richness in the Brazilian Cerrado and its environmental–climatic associations

II. Sampling bias and the use of ecological niche modeling in conservation planning: A field evaluation in a biodiversity hotspot

III. Detecting the influence of climatic variables on species' distributions: a test using GIS niche-based models along a steep longitudinal environmental gradient

IV. Biogeography of the Amazon molly: ecological niche and range limits of an asexual hybrid species

Abstract

The concept of the niche has been a central theme of ecology for over half a century. With recent advances in Geographic Information Systems (GIS) and statistical procedures, the inferential method of ecological niche modeling (ENM) has emerged as an important tool for studies in ecology and evolutionary biology. This method consists of modeling species' distributions based on attributes of the environment that should be correlated with species' niche requirements. The models assume that geographic distribution and characteristics of the environment are related to the species niche, and therefore are tied to ecological niche theory. Over the past several years, ENM has been widely used to address numerous research questions in ecology, evolutionary biology, and conservation. Here I applied ENM in novel ways, focusing on central themes in Macroecology, Biogeography, and Conservation. First, I analyzed species richness in the Brazilian Cerrado, identifying areas of high richness and analyzing the relationship between richness and climatic variables to test predictions of large-scale hypotheses concerning determinants of species richness. Second, I assessed the performance of two ENM algorithms in predicting species richness and composition of an unsampled area of conservation interest in the Brazilian Cerrado. These analyses provided an overall assessment of the weaknesses and strengths of ENM for conservation planning. Lastly, I applied ENM to two different situations in an effort to understand and disentangle the relative contributions of abiotic factors, competition, dispersal barriers, and biotic interactions in determining species' range limits. Overall, I was able to show that by applying ENM techniques it is possible to gain insights into a vast number of ecological

questions. These studies highlight the complexity of studying species range limits, but reiterate that an integrative approach is necessary to understand this phenomenon.

Chapter I

Squamate richness in the Brazilian Cerrado and its environmental–climatic associations

(formatted for *Diversity and Distributions*)

ABSTRACT

We investigated patterns of squamate species richness in the Brazilian Cerrado, identifying areas of particularly high richness, and testing predictions of large-scale richness hypotheses by analyzing the relationship between species richness and environmental climatic variables. We used point localities from museum collections to produce maps of the predicted distributions for 237 Cerrado squamate species, using niche-modeling techniques. We superimposed distributions of all species on a composite map, depicting richness across the ecosystem. Then, we performed a multiple regression analysis using eigenvector-based spatial filtering (Principal Coordinate of Neighbor Matrices – PCNM) to assess environmental–climatic variables that are best predictors of species richness. We found that the environmental–climatic and spatial filters multiple regression model explained 78% of the variation in Cerrado squamate richness ($r^2 = 0.78$; $F = 32.66$; $P < 0.01$). Best predictors of species richness were: annual precipitation, precipitation seasonality, altitude, net primary productivity, and precipitation during the driest quarter. A model selection approach revealed that several mechanisms related to the different diversity hypothesis might work together to explain richness variation in the Cerrado. Areas of higher species richness in Cerrado were located mainly in the southwest, north, extreme east, and scattered areas in the northwest portions of the biome. Partitioning of energy among species, habitat differentiation, and tolerance to variable environments may be the primary ecological factors determining variation in squamate richness across the Cerrado. High richness areas in northern Cerrado, predicted by our models, are still poorly sampled and biological surveys are warranted in that region. The southwestern region of the Cerrado exhibits high species richness and is also undergoing high levels of deforestation. Therefore, maintenance of existing reserves, establishment of ecological corridors among reserves, and creation of new reserves are urgently needed to ensure conservation of species in these areas.

Key words: Conservation, niche-modeling, reptiles, richness gradients, spatial statistics, species distribution.

INTRODUCTION

The Cerrado is the second largest biome in South America, originally occupying about 20% of Brazil's land surface (Ab'Saber, 1977; Ratter *et al.*, 1997; Silva & Bates, 2002). It is located mainly in central Brazil and shares contact zones with almost all other Brazilian biomes. The Cerrado consists of savanna vegetation distributed in a gradient from open grassland to forested sites with closed canopy (Eiten, 1972). A recent overview of the Cerrado fauna and flora is available in Oliveira & Marquis (2002).

Along with 34 other ecosystems, the Cerrado is considered a global biodiversity "hotspot," as defined by diversity, endemism, and human threats (Myers, 2003; Mittermeier *et al.*, 2005). A recent estimate based on satellite remote sensing shows that approximately 55% of the Cerrado's original vegetation has been removed, and the annual deforestation rate is higher than that in the Amazon (Machado *et al.*, 2004b). If the current destruction rate is not reversed, no natural areas will remain in the Cerrado by 2030 (Machado *et al.*, 2004b). Agriculture and cattle farming, the major economic activities in Brazil, are the main threats to Cerrado, and the pace of deforestation is not likely to change in the near future (Klink & Moreira, 2002; Klink & Machado, 2005). Hence, Cerrado conservationists face the challenge of preserving the maximum amount of biodiversity using an ever-decreasing available space. Immediate action is needed because every year vast areas of the Cerrado are lost (Machado *et al.*, 2004b).

Despite its great diversity, endemism, and level of deforestation, less than 3% of the Cerrado's land area is protected in reserves (Machado *et al.*, 2004a; Rylands *et al.*, 2005). To identify priority areas for conservation, detailed information on species distributions is necessary. Vast areas of Cerrado are still poorly sampled (Silva, 1995;

Ratter *et al.*, 1997; Aguiar, 2000; Felfili *et al.*, 2004) and biological surveys require significant investments of both time and money. Unfortunately, the speed at which the Cerrado is being destroyed does not allow sufficient time to adequately survey the entire region.

Recently, new advances in Geographical Information Systems (GIS) allow modeling of species' distributions based on attributes of the environment that should be correlated with species' niche requirements (Peterson, 2001; Guisan & Thuiller, 2005). These models identify previously unsampled locations where the species has a high probability of occurrence because the environment is similar to conditions at known occurrence localities. These techniques do not account for species interactions or historical factors, but they combine spatially explicit factors to yield potential geographic distributions of species (Araujo & Guisan 2006). These methods have been successfully applied to several animal groups in different ecosystems, including butterflies in Finland (Luoto *et al.*, 2002), birds in North America (Peterson, 2001), and lizards in Madagascar (Raxworthy *et al.*, 2003). A recent study comparing different methods for predicting species distributions using presence-only data concluded that overall, these techniques produce good results (Elith *et al.*, 2006).

Additionally, niche modeling can be used to generate distribution maps for individual species, which can be superimposed to generate maps showing gradients in species richness. The occurrence and determinants of large-scale patterns of species richness are fundamental questions in ecology that are still far from being resolved (Ricklefs, 2004). Several hypotheses exist to explain such patterns, some of which relate richness to particular environmental–climatic parameters (see Table 1 for a list of current hypotheses and how they relate to environmental–climatic variables). For example, the "available energy hypothesis" states that energy partitioning among species is the most important factor limiting species richness (Wright, 1983). According to the predictions of this hypothesis, richness should be highly correlated with temperature, precipitation, and potential evapotranspiration because higher available energy can support more species (Fraser & Currie, 1996; Moser *et al.*, 2005).

The goal of our study was two-fold. First, we applied niche-modeling techniques to a large dataset on Cerrado squamate (lizards, snakes, and amphisbaenians) occurrences to produce distribution maps for all Cerrado species. We superimposed these maps to make a single map that depicts patterns of squamate species richness across the biome. After that, we analyzed the richness map generated using multiple regression and spatial analysis techniques to determine which environmental–climatic variables exert a greater influence on species richness, and we used model selection to evaluate which hypothesis best explains species richness gradients in the Brazilian Cerrado.

Second, we provided biological information to support conservation decisions. Using the richness map produced, we identified areas of high richness within the Cerrado. Although we used only squamate reptiles, previous works have shown that richness across large geographical scales is correlated among several different animal groups (Lamoreux *et al.*, 2006); thus, patterns that emerge from this analysis can be extrapolated for other animal groups in Cerrado. This analysis will provide substantial data for further studies on identifying priority areas for conservation.

METHODS

Data collection

We collected locality data for squamate species occurring in the Cerrado from museums, literature, and fieldwork, and created a database with species names, localities, and geographical coordinates. The database consists of records from the major collections for Cerrado squamates: Coleção Herpetológica da Universidade de Brasília (CHUNB), Museu de Zoologia da Universidade de São Paulo (MZUSP), and Instituto Butantan (IB). When available, we used geographical coordinates from museum databases or published studies, and in other instances, we approximated point localities from locality descriptions using georeferencing techniques and gazetteers (NGA, 2005).

In the analysis, we used only those species that had at least one data point within the Cerrado biome. However, for species whose distributions spanned multiple biomes, we also included data points outside the Cerrado, because characteristics of these points can help identify suitable regions for species occurrence within the Cerrado. At the time of the analysis the database contained a total of 237 species known to occur in the Cerrado (Appendix A).

Niche-modeling

For each of the 237 species, we produced predicted distribution maps, using the software DesktopGARP®. This software uses the GARP algorithm (Genetic Algorithm for Rule-set Prediction), which includes several distinct algorithms in an iterative, artificial intelligence approach based on species presence data points (Stockwell & Peters, 1999). DesktopGARP® software generates each species' predicted distribution based upon characteristics of environmental–climatic variables for localities in which a given species has been previously collected. We used a total of 22 variables (Supplementary material). Variables were downloaded from the Worldclim project (Hijmans *et al.*, 2005). Details, descriptions, and files for downloading are available free on-line at: http://www.worldclim.org/.

We used the following options while running the software: Optimization parameters – 20 runs, 0.001 convergence limit, and 1000 maximum interactions; Rule Types – Atomic, Range, Negated Range, and Logistic regression; Best subset active, 5% omission error, 40% commission error, and 50% of points for training; Omission measure = extrinsic, and Omission threshold = Hard; 10 models under hard omission threshold.

The output of DesktopGARP® consists of Arc/Info grid maps with 'zeros' where the species do not occur and 'ones' where the species are predicted to occur. We used the area covered by the coincidence of 7 of the 10 models in the best subset selection (optimum models considering omission/commission relationships; Anderson *et al* 2003) as the estimation of the distribution of each species. This approach is called ensemble forecasting and has been recently reviewed by Araujo & New (2007). By doing that and by setting the commission error to 40%, we believe we added a component of conservatism in the predictions of GARP that otherwise could extrapolate too much in the direction of areas too far from where the species have previously been collected. After producing such maps using the same criteria for all 237 species, we overlaid the distributions of all species into a composite map using the software ArcGis™ (Environment Systems and Research Institute Inc. Redlands, California). This final map consisted of a grid where the value of each cell was the predicted number of species

(species richness), calculated by summing the number of overlaid corresponding cells with positive occurrence results.

Multiple Regression

We performed a multiple regression analysis to evaluate which environmental–climatic variables are the best predictors of squamate richness. We selected 300 random points in the Cerrado region using a filter option in the software IDRISI Kilimanjaro (Eastman, 2003). Values of richness (dependent variable) and environmental variables (independent variables) for the selected points were used to build a matrix. We did not use all environmental variables used to run GARP, because including many highly correlated variables in a multiple regression creates several theoretical and statistical problems, especially in the estimation of partial regression coefficients (Tabachnick, 2000). We selected variables that were previously identified as influencing species richness patterns and were not highly correlated ($r < 0.9$) (Table 1, Appendix B). We add to the regression the variable net primary productivity downloaded from the Atlas of Biosphere website: http://www.sage.wisc.edu/atlas. This variable has been previously suggested to influence species richness (Table 1). We performed an exploratory analysis of the data matrix where we identified and eliminated univariate and multivariate outliers. The spatial distribution of points used in the analysis can be seen in Figure 1.

Ecologists have long recognized that macroecological and biodiversity data show strong spatial patterns, which are driven by structured biological processes and, therefore, are usually spatially autocorrelated (Legendre, 1993; Diniz-Filho *et al.*, 2003). Spatial autocorrelation occurs when variable values at a certain distance apart are more (positive autocorrelation) or less similar (negative autocorrelation) than expected by chance (Legendre, 1993). Failure to account for spatial autocorrelation in multiple regression analysis results in inflation of Type I error (Legendre, 1993; Diniz-Filho *et al.*, 2003). To alleviate this problem, we used spatial filters obtained by Principal Coordinates Neighbour Matrices (PCNM). This method partitions variation between spatial and environmental components, and works well at different spatial scales (Borcard & Legendre, 2002; Borcard *et al.*, 2004; Diniz-Filho & Bini, 2005). In the analysis, we treated the filters as candidate predictor variables, along with other environmental predictors. Using this approach, the effects of environmental predictors are evaluated as partial effects, taking space into account explicitly (Rangel *et al.*, 2006). We selected filters in an iterative process, by examining the pattern of spatial autocorrelation of regression residuals. Filters were selected to minimize both the autocorrelation among residuals and the number of filters used in regression. To investigate the presence of spatial autocorrelation, we used Moran's coefficient, the most commonly used statistic for autocorrelation analysis in macroecology and biogeography (Diniz-Filho *et al.*, 2003).

In addition, based on r square values for partial regressions made with filters alone, and environmental predictors alone we divided the total variation explained by the model into: explained by space only, explained by the environmental variables only, and shared explained variance. All spatial analyses were performed in SAM – Spatial Analysis in Macroecology (Rangel *et al.*, 2006), which is freely available at: www.ecoevol.ufg.br/sam.

In order to determine which hypotheses best explained variation in species richness in the Cerrado squamates, we conducted separate regressions to fit each of the hypothesis presented in Table 1 (with the addition of two mixed models, one incorporating all variables related to each hypothesis, and other using only the variables pointed out as significant by the multiple regression). We used model selection based on the sample corrected Akaike Information Criteria (AICc). We used AIC because information–theoretic approaches are strongly recommended alternatives to traditional hypothesis testing. Particularly, AIC is more tolerant to violations of the assumptions of parametric statistics that are commonly encountered in ecological data (e.g. normality), it is specially useful when comparing multiple working hypotheses, and it does not rely solely on the use of random *P*-values for determining significance (Anderson *et al.*, 2000, Burnham & Anderson, 2004)

Finally, to compare the pattern predicted by niche modeling with the actual available data, we mapped species locality points and determined species richness for a Cerrado grid consisting of 100,000 ha cell size (Figure 2). This approach allowed us to identify whether a spatial bias in sampling effort was present in the final modeling map (i.e., areas that have more species collected coincide with the areas the model pointed out as having higher richness).

RESULTS

The composite map summing all 237 species individual maps is shown in Figure 3. In this map, areas with greater species richness are located mainly in the southwest portion of the Cerrado biome, in areas corresponding to the southern portion of Góias state (GO), northern portion of Mato Grosso do Sul state (MS), and the very southern part of Mato Grosso state (MT). An interesting pattern is the presence of an area with high richness in northern Cerrado, corresponding to the border between the states of Tocantins (TO) and Maranhão (MA). In addition, areas of high richness exist in the extreme west portion of Cerrado in Minas Gerais state (MG) and scattered areas in Mato Grosso and Rondônia (RO) states (Figure 3).

Mapping of raw data shows a slight sampling bias toward the southeastern portion of Cerrado, where the largest biological collections (MZUSP and IB) are located (Figure 2A). However, the results of our modeling are not highly influenced by this bias, since areas with the greatest diversity in Cerrado do not overlay completely with this pattern. In addition, high richness areas were found by the modeling in regions where the sampling effort was extremely poor, such as the northern portion of Cerrado (Figure 2B).

In the iterative process of filter selection for the multiple regression analysis, nine eigenvector spatial filters (PCNM) were sufficient to remove most spatial autocorrelation in regression residuals (Figure 4). An examination of the spatial correlogram based on Moran's coefficient of the dependent variable (species richness) and regression residuals confirmed this pattern (Figure 4). Addition of more filters did not significantly reduce autocorrelation in the residuals; therefore, we used only the best nine filters to minimize the number of independent variables.

The climatic–environmental and spatial filters (PCNM) multiple regression model explained 78% of the variation in Cerrado squamate richness ($r^2 = 0.78$; $F = 32.66$; $P < 0.01$). The partial regression approach revealed that the variance explained by space alone was 37%, the variance explained by the environmental variables alone was 12%, and the shared explained variance was 28%.

Based on the analysis including both climatic–environmental variables and the PCNM spatial filters, annual precipitation, precipitation seasonality, altitude, net primary productivity, and precipitation of the driest quarter were the best predictors of species richness (Table 2). Using spatial filters in this analysis ensured that environmental variables were evaluated while taking into account spatial autocorrelation.

The model with the lowest AICc, and therefore the most parsimonious, was the mixed model, which contained all variables related to several different hypotheses (Table 3). All other models had considerably less support (ΔAICc > 10 and low values of Akaike weight, see Burnham & Anderson, [2004]). These results suggest that an interaction of factors related to the different hypotheses may be the best explanation for the variation on squamates species richness in the Cerrado.

DISCUSSION

Our results based on the multiple regression and the model selection suggest that several mechanisms related to the different diversity hypotheses might work together to explain richness variation in the Cerrado. Annual precipitation was the most important climatic variable that predicted Cerrado squamate richness, supporting the species-energy hypothesis. Several studies on different organisms and at different geographical scales have found that species richness was correlated with available energy (Currie *et al.*, 2004, Hawkins *et al.*, 2003). Results of our study join the body of literature that corroborates the species-energy hypothesis.

The species-energy hypothesis posits that the amount of energy available and energy partitioning among species limit species richness (Wright, 1983). A mechanism explaining how diversity gradients are produced based on energy is lacking but several hypotheses exist (Clarke & Gaston, 2006). Exciting recent studies suggest that variation in temperature results in variation in DNA substitution rates and may influence rates of evolution (Wright *et al.*, 2003; Allen & Gillooly, 2006; Allen *et al.*, 2006). Although this mechanism may be operating on a global scale, whether it is important at smaller geographical scales such as the Cerrado region remains undetermined.

The environmental stability hypothesis posits that physiological tolerances of species to variable environments may limit species richness, leading to higher diversity in more stable environments (Pianka, 1966; Currie, 1991). Although the rationale behind the hypothesis seems reasonable, recent studies failed to support this hypothesis (Velho *et al.*, 2004; Mora & Robertson, 2005). We found supporting evidence for the environmental stability hypothesis in that precipitation seasonality was the second best climatic variable in predicting squamate richness. The Cerrado is a highly seasonal environment with two well-defined seasons, one dry and cold, and the other wet and warm. This seasonality may influence variation in species richness because some species may not be able to physiologically tolerate the harsh dry season, and/or areas with less severe conditions may be colonized by species that primarily occur in other biomes, such as the Amazon forest.

Altitude was important in predicting Cerrado squamate species richness. Correlations of species richness with altitude suggest that the environmental heterogeneity hypothesis, which posits that habitat differentiation and resource partitioning facilitate coexistence and enhance species richness (Richerson & Lum, 1980), may play a role in large-scale patterns of species richness. Relationships of species richness to topographic and environmental heterogeneity have been shown previously, especially at smaller spatial scales (Bohning-Gaese, 1997; Cleary *et al.*, 2005; Triantis *et al.*, 2005). In addition, some studies supporting the species-energy hypothesis have also found a significant effect of topography or landscape heterogeneity as in South American birds (Diniz-Filho & Bini, 2005) and plants (Lavers & Field, 2006).

Altitude may have had an historical influence on the distribution and composition of Cerrado fauna and flora (da Silva, 1996; Meio *et al.*, 2003; Colli, 2005). In birds Amazonian elements are restricted to gallery forests at lower altitude, whereas Atlantic elements are found in higher altitudes (da Silva, 1996). The same pattern is repeated for species of trees and shrubs (Meio *et al.*, 2003). A crucial event responsible for current patterns of species distribution of Cerrado squamates was the uplift of the Central Brazilian Plateau in the Tertiary (Colli, 2005), again providing evidence for the importance of altitude and topography.

The above hypotheses are not mutually exclusive and, based on results of the model selection, which pointed out that the best model to explain species richness is a combination of the different hypothesis, we advance the idea that partitioning of energy among species, habitat differentiation, and tolerance to variable environments are the primary ecological factors determining variation in squamate richness across Cerrado. Others studies in the literature also suggested that a combination of different hypotheses may best explain species richness patterns (Bohning-Gaese, 1997; Hurlbert, 2004; Diniz-Filho & Bini, 2005). Additional studies that examine richness at different spatial scales are necessary, as well as studies that explore other factors known to affect species richness, such as historical factors (e.g. geographical barrier and/or effect of regional pool of species) and ecological interactions such as competition and predation. The Cerrado is a highly heterogeneous landscape, composed of a mosaic of edaphic conditions and different vegetation types, and it would be constructive to investigate how finer scale environmental and landscape variables affect species richness. For instance, Nogueira (2005) studied a lizard assemblage in the central Cerrado and found that richness is higher in the open landscape when compared with forested habitats.

Deforestation of Cerrado has accelerated during the past 20 years, and landcover has shifted greatly toward planted pastures and agriculture (Klink & Moreira, 2002; Klink & Machado, 2005). The central portion of Cerrado and areas of the southern portion of Góias state (GO) and northern portion of Mato Grosso do Sul state (MS), all of which coincide with large patches of high squamate richness (Figure 3), are of particular concern because these areas have been largely converted into pasture and agriculture (Silva *et al.*, 2006). Due to its fragmentation, maintenance and protection of current reserves, establishment of ecological corridors among existing reserves, and creation of new reserves are urgently needed to ensure the conservation of squamate richness in these regions.

The northern portion of Cerrado also exhibited high richness of squamate species. Luckily, most of the remaining pristine native Cerrado areas are located in this region. Recent infrastructure development is changing this situation, and large portions of native vegetation are being replaced by soybean plantations. However, these areas remain poorly sampled (Figure 2A). Lack of sampling is a major issue in the Neotropics and particularly in the Cerrado. Bini et al. (2006) modeled range distributions of Cerrado amphibian species based on habitat suitability and number of inventories, and reached conclusions similar to this study regarding the importance for conservation of the northern areas of the Cerrado. Consequently, priorities for this region should begin with biological surveys, followed by the creation of new reserves where deemed necessary.

Overall, the level of deforestation and threats to Cerrado biodiversity are so alarming that any action toward conservation is important. Because both time and monetary resources are limited, decisions often must be based upon available data. Thus it is crucial to maximize utility of these data. However, results of species niche modeling studies cannot be considered unequivocally true, and repeated verification based on fieldwork is necessary to ensure that proper decisions are being made. Biodiversity data, in the form of compilations of revised point-localities, are crucial for interpreting richness patterns and are highly informative to conservation. In addition, niche-based models cannot account for factors that may have limited species distributions historically, such as geographical barriers, resulting in speciation events. In Madagascar, niche models predicted occurrence of known species, but fieldwork revealed that closely related and ecologically similar undescribed species lived in the predicted areas (Raxworthy *et al.*, 2003).

To complement and add utility of the information provided by our study, additional studies on the geographic patterns of endemism in Cerrado species and the temporal and spatial pattern of deforestation are needed. Analyzing patterns of species richness and endemism and correlating them with patterns of deforestation and human occupation can help define conservation strategies and minimize conflicts between development and conservation (Whittaker *et al.*, 2005; Diniz-Filho *et al.*, 2006). Our results are an initial step toward the development of a large comprehensive dataset on richness, endemism, and patterns of deforestation that will provide the necessary information to support conservation decisions for the Cerrado.

ACKNOWLEDGMENTS

We thank C. Wolfe, D. Shepard, and two anonymous reviewers for comments and suggestions on earlier versions of the manuscript. This work was submitted in partial fulfillment of the PhD degree at the University of Oklahoma to GCC. We thank T. Rashed, L. Vitt, J. Caldwell, J. Kelly, and M. Kaspari for serving on GCC doctoral committee and providing suggestions on the manuscript. We also thank Thiago F.L.V.B. Rangel for advice on how to use the PCNM spatial filters in SAM, and Ingo Schlupp for advice on the model selection approach. This work was supported by a CAPES/Fulbright PhD fellowship to GCC. FAPESP and Conservation International provided funds to CN for the compilation of Cerrado Squamate Reptiles point locality database. We thank the curators and staff of CHUNB, MZUSP, and IB for allowing the examination of specimens housed at these institutions.

REFERENCES

Ab'Saber, A.N. (1977) Os domínios morfoclimáticos da América do Sul. Primeira aproximação. *Geomorfologia*, **52**, 1-21.

Aguiar, L. (2000) Comunidades de morcegos do Cerrado no Brasil Central. Tese de Doutorado. Brasília, DF. Universidade de Brasília.

Allen, A.P. & Gillooly, J.F. (2006) Assessing latitudinal gradients in speciation rates and biodiversity at the global scale. *Ecology Letters*, **9**, 947-954.

Allen, A.P., Gillooly, J.F., Savages, V.M. & Brown, J.H. (2006) Kinetic effects of temperature on rates of genetic divergence and speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 9130-9135.

Anderson, D. R., Burnham, K. P. & Thompson, W. L. (2000) Null hypothesis testing: Problems, prevalence, and an alternative. *Journal of Wildlife Management.* **64**, 912-923.

Anderson, R. P., Lew, D. & Peterson, A. T. (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling.* **162**, 211-232.

Araujo, M. B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography.* **33**, 1677-1688.

Araujo, M. B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution.* **22**, 42-47.

Bini, L. M., Diniz-Filho, J. A. F., Rangel, T., Bastos, R. P. & Pinto, M. P. (2006) Challenging Wallacean and Linnean shortfalls: knowledge gradients and conservation planning in a biodiversity hotspot. *Diversity and Distributions.* **12**, 475-482.

Bohning-Gaese, K. (1997) Determinants of avian species richness at different spatial scales. *Journal of Biogeography*, **24**, 49-60.

Borcard, D. & Legendre, P. (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51-68.

Borcard, D., Legendre, P., Avois-Jacquet, C. & Tuomisto, H. (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology*, **85**, 1826-1832.

Burnham, K. P. & Anderson, D. R. (2004) Multimodel inference - understanding AIC and BIC in model selection. *Sociological Methods & Research.* **33**, 261-304.

Clarke, A. & Gaston, K.J. (2006) Climate, energy and diversity. *Proceedings of the Royal Society B-Biological Sciences*, **273**, 2257-2266.

Cleary, D.F.R., Genner, M.J., Boyle, T.J.B., Setyawati, T., Angraeti, C.D. & Menken, S.B.J. (2005) Associations of bird species richness and community composition with local and landscape-scale environmental factors in Borneo. *Landscape Ecology*, **20**, 989-1001.

Colli, G.R. (2005) As origens e a diversificação da herpetofauna do Cerrado. *Biodiversidade, Ecologia e Conservação do Cerrado* (ed. by A. Scariot, J.C. Souza-Silva and J.M. Felfili), pp. Brasília, DF.

Currie, D.J. (1991) Energy and Large-Scale Patterns of Animal-Species and Plant-Species Richness. *American Naturalist*, **137**, 27-49.

Currie, D. J., Mittelbach, G. G., Cornell, H. V., Field, R., Guegan, J. F., Hawkins, B. A., Kaufman, D. M., Kerr, J. T., Oberdorff, T., O'Brien, E. & Turner, J. R. G. (2004) Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters.* **7**, 1121-1134.

Diniz-Filho, J.A.F. & Bini, L.M. (2005) Modelling geographical patterns in species richness using eigenvector-based spatial filters. *Global Ecology and Biogeography*, **14**, 177-185.

Diniz-Filho, J.A.F., Bini, L.M. & Hawkins, B.A. (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, **12**, 53-64.

Diniz-Filho, J.A.F., Bini, L.M., Pinto, M.P., Rangel, T., Carvalho, P. & Bastos, R.P. (2006) Anuran species richness, complementarity and conservation conflicts in Brazilian Cerrado. *Acta Oecologica-International Journal of Ecology*, **29**, 9-15.

Eastman, J. (2003) IDRISI Kilimanjaro – Guide to GIS and Image Processing. Clark Labs – Clark University. Worcester, MA.

Eiten, G. (1972) The Cerrado vegetation of Brazil. *The Botanical Review*, **38**, 201-341.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151.

Felfili, J.M., Junior, M.C.D., Sevilha, A.C., Fagg, C.W., Walter, B.M.T., Nogueira, P.E. & Rezende, A.V. (2004) Diversity, floristic and structural patterns of cerrado vegetation in Central Brazil. *Plant Ecology*, **175**, 37-46.

Fraser, R.H. & Currie, D.J. (1996) The species richness-energy hypothesis in a system where historical factors are thought to prevail: Coral reefs. *American Naturalist*, **148**, 138-159.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters.* **8**, 993-1009.

Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J., Guegan, J. F., Kaufman, D. M., Kerr, J. T., Mittelbach, G. G., Oberdorff, T., O'Brien, E. M., Porter, E. E. &

Turner, J. R. G. (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*. **84**, 3105-3117.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Hurlbert, A.H. (2004) Species-energy relationships and habitat complexity in bird communities. *Ecology Letters*, **7**, 714-720.

Klink, C.A. & Machado, R.B. (2005) Conservation of the Brazilian Cerrado. *Conservation Biology*, **19**, 707-713.

Klink, C.A. & Moreira, A.G. (2002) Past and Current Human Occupation, and Land Use. *The Cerrados of Brazil: Ecology and Natural History of a Neotropical Savanna* (ed. by P.S. Oliveira and R.J. Marquis), pp. 69-88. Columbia University Press, New York, NY.

Lamoreux, J.F., Morrison, J.C., Ricketts, T.H., Olson, D.M., Dinerstein, E., McKnight, M.W. & Shugart, H.H. (2006) Global tests of biodiversity concordance and the importance of endemism. *Nature*, **440**, 212-214.

Lavers, C. & Field, R. (2006) A resource-based conceptual model of plant diversity that reassesses causality in the productivity-diversity relationship. *Global Ecology and Biogeography*, **15**, 213-224.

Legendre, P. (1993) Spatial Autocorrelation - Trouble or New Paradigm. *Ecology*, **74**, 1659-1673.

Luoto, M., Kuussaari, M. & Toivonen, T. (2002) Modelling butterfly distribution based on remote sensing data. *Journal of Biogeography*, **29**, 1027-1037.

Machado, R., Aguiar, L., Ramos Neto, M., Rodrigues, F., Hass, A. & Aquino, F. (2004a) *Atlas das unidades de conservação federais do Brasil*, edn. Metalivros, São Paulo.

Machado, R.B., Ramos Neto, M.B., Pereira, P.G.P., Caldas, E.F., Gonçalves, D.A., Santos, N.S., Tabor, K. & Steininger, M. (2004b) Estimativas de perda da área do Cerrado brasileiro. In Brasilia, DF: Conservation International.

Meio, B.B., Freitas, C.V., Jatoba, L., Silva, M.E.F., Ribeiro, J.F. & Henriques, R.P.B. (2003) Influência da flora da floresta Amazônica e Atlântica na vegetação do cerrado sensu stricto. *Revista Brasileira de Botânica*, **26**, 437-444.

Mittermeier, R.A., Gil, P.R., Hoffman, M., Pilgrim, J., Brooks, T., G, M.C., J, L. & Fonseca, G.A.B. (2005) *Hotspots Revisited : Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions*, Second edn. Conservation International.

Mora, C. & Robertson, D.R. (2005) Causes of latitudinal gradients in species richness: A test with fishes of the Tropical Eastern Pacific. *Ecology*, **86**, 1771-1782.

Moser, D., Dullinger, S., Englisch, T., Niklfeld, H., Plutzar, C., Sauberer, N., Zechmeister, H.G. & Grabherr, G. (2005) Environmental determinants of vascular plant species richness in the Austrian Alps. *Journal of Biogeography*, **32**, 1117-1127.

Myers, N. (2003) Biodiversity hotspots revisited. *Bioscience*, **53**, 916-917.

NGA (2005) Name Files of Selected Countries - Geonet names server. In: available at : http://earth-info.nga.mil/gns/html/cntry_files.html.

Nogueira, C., Valdujo, P.H. & Franca, F.G.R. (2005) Habitat variation and lizard diversity in a Cerrado area of Central Brazil. *Studies on Neotropical Fauna and Environment*, **40**, 105-112.

Oliveira, P.S. & Marquis, R.J. (2002) *The Cerrados of Brazil: Ecology and Natural History of a Neotropical Savanna*, edn. Columbia University Press, New York, NY.

Peterson, A.T. (2001) Predicting species' geographic distributions based on ecological niche modeling. *Condor*, **103**, 599-605.

Pianka, E.R. (1966) Latitudinal gradients in species diversity: a review of concepts. *The American Naturalist*, **100**, 33-46.

Rangel, T.F.L.V.B., Diniz-Filho, J.A.F. & Bini, L.M. (2006) Towards an integrated computational tool for spatial analysis in macroecology and biogeography. *Global Ecology and Biogeography*, **15**, 321-327.

Ratter, J., Ribeiro, J. & Bridgewater, S. (1997) The Brazilian Cerrado vegetation and threats to its biodiversity. *Annals of Botany*, **80**, 223-230.

Raxworthy, C.J., Martinez-Meyer, E., Horning, N., Nussbaum, R.A., Schneider, G.E., Ortega-Huerta, M.A. & Peterson, A.T. (2003) Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, **426**, 837-841.

Richerson, P.J. & Lum, K. (1980) Patterns of Plant-Species Diversity in California - Relation to Weather and Topography. *American Naturalist*, **116**, 504-536.

Ricklefs, R.E. (2004) A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, **7**, 1-15.

Rylands, A.B., Fonseca, M., Machado, R. & Cavalcanti, R. (2005) Brazil. *The state of the world's protected areas* (ed. by C.S. Spalding and M. Jenkins). United Nations Environment Programme (UNEP) and World Conservation Monitoring Center (WCMC), Cambridge, UK.

Silva, J.F., Farinas, M.R., Felfili, J.M. & Klink, C.A. (2006) Spatial heterogeneity, land use and conservation in the cerrado region of Brazil. *Journal of Biogeography*, **33**, 536-548.

Silva, J.M.C. (1995) Birds of the Cerrado Region, South America. *Steenstrupia*, **21**, 69-92.

Silva, J.M.C. (1996) Distribution of Amazonian and Atlantic birds in gallery forests of the Cerrado region, South America. *Ornitologia Neotropical*, **7**, 1-18.

Silva, J.M.C. & Bates, J.M. (2002) Biogeographic patterns and conservation in the South American Cerrado: A tropical savanna hotspot. *Bioscience*, **52**, 225-233.

Stockwell, D. & Peters, D. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, **13**, 143-158.

Tabachnick, B.G. & Fidell, L.S. (2000) *Using multivariate statistics*, 4th edn. Allyn and Bacon, Boston, MA.

Triantis, K.A., Mylonas, M., Weiser, M.D., Lika, K. & Vardinoyannis, K. (2005) Species richness, environmental heterogeneity and area: a case study based on land snails in Skyros archipelago (Aegean Sea, Greece). *Journal of Biogeography*, **32**, 1727-1735.

Velho, L.F.M., Bini, L.M. & Lansac-Toha, F.A. (2004) Testate amoeba (Rhizopoda) diversity in plankton of the Upper Parana River floodplain, Brazil. *Hydrobiologia*, **523**, 103-111.

Whittaker, R.J., Araujo, M.B., Paul, J., Ladle, R.J., Watson, J.E.M. & Willis, K.J. (2005) Conservation Biogeography: assessment and prospect. *Diversity and Distributions*, **11**, 3-23.

Wright, D.H. (1983) Species-Energy Theory - an Extension of Species-Area Theory. *Oikos*, **41**, 496-506.

Wright, S.D., Gray, R.D. & Gardner, R.C. (2003) Energy and the rate of evolution: Inferences from plant rDNA substitution rates in the western pacific. *Evolution*, **57**, 2893-2898.

Table 1. General hypotheses explaining species richness patterns, environmental characteristics traditionally used to test these hypotheses, and variables used in this paper in the regression models to access the importance of each hypothesis for Cerrado squamates. Adapted from: (Moser *et al.*, 2005).

Hypothesis	Argument	Factor used to test	In this paper
Available energy[*]	Partitioning of energy among species limits richness	Temperature, potential evapotranspiration, and precipitation	Annual Precipitation, Net Primary Productivity
Environmental stress[†]	Fewer species are physiologically equipped to tolerate harsh environments	Minimum values of temperature and potential evapotranspiration	Precipitation of Driest Quarter, Mean Temperature of Driest Quarter
Environmental favorableness[‡]	Better life conditions promote higher species numbers	Maximum values of temperature and potential evapotranspiration	Mean Temperature of Warmest Quarter, Precipitation of Warmest Quarter,

Environmental stability[§]	Fewer species are physiologically equipped to tolerate variable environments	Annual variation in temperature	Temperature Seasonality, Temperature Annual Range, Precipitation Seasonality
Environmental heterogeneity[¶]	Habitat differentiation and resource partitioning facilitate coexistence and enhance species richness	Topographic, spatial climatic, edaphic and land-use heterogeneity	Altitude, Terrain declivity

[*] (Wright, 1983)

[†] (Fraser & Currie, 1996)

[‡] (Pianka, 1966; Richerson & Lum, 1980)

[§] (Pianka, 1966; Fraser & Currie, 1996)

[¶] (Richerson & Lum, 1980)

Table 2. Partial regression coefficients of the multiple regression model (*b*), t statistics and associated P-values for species richness of Brazilian Cerrado squamates regressed against environmental variables. Spatial structure was accounted for in the multiple regression by adding 9 eigenvector filters produced with the method of Principal Coordinates of Neighbour Matrices (PCNM). Filters were omitted from table.

Variables	*b*	T	P
Annual precipitation	-0.04	-5.08	<0.01
Precipitation seasonality	-1.23	-4.75	<0.01
Altitude	0.03	3.28	<0.01
Precipitation of driest quarter	-0.28	-2.93	<0.01
Net primary productivity	23.6	2.52	0.01
Mean temperature of driest quarter	0.92	2.39	0.02
Temperature annual range	0.21	2.04	0.04
Precipitation of warmest quarter	0.02	1.87	0.06
Terrain declivity	0.11	1.32	0.20
Mean temperature of warmest quarter	-0.20	-0.45	0.65
Temperature seasonality	0.004	0.425	0.67

Table 3. Summary of the model selection procedure. The model with the lowest AICc value is the most parsimonious one among the fitted models and is selected (marked in bold). ΔAICc is the difference in AICc to the selected model. *Wi* is the Akaike weight and it indicates the relative support a given model has when compared with the other models. K is the number of parameters of the model (no. of variables + intercept).

Model	AICc	ΔAICc	K	*Wi*
Available energy	1290.74	65.56	3	0.00
Environmental favorableness	1301.10	75.92	3	0.00
Environmental heterogeneity	1289.53	64.35	3	0.00
Environmental stability	1256.76	31.58	3	0.00
Environmental stress	1293.06	67.88	3	0.00
*** Mixed model**	**1225.18**	**0.00**	**12**	**1.00**
† Only significant	1262.44	37.26	8	0.00

* All 11 variables used in the multiple regression.

† Only the significant variables ($p < 0.05$) from the multiple regression.

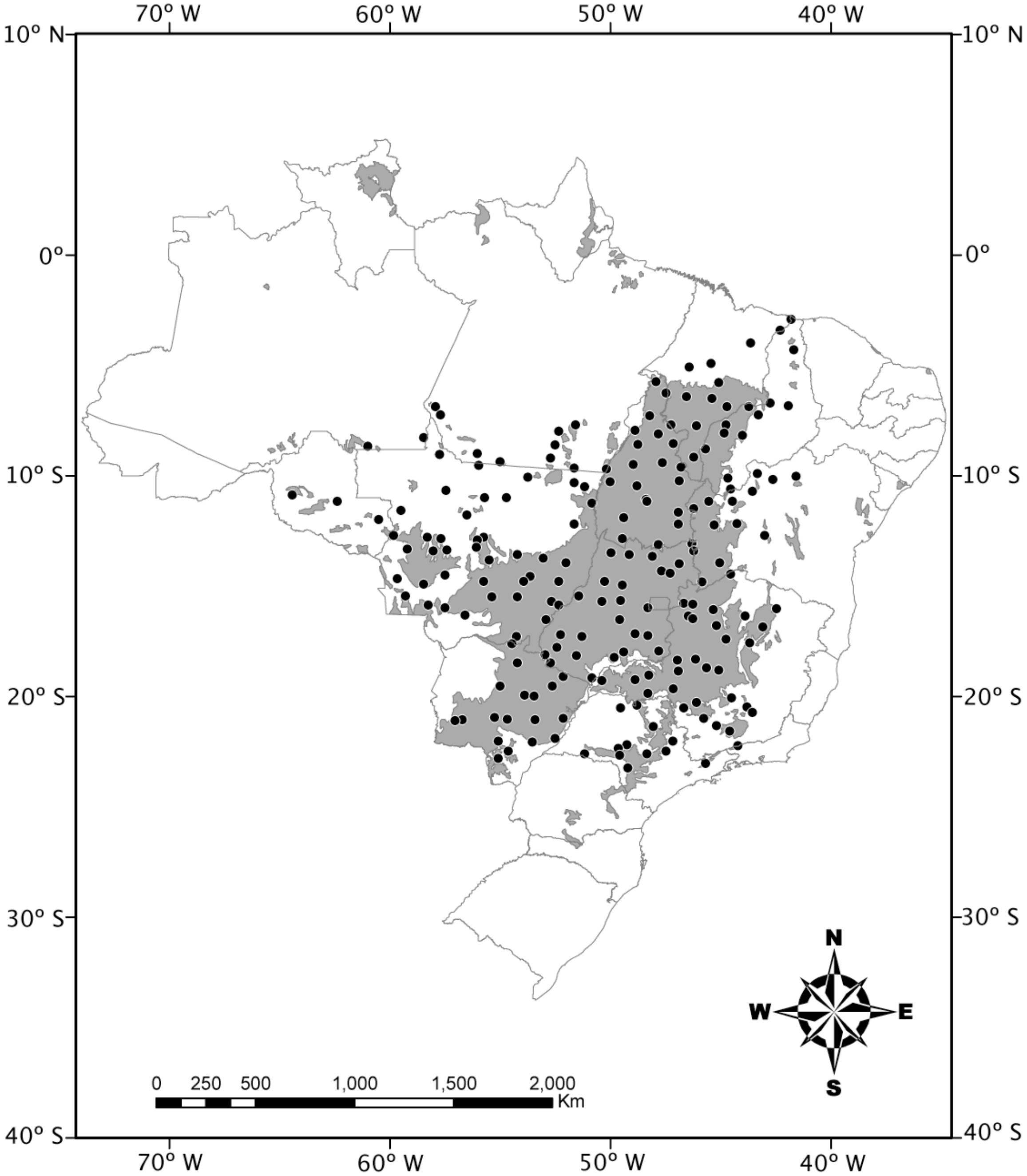
FIGURE LEGENDS

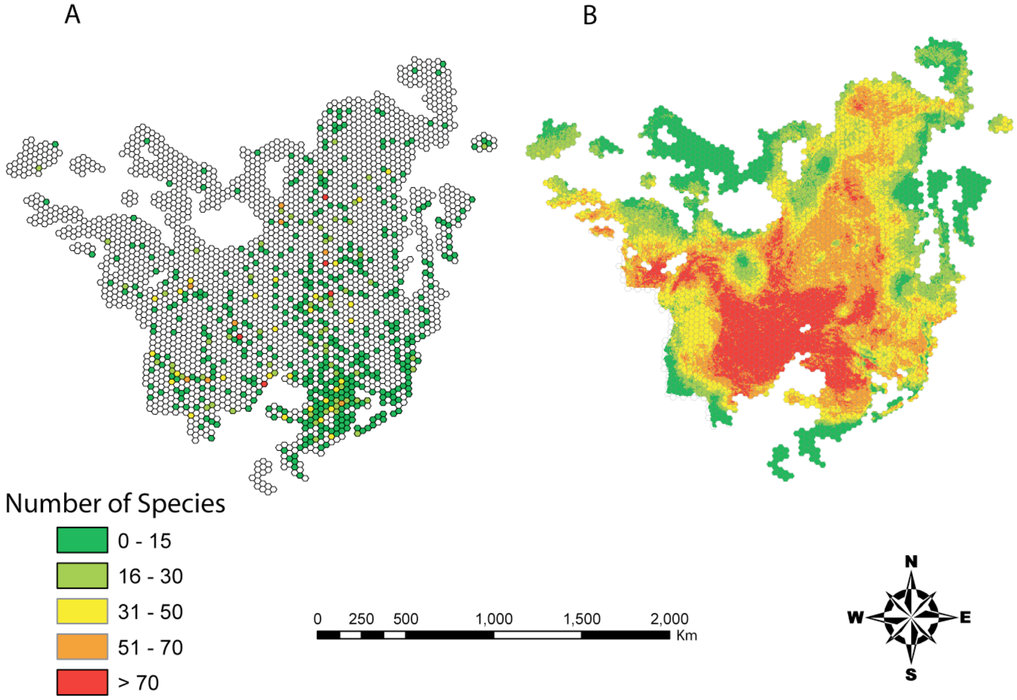
Figure 1 – Spatial distribution within Cerrado of points used in the multiple regression analysis. The area in gray corresponds to the Cerrado limits and transition areas with other biomes.

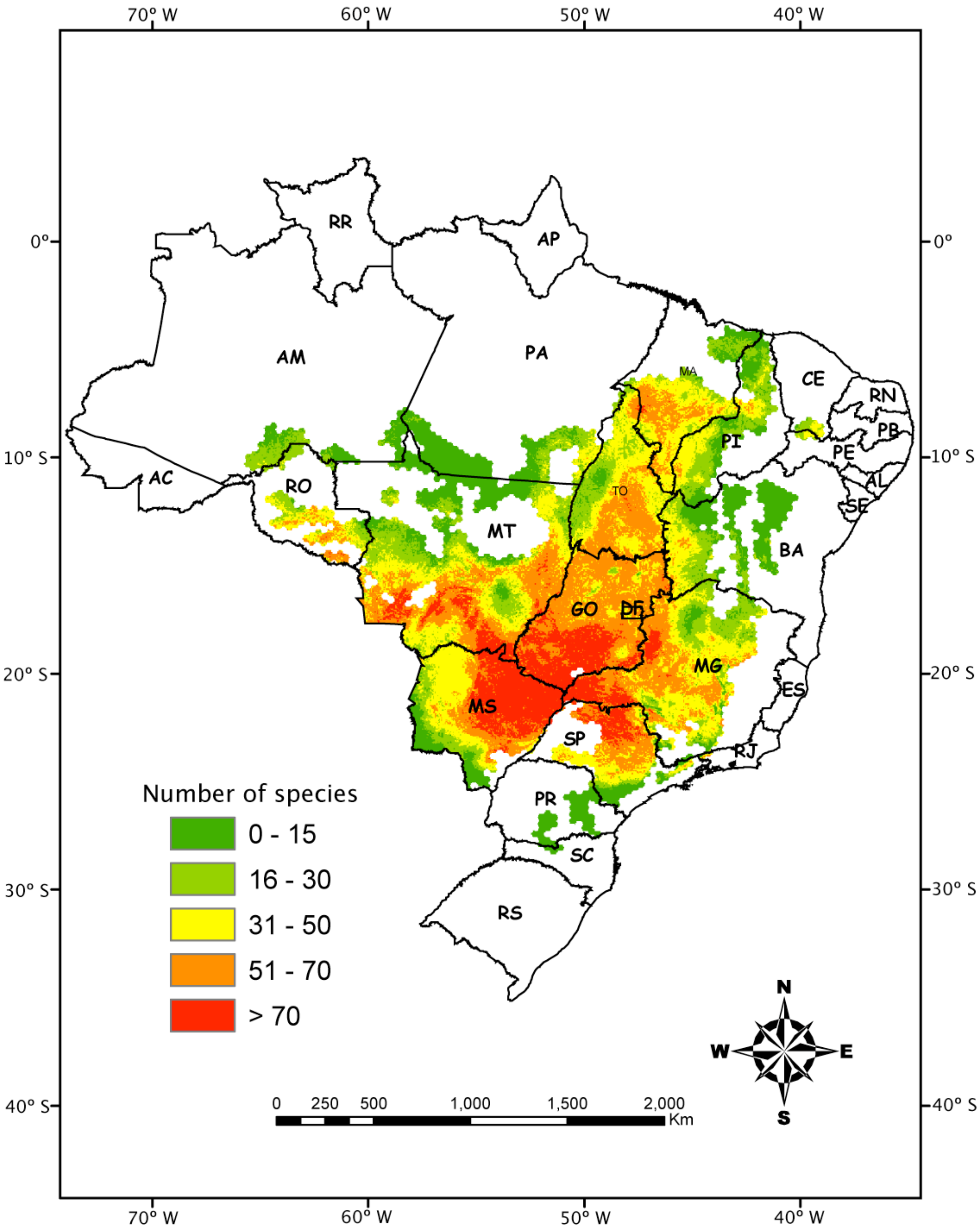
Figure 2 – Number of species per unit area (cell) based on museum collections. The grid corresponds to the approximate area of the Cerrado biome. The area of each cell is 100,000 ha. (A) Number of species based on the raw data. Blank cells have no specimen based on the major Cerrado collections; (B) Number of species based on the niche modeling of 237 species.

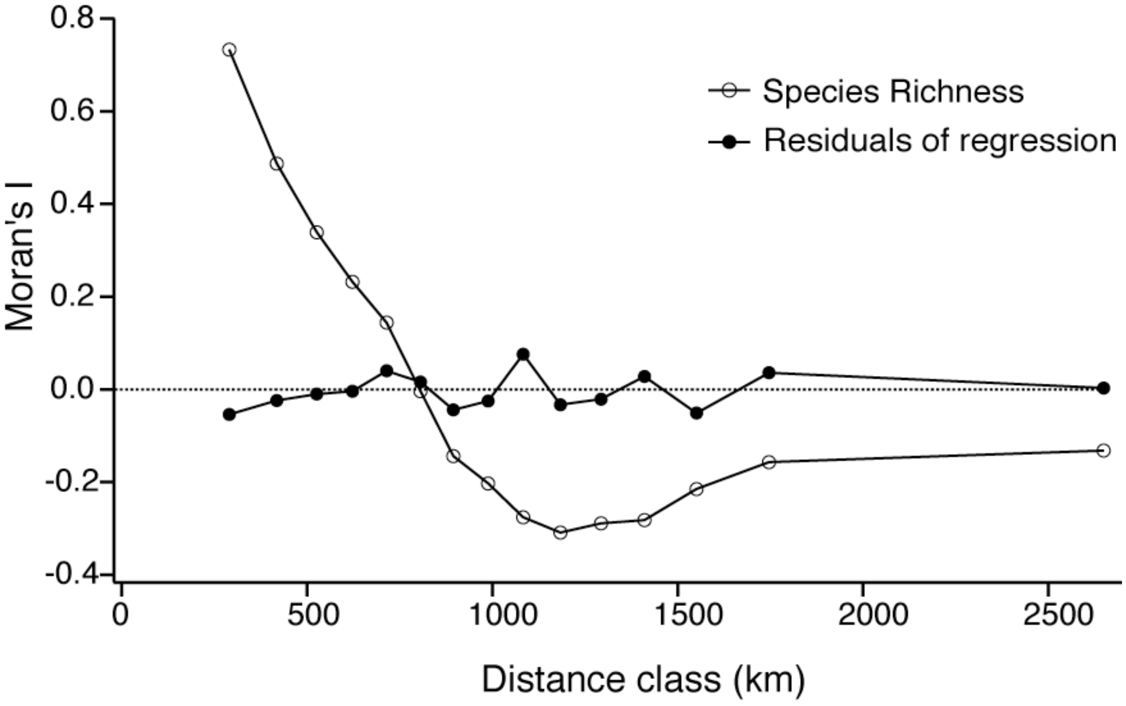
Figure 3 – Raster grid of Cerrado squamate species richness based on the sum of 237 individual species maps.

Figure 4 – Moran's index correlogram for squamate species richness and the residuals of multiple regression with environmental predictors.


70° W
60° W
50° W
40° W
10° N
0°
10° S
20° S
30° S
40° S
N
W
E
S
0
250
500
1,000
1,500
2,000
Km


A
B
Number of Species
0 - 15
16 - 30
31 - 50
51 - 70
> 70
0
250
500
1,000
1,500
2,000
Km
N
W
E
S


70° W
60° W
50° W
40° W
0°
10° S
20° S
30° S
40° S
RR
AP
AM
PA
MA
CE
RN
PB
PE
AL
SE
PI
AC
RO
TO
MT
BA
GO
DF
MG
ES
MS
SP
RJ
PR
SC
RS
Number of species
0 - 15
16 - 30
31 - 50
51 - 70
> 70
N
W
E
S
0
250
500
1,000
1,500
2,000
Km



0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
Moran's I
0
500
1000
1500
2000
2500
Distance class (km)
Species Richness
Residuals of regression

Appendix A. List of the 237 species used in the analysis.

Ameiva ameiva, Amphisbaena alba, A. anaemariae, A. camura, A. crisae, A. fuliginosa, A. leeseri, A. mensae, A. mertensi, A. miringoera, A. neglecta, A. sanctaeritae, A. silvestrii, A. talisiae, A. vermicularis, Anilius scytale, Anisolepis grillii, Anolis chrysolepis , Anolis meridionalis, Anops bilabialatus, A. albicollaris, Apostolepis ammodytes, A. assimilis, A. cearensis, A. cerradoensis, A. cf. longicaudata, A. christineae, A. dimidiata, A. flavotorquata, A. goiasensis, A. intermedia, A. lineata, A. longicaudata, A. nigroterminata, A. sp., *A. polylepis, A. vittata, Atractus albuquerquei, A. latifrons, A. pantostictus, Bachia bresslaui, B. cacerensis, Boa constrictor, Boiruna maculata, Bothrops alternatus, B. itapetiningae, B. lutzi, B. mattogrossensis, B. moojeni, B. neuwiedi, B. pauloensis, Bothrops* sp., *Briba brasiliana, Bronia bedai, B. kraoh, B. saxosa, Cercolophia roberti, C. steindachneri, Cercosaura albostrigatus, C. ocellata, C. parkeri, C. schreibersii, Chironius bicarinatus, C. exoletus, C. flavolineatus, C. laurenti, C. quadricarinatus, C. scurrulus, C. laevicollis, C. multiventris, Clelia bicolor, C. clelia, C. plumbea, C. quimi, C. rustica, Cnemidophorus aff. parecis, C. mumbuca, C. gr. ocellifer, C. parecis, Coleodactylus brachystoma, C. meridionalis, Colobosaura modesta, Corallus hortulanus, Crotalus durissus, Dipsas indica, Dracaena paraguayensis, Drymarchon corais, Drymoluber brazili, Taeniophallus occipitalis, Enyalius bilineatus, E. brasiliensis, E. catenatus, Epicrates cenchria, Erythrolamprus aesculapii, Eunectes murinus, E. notaeus, Gomesophis brasiliensis, Gonatodes humeralis, Gymnodactylus amarali, G. guttulatus, Helicops angulatus, H. gomesi, H. infrataeniatus, H. leopardinus, H. modestus, H. polylepis, H. trivittatus, Hemidactylus mabouya, Heterodactylus lundii, Hoplocercus spinosus, Hydrodynastes bicinctus, H. gigas, Iguana iguana, Imantodes cenchoa, Kentropyx aff. paulensis, K. calcarata, K. paulensis, K. vanzoi, K. viridistriga, Leposternon infraorbitale, L. microcephalum, L. polystegum, Leptodeira annulata, Leptophis ahaetulla, Leptotyphlops brasiliensis, L. dimidiatus, L. koppesi, Liophis almadensis, L. dilepis, L. frenatus, L. jaegeri, L. maryellenae, L. meridionalis, L. miliaris, L. paucidens, L. poecilogyrus, L. reginae, L. taeniogaster, L. typhlus, Liotyphlops beui, L. schubarti, L. ternetzii, Lygodactylus wetzeli, Lystrophis matogrossensis, L. nattereri, Mabuya dorsivittata, M. frenata, M. guaporicola, M. heathi, M. nigropunctata, Mastigodryas bifossatus, M. boddaerti, Micrablepharus atticolus, M. maximiliani, Micrurus brasiliensis, M. aff. ibiboboca, M. frontalis, M. lemniscatus, M. spixii, M. surinamensis, M. tricolor, Ophiodes fragilis, O. "striatus", Ophiodes sp 1, Ophiodes sp 2, Oxybelis aeneus, O. fulgidus, Oxyrhopus guibei, O. petola, O. rhombifer, O. trigeminus, Phalotris concolor, P. labiomaculatus, P. lativittatus, P. mertensi, P. multipunctatus, P. nasutus, P. tricolor, Philodryas aestiva, P. livida, P. mattogrossensis, P. nattereri, P. olfersii, P. patagoniensis, P. psammophidea, P. viridissima, Phimophis guerini, P iglesiasi, Phyllopezus pollicaris, Polychrus acutirostris, P. marmoratus, Psedablabes agassizii, Pseudoboa coronata, P. neuwiedii, P. nigra, Pseudoeryx plicatilis, Pseustes sulphureus, Psomophis genimaculatus, P. joberti, Rhachidelus brazili, Sibynomorphus mikanii, S. turgidus, S. ventrimaculatus, Simophis rhinostoma, S. leucocephalus, S. longicaudatus, S. worontzowi, Spilotes pullatus, Stenocercus aff. dumerilii, Stenocercus aff. tricristatus, Stenocercus caducus, S. dumerilii, Tantilla boipiranga, T. melanocephala, Teius teyou, Thamnodynastes chaquensis, T. hypoconia, T. rutilus, Tropidurus etheridgei, T. guarani, T. hispidus, T. insulanus, T. itambere, T. montanus, T.*

oreadicus, T. semitaeniatus, T. torquatus, Tupinambis duseni, T. merianae, T. quadrilineatus, T. teguixin, Typhlops brongersmianus, Vanzosaura rubricauda, Waglerophis merremii, Xenodon rhabdocephalus, X. severus, Xenopholis undulatus.

Appendix B Correlation matrix of the variables selected for the multiple regression analysis. The variables were selected in a way to minimize the correlation among each other and to relate to different hypothesis to explain patterns of species richness.

					Variables						
Variables	ALT	BIO4	BIO7	BIO9	BIO10	BIO12	BIO15	BIO17	BIO18	DECL	NPP
ALT	–	0.44**	0.18*	-0.72**	-0.84**	-0.28**	0.19**	-0.10	0.37**	0.21**	0.22**
BIO4		–	0.40**	-0.86**	-0.56**	-0.40**	-0.38**	0.57**	0.60**	0.11	0.29**
BIO7			–	-0.49**	-0.38**	0.08	-0.17*	0.14*	0.52**	0.01	0.30**
BIO9				–	0.89**	0.24**	0.25**	-0.41**	-0.71**	-0.15*	-0.48**
BIO10					–	0.07	0.05	-0.18*	-0.61**	-0.17*	-0.52**
BIO12						–	-0.28**	0.16*	0.23**	-0.18**	0.44**
BIO15							–	-0.90**	-0.37**	0.01	-0.28**
BIO17								–	0.43**	0.00	0.34**
BIO18									–	-0.09	0.55**
DECL										–	0.05
NPP											–

** Correlation is significant at the 0.01 level.

* Correlation is significant at the 0.05 level

Chapter II

Sampling bias and the use of ecological niche modeling in conservation planning: A field evaluation in a biodiversity hotspot

(formatted for *Conservation Letters*)

Abstract

Ecological niche modeling (ENM) has become an important tool in conservation biology. Despite its recent success, several basic issues related to the performance of the algorithms are still being debated. We assess the ability of two of the most popular algorithms, GARP and Maxent to predict distributions when sampling is geographically biased. We use an extensive data set collected in the Brazilian Cerrado, a biodiversity hotspot in South America. We found that both algorithms underestimate species richness from a study site far removed from the region for which we have the best sampling data. We also found that Maxent tends to be more sensitive to sampling bias than GARP. However, Maxent performs better when sampling is poor (e.g., low number of data points). Our results demonstrate that extreme care should be applied when examining outputs from ENM for conservation planning and decision-making. This is particularly critical when dealing with regions where sampling is geographically biased or poor.

Introduction

Sound conservation strategies depend heavily on biodiversity information, especially species distributions. However, knowledge about biodiversity remains inadequate, particularly in the highly speciose Tropics, where many species remain formally undescribed (Linnean shortfall) and poorly understood in terms of their geographical distribution (Wallacean shortfall) (Lomolino 2004; Whittaker *et al.* 2005). As a result, biodiversity databases, although extremely useful, may suffer strong limitations even for groups and/or regions that have been well studied (Soberón *et al.* 2000; Hortal *et al.* 2007; Soberón *et al.* 2007). Recently a new methodological approach, ecological niche modeling (ENM), has emerged as a powerful tool to reconstruct or predict species distributions. The method uses geo-referenced known occurrence points of the species under study that are linked with abiotic and/or biotic variables from each point locality. A particular algorithm processes information and then a predicted 'niche' envelope in which the species is likely to occur is produced (see Elith *et al.* 2006 for a review on the methods).

ENM has been applied in conservation biology to identify species richness 'hotspots' (Garcia 2006; Costa *et al.* 2007), sample for rare species (Guisan *et al.* 2006), predict effects of climate change on species' distributions (Araújo & Rahbek 2006; Hijmans & Graham 2006), and assess potential invasion and proliferation of exotic species (Peterson & Vieglais 2001). Despite the recent growth and diversity of studies that apply ENM to address conservation and/or evolutionary questions, several basic issues related to the performance of the algorithms remains unsettled. Among the most important issues is how the accuracy of ENM is influenced by factors such as sample size

(Stockwell & Peterson 2002; Hernandez *et al.* 2006), spatial scale (Lassueur *et al.* 2006; Guisan *et al.* 2007a; Trivedi *et al.* 2008), the nature of the environmental data set (Parra *et al.* 2004; Peterson & Nakazawa 2008), species traits (Poyry *et al.* 2008), biotic interactions (Araújo & Luoto 2007; Heikkinen *et al.* 2007), and finally, which particular algorithm is being used (Segurado & Araújo 2004; Elith *et al.* 2006).

Another important issue is how ENM models are influenced by geographical bias in the sampling points used to train the models. For example, a previous study found that the frequency of plant observations near roads was greater than that expected from a spatially random distribution such that predictive maps based on near-road observations were less accurate than those based on observations corrected for roadside bias (Kadmon *et al.* 2004). On a larger spatial scale, Loiselle *et al.* (2008) found that although localities based on herbarium collections did not represent well the entire climatic gradient in which most species occur, this existing climatic bias however, did not greatly affect distribution predictions when compared with an unbiased data set. Therefore, determining how well ENM can reconstruct a species distribution providing only with a biased subset of the whole species range is a crucial matter to establish ENM utility as a conservation tool.

We use two of the most commonly used ENM algorithms (GARP and Maxent) to predict the distribution of squamate reptiles (lizards, snakes, and amphisbaenas) in the Brazilian Cerrado, one of the 34 world biodiversity hotspots (Myers 2003; Mittermeier *et al.* 2005), a region for which a strong sampling bias exists (Costa *et al.* 2007). We tested performance of these two methods by first predicting species richness and composition of an unsampled area of conservation interest using ENM, and then conducting field surveys to determine actual species richness and composition. The poorly sampled region lies near the northern edge of the Cerrado region. We identify both limitations and strengths of ENM as a tool in conservation planning and biodiversity studies.

Methods

Ecological Niche Modeling

We used GARP and Maxent to model the distributions of all known (at the time of analysis) squamate species occurring in the Cerrado, a total of 237 species based on an extensive existing database. We used only species for which at least one data point existed within the Cerrado biome. For species whose distributions spanned multiple biomes, we also included data points outside of the Cerrado, because characteristics of these points can help identify suitable regions for species occurrence within Cerrado. Locality data for each species were collected from museums, literature, and previous fieldwork (see Costa *et al.* 2007; Nogueira *et al.* In press for details). All specimen records were checked for accurate taxonomy and the most precise locality information, a critical need, as museum data can be error-prone. Locality data varied between 3–256 (mean = 35.58, standard deviation = 39.32) unique point localities per species. The dataset contains a clear geographical sampling bias; most records come from the central and southeastern portion of the Cerrado, where the majority of research institutions are located, and very few inventories have been made in the Northern parts of the Biome (Fig. 1).

We used the implementation of GARP provided by the software OpenModeller. The algorithm divides occurrence points into training and extrinsic test data. The extrinsic test dataset is divided evenly into true training data (for model rule development) and intrinsic test data (for model rule evaluation and refinement). Models are based on presence-only data, with absence information included via random sampling of 1250 pseudo-absence points from the set of pixels at which the individual species were not collected. The algorithm works in an iterative process of rule selection, testing, and incorporation or rejection. More details on algorithm function are provided by Stockwell & Noble (1992). We used the default parameters of the OpenModeller version of GARP with the best subset selection option (optimum models considering omission/commission relationships; see Anderson *et al.* 2003).

Maxent fits a probability distribution for species occurrence to the set of pixels across the region of interest. The algorithm is based on the principle that, given the appropriate constraints, the best explanation to unknown phenomena will maximize the entropy of the probability distribution. For ecological niche modeling, these constraints derive from the values of those pixels at which the species has been detected. More details on Maxent function are provided by Phillips *et al.* (2004) and Phillips *et al.* (2006). We used the default parameters for Maxent v.3.2.1, which were adjusted based on a recent comprehensive evaluation (Phillips & Dudik 2008). The output format for Maxent and GARP are raster grids with values ranging from 0-1 for Maxent and 0-100 for GARP. To transform the models into discrete presence or absence, selection of a threshold is necessary. We selected threshold values where sensitivity (proportion of true positive predictions vs. the number of actual positive sites) is equal to specificity (the proportion of true negative predictions vs. the number of actual negative sites). This approach maximizes agreement between observed and modeled distributions balancing the cost arising from an incorrect prediction against the benefit gained from correct prediction (Manel *et al.* 2001; Pearson *et al.* 2006). In addition, to evaluate effects of threshold selection on the ability of the models to predict species presence in our study site, we determined the largest predicted value (if any) that would ensure presence of the species that were collected (i.o.w. which minimum threshold value would be necessary to assure species presence in the study site). We then plotted these values against the number of training points to address the question of whether the choice of threshold is influenced by the number of training points.

For both GARP and Maxent we used environmental variables from the Worldclim project (Hijmans *et al.* 2005), which are available for download at http://www.worldclim.org. We constructed a correlation matrix among all variables and selected for the modeling only the variables that were not highly correlated ($r > 0.9$). After applying this criterion we used the following environmental variables: altitude, annual precipitation, isothermality, maximum temperature of warmest month, mean diurnal range, mean temperature of warmest quarter, mean temperature of wettest quarter, minimum temperature of coldest month, precipitation of coldest quarter, precipitation of driest month, precipitation of warmest quarter, precipitation of wettest month, precipitation seasonality, temperature annual range, and temperature seasonality. All variables were at 1 km resolution.

Model Evaluations

To statistically evaluate model performance we used the area under the curve (AUC) on receiver operating characteristic (ROC). ROC analysis is a method designed to evaluate the specificity (absence of commission error) and sensitivity (absence of omission error) of a diagnostic test (Zweig & Campbell 1993; Fielding & Bell 1997). The AUC provides a threshold-independent measure of model performance as compared with that of null expectations (Fielding & Bell 1997), and it is the most commonly used statistic to evaluate ENM performance (Elith *et al.* 2006; Guisan *et al.* 2007b; Peterson *et al.* 2007). When the AUC is 0.50, the model is performing no better than random. Higher AUC values indicates better model performance; and a perfect prediction would have the value 1.0 (Hanley & Mcneil 1982).

After constructing niche models and calculating the AUC statistics, we tested performance of ENM in predicting species diversity and distributions by surveying a remote and previously unsampled area. This allowed us to evaluate the effect of sampling bias on the ability of ENM to project distributions into unsampled regions, and to determine whether GARP and/or Maxent are differentially affected by sampling bias. Using this approach, several scenarios are possible. First, when sampling points are concentrated in a subset of the species range, ENM is (a) capable of predicting the species occurrence or (b) not able to predict the occurrence of the species in the unsampled region outside of the major concentration of sampling (Fig. 2a, b). Second, when the sampling points are more dispersed throughout a species' range, ENM is (c) capable of predicting the species occurrence or (d) not able to predict the species occurrence (Fig. 2c, d) in the area.

Study Area and Field Sampling

We chose a study site located within the northern portion of the Cerrado Biome in the "Parque Nacional da Chapada das Mesas" (PNCM -7º10'S, 47º9'W), a recently created 160,000 ha conservation unit in the Brazilian state of Maranhão (Fig. 1). This area is ideal for evaluating sampling bias in ENM because it is relatively undisturbed, poorly sampled, and a recent niche modeling exercise predicted high squamate species diversity (Costa *et al.* 2007).

We collected squamates from November 30th to December 17th 2007, using 48 arrays of pitfall traps and 24 arrays of funnel traps resulting in 5,184 trap days. Traps were divided among six sampling points, which were located inside PNCM and were chosen in order to sample the full range of landscape and vegetation cover variation within PNCM. Each array of pitfall traps consisted of four 35 l buckets arranged in a Y-shape (one at the center and one on each of the three ends). Buckets were 5 m from each other, and 50 cm high plastic fences (bottom edge buried) spanned the distance between buckets. The funnel trap arrays consisted of a single 5 m long, 50 cm high plastic fence with a pair of funnel traps at each end (one on each side). Arrays were spaced approximately 20 m apart. All traps were checked daily. All specimens collected were deposited at the Coleção Herpetológica da Universidade de Brasília (CHUNB). In addition to our trapping methods, we collected animals by hand, noose, or using a shotgun during haphazard searches of various habitats within PNCM. We also routinely drove roads both during the day and night for snakes in the process of crossing or that had been killed by vehicles. Road collecting is a common and effective survey method for snakes (Sullivan 1981).

To estimate species richness of the region based on our sampling, we produced a species accumulation curve using the software EstimateS v.8 (Colwell 2005). EstimateS randomizes the sampling order to generate smooth species accumulation curves and species richness estimators. Resulting values are numbers of species expected based on empirical data (Colwell *et al.* 2004). We used the Abundance-based Coverage Estimator (ACE) to estimate species richness based on the sampling (Chazdon *et al.* 1998; Chao *et al.* 2000), and performed 10,000 randomizations without replacement. In addition, we fitted our data to three different accumulation curve mathematical models, Clench, Logarithmic, and Exponential. Model fitting was performed using methods and software described by Díaz-Francés & Soberón (2005). The model providing the best fit can then be used to estimate the asymptote (i.e., total species richness) of the species accumulation curve.

Statistical Analysis

We used the statistical package R to perform a two-sample test for equality of proportions with continuity correction to determine whether a difference exists in the proportion of species successfully predicted between GARP and Maxent. We also developed a multiple logistic regression model to explore different factors that may influence the probability of GARP and Maxent to successfully predict species occurrence in our study site. The dependent variable was the prediction success (0 = fail, 1 = success), and our independent variables were: 1 – Nearest neighbor index, which is calculated based on the average distance of each point to its nearest point. Low values of the index indicate a distribution more clumped than expected by chance whereas high values indicate a more dispersed distribution; 2 – Number of locality points used in the modeling exercise; and 3 –

Distance from the nearest locality point to PNCM. We ran the regression with all species collected in our field survey (N = 48) and also using only the species that had more than 15 known locality points (N = 42).

For both methods, a species was considered present in PNCM if any pixel of the final predicted distribution map (see above for details on how we obtained the final presence/absence maps) for that species lied within the PNCM limits (Fig. 1). Because we cannot distinguish between species that do not occur in the region from species that do occur but were not collected because of sampling deficiency, we restrict our evaluations to only the species that we collected during the field survey. One species (*Amphisbaena* sp.) was removed from all analyses due to taxonomic uncertainties. Calculations of the nearest neighbor index, and distance to the nearest point, were performed in ArcGIS 9.2. The multiple logistic regression was performed in SAS 9.1.

Results

We collected a total of 49 species of squamates in PNCM (Table 1). Our accumulation curve analysis indicated that our sampling efforts were far from stabilizing and the true richness of squamates in the region may be over 70 species (Fig. 3). The ACE richness estimation was 74 species and the model that produced the best fit was the logarithmic. Usually, when this model produces the best fit it is because the sample area is too large and/or the taxa are poorly known (Soberón & Llorente 1993). Such results are well known for Neotropical squamates, which require long-term fieldwork in order for sampling to stabilize (Duellman 1978), often because of snake species that are rare or difficult to sample.

GARP predicted 59 species within PNCM; we collected 35. The method failed to predict the presence of 13 species that were collected in our survey. Maxent predicted 51 species within PNCM and, among those, we collected 22. Maxent failed to predict the presence of 26 species that we collected in our fieldwork (Table 1). In addition, the two-sample test for equality of proportions with continuity correction showed that the ratio between predicted and surveyed species was higher for GARP ($\chi^2 = 4.46$, $P = 0.03$). However, Maxent models had higher AUC values (GARP $\overline{x} = 0.78 \pm 0.12$, Maxent $\overline{x} = 0.91 \pm 0.04$; $F = 51.7$, $P < 0.01$, all AUC values are in Table 1).

GARP predicted 15 species we collected in the area that Maxent failed to predict. Of those 15 species, seven followed the pattern described in Figure 2a, where the known sampled localities were concentrated in the central part of the Cerrado. None or very few known localities were in the northern part of the Cerrado where PNCM is located. We illustrate three of those cases in Fig. 4a-c. The remaining eight species follow a pattern similar to the one described in Fig. 2c, where sampling is more spread throughout the Cerrado and PNCM was surrounded by known sampled localities. We illustrate three of those cases in Fig. 4d-f. For some species such as Fig. 4a-f, the prediction of GARP that includes PNCM is a narrow extension from the main predicted area for the species, which is located in Central Brazil. This kind of pattern is particularly relevant for the issue of how different ENM algorithms deal with sampling bias on the training points, and we will comment on that further in the discussion.

Maxent successfully predicted three species in PNCM that GARP failed to predict, two of those cases were species that had low numbers of known localities (e.g. *Leptotyphlops brasiliensis* Fig. 5b). The other species showed a pattern similar to Fig. 2a where the majority of points used to train the model are away from PNCM, however the highest concentration of points was not on the south central part of the Cerrado (*Bothrops lutzi* Fig. 5a). Both methods successfully predicted 20 species and failed to predict 10 species that we actually collected in the area. Some of the species that both methods failed to predict showed the pattern described in Figure 2b, where the model's predictions were concentrated near the known localities (e.g. *Thamnodynastes hypoconia* Fig 5f). The other species that both methods failed to predict were likely affected by low numbers of known localities available for the modeling (e.g. *Apostolepis polylepis* Fig 5c).

The list of species successfully predicted in PCNM by ENM would not change much with selection of different thresholds. For GARP, only with the selection of much lower thresholds the species in which the method failed would be included, and that would result in models with very large commission errors. For Maxent, a few species would be added to the list with the selection of slightly lower thresholds, but for the majority, a much lower threshold would be necessary (see Table 1, maximum values column). For GARP the largest predicted values were significantly related to the number of points used to train the models ($F = 24.6$, $P < 0.01$, $r^2 = 0.35$; Fig. 6), whereas for Maxent there was no relationship ($F = 0.1$, $P = 0.77$; Fig. 6). This result is a consequence of many zero predicted values in PNCM for GARP models produced by low number of training points (see Table 1 maximum values column).

The multiple logistic regression results show that GARP models were not significantly influenced by any of the variables in the regression model. The same result was found when species with low known locality points were removed from the analysis (Table 2). Maxent predictions were influenced by the distance to the nearest point. After eliminating the species with few known localities, predictions were still influenced only by the distance to the nearest point.

Discussion

Based on our accumulation curve analysis, both ENM methods underestimated species richness. Our richness estimator (ACE) and accumulation curve predicted that the richness for the region should be more than 70 species, which is still a rather conservative estimate considering other well-sampled Cerrado localities (Colli *et al.* 2002; França & Araújo 2007). Therefore, even if all species predicted by both ENM methods were collected in the region, ENM would have still underpredicted species richness. Although ENM has been successfully used for various conservation applications (Domínguez-Domínguez *et al.* 2006; García 2006; Pawar *et al.* 2007), our results indicate that predictions from ENM should be examined carefully when working with regions where sampling is geographically biased or low.

Although Maxent models produced higher AUC values, GARP models better predicted species richness and composition of our study area. Previous work has suggested that Maxent may be more sensitive to geographical bias in the training points (Peterson *et al.* 2007). Two alternative hypotheses may explain why GARP better predict species occurrence in PCNM despite having lower AUC models. First, Maxent models are better than GARP models but our approach of evaluating the presence of the species in a specific region does not characterize well the performance of the models in their entire distribution. Second, the AUC statistics does not provide the best possible evaluation of the models. A lot more data and analysis would be necessary to assess the first hypothesis. However, previous studies have provided support for the second (Raes & ter Steege 2007; Lobo *et al.* 2008). Some recent work shows that reliance on AUC as the only estimate of model success needs to be re-examined (Austin 2007). Either way, better methods to statistically evaluate ENM models are likely to be a major topic of future research (Raes & ter Steege 2007; Lobo *et al.* 2008; Peterson *et al.* 2008).

In some cases, even when most of the known locality points where far away from PNCM, GARP was able to correctly predict the occurrence of species. This ability of GARP may be desirable in different applications of ENM, including the discovery of new populations and/or species. For example, in Madagascar, field survey of areas with similar sampling characteristics lead researchers to the discovery of several undescribed species of chameleons (Raxworthy *et al.* 2003). In our system, we discovered no obvious undescribed closely-related species; nevertheless, future genetic studies may reveal hidden diversity because populations of some species appear to be separated by areas where environmental conditions are predicted to be unsuitable. Recent studies in other Cerrado areas have been revealing new Squamate species, including some with restricted ranges, and from poorly studied taxa. (Nogueira & Rodrigues 2006; Rodrigues *et al.* 2007; Rodrigues *et al.* 2008). Because these species show high endemism and restricted ranges, they have are of special concern for conservation. Modeling of closely related species may help to identify regions where these species occur.

The ability to project distributions in areas distant from known localities may also be useful in ecosystems such as the Cerrado in which species' range extensions of several hundred of kilometers are commonly recorded (e.g. Filho & Montigelli 2006; Freitas *et al.* 2007; Silveira 2007). This may also be important in other uses of ENM. For some applications of ENM in ecology and evolutionary biology, precisely reconstructing species' distributions is not expected or desired from ENM; rather ENM is used to estimate a map of the environmental space in which the species is likely to occur. In these cases, contrasting where the species is predicted to occur with where the species does occur can provide insights into interesting biogeographical or ecological factors shaping the species' distribution (Anderson *et al.* 2002; Costa *et al.* 2008). A method that is too sensitive to sampling bias will be less useful to address such questions.

The multiple logistic regression models showed no effect of the nearest neighbor index in the probability of GARP or Maxent to successfully predict species' distribution. However, the distance to the nearest point influenced Maxent. This suggests that as long as a known locality exists close to the region, the algorithm will successfully predict species' presence even if the distribution of points is clustered. We found that Maxent tended to produce better estimates of species' distributions than GARP when a low number of localities are used in modeling. Also, the largest predicted values at PCNM were significantly related to the number of points used to train the GARP models, whereas for Maxent there was no relationship. This result is in agreement with a recent study using geckos in Madagascar, which showed Maxent performing better than GARP when sample size was smaller than 10 points (Pearson *et al.* 2007).

The Cerrado is a global biodiversity "hotspot" as defined by species richness, endemism, and human threats (Myers *et al.* 2000; Mittermeier *et al.* 2005). The region is being destroyed at a high rate with 55% of its original vegetation already removed (Machado *et al.* 2004; Klink & Machado 2005). Given the urgency to conserve habitats and species, time to conduct adequate surveys of the entire region is not available. In such a scenario, ENM may prove to be a useful tool in conservation planning. However, our results indicate that relying only on maps provided by ENM may underestimate species diversity, especially if strong geographical biases exist in the dataset used to generate the models. ENM may be a useful tool to guide survey efforts but may not be sufficient to justify management decisions and the design of protected area systems. As in most of the Neotropical region most conservation opportunities lies in remote and generally poorly sampled regions, where data generated by ENM provide a useful first evaluation. However, for reliable conservation decisions ENM data must be followed by well-designed field inventories.

Acknowledgements

We thank D. Shepard and A. T. Peterson for comments on the manuscript. Fieldwork was funded by Conservation International – Brazil, and field support was provided by Pequi, a Brazilian non governmental organization. Work on PNCM was authorized by IBAMA permit # 13204-1. We thank, P. Valdujo, S. Balbino, R. Recoder, and R. Bosque for help during fieldwork. This study was submitted in partial fulfilment of G.C.C.'s PhD degree at the University of Oklahoma. The species locality database was assembled as part of doctoral studies conducted by CN, and supported by FAPESP. We thank J. Caldwell, M. Kaspari, J. Kelly, T. Rashed, and L. Vitt, for serving on G.C.C.'s doctoral committee and providing comments. GCC is supported by a Fulbright/CAPES PhD fellowship (15053155–2018/04-7). GRC by CNPq grant (# 302343/88–1). Portions of this research were supported by a National Science Foundation grant to Laurie J. Vitt and Janalee P. Caldwell (DEB-0415430).

References

Anderson R.P., Lew D. & Peterson A.T. (2003). Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecol. Model.*, 162, 211–232.

Anderson R.P., Peterson A.T. & Gómez-Laverde M. (2002). Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos*, 98, 3–16.

Araújo M.B. & Luoto M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecol. Biogeogr.*, 16, 743–753.

Araújo M.B. & Rahbek C. (2006). How does climate change affect biodiversity? *Science*, 313, 1396–1397.

Austin M.P. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecol. Model.*, 200, 1–19.

Chao A., Hwang W.H., Chen Y.C. & Kuo C.Y. (2000). Estimating the number of shared species in two communities. *Statistica Sinica*, 10, 227–246.

Chazdon R.L., Colwell R.K., Denslow J.S. & Guariguata M.R. (1998). Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of NE Costa Rica. In: *Forest Biodiversity Research, Monitoring and Modeling: Conceptual Background and Old World Case Studies* (eds. Dallmeier F & Comiskey JA). Parthenon Publishing Paris, pp. 285–309.

Colli G.R., Bastos R.P. & Araújo A.F.B. (2002). The character and dynamics of the Cerrado herpetofauna. In: *The Cerrados of Brazil: Ecology and Natural History of a Neotropical Savanna* (eds. Oliveira PS & Marquis RJ). Columbia University Press New York, NY.

Colwell R.K. (2005). EstimateS: Statistical estimation of species richness and shared species from samples. User's Guide and application published at: http://purl.oclc.org/estimates.

Colwell R.K., Mao C.X. & Chang J. (2004). Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology*, 85, 2717–2727.

Costa G.C., Nogueira C.C., Machado R.B. & Colli G.R. (2007). Squamate richness in the Brazilian Cerrado and its environmental–climatic associations. *Divers. Distrib.*, 13, 714–724.

Costa G.C., Wolfe C., Shepard D.B., Caldwell J.P. & Vitt L.J. (2008). Detecting the influence of climatic variables on species' distributions: a test using GIS niche-based models along a steep longitudinal environmental gradient. *J. Biogeogr.*, 35, 637–646.

Díaz-Francés E. & Soberón J. (2005). Statistical estimation and model selection of species-accumulation functions. *Conserv. Biol.*, 19, 569–573.

Domínguez-Domínguez O., Martinez-Meyer E., Zambrano L. & De Leon G.P.P. (2006). Using ecological-niche Modeling as a conservation tool for freshwater species: Live-bearing fishes in central Mexico. *Conserv. Biol.*, 20, 1730–1739.

Duellman W.E. (1978). The biology of an equatorial herpetofauna in Amazonian Ecuador. *Miscellaneous Publications of the University of Kansas, Museum of Natural History, Lawrence*, 1–352.

Elith J., Graham C.H., Anderson R.P., Dudik M., Ferrier S., Guisan A., Hijmans R.J., Huettmann F., Leathwick J.R., Lehmann A., Li J., Lohmann L.G., Loiselle B.A., Manion G., Moritz C., Nakamura M., Nakazawa Y., Overton J.M., Peterson A.T., Phillips S.J., Richardson K., Scachetti-Pereira R., Schapire R.E., Soberón J., Williams S., Wisz M.S. & Zimmermann N.E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129–151.

Fielding A.H. & Bell J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.*, 24, 38–49.

Filho G.A.P. & Montigelli G.G. (2006). Geographic Distribution: *Hydroginastes gigas*. *Herpetological Review*, 37, 497–497.

França F.G.R. & Araújo A.F.B. (2007). Are there co-occurence patterns that structure snake communities in central Brazil? *Braz. J. Biol.*, 67, 33–40.

Freitas M.A., Silva T.F.S. & Rodriguez M.T. (2007). Geographic Distribution: *Chironius quadrilineatus*. *Herpetological Review*, 38, 354–354.

García A. (2006). Using ecological niche modelling to identify diversity hotspots for the herpetofauna of Pacific lowlands and adjacent interior valleys of Mexico. *Biol. Conserv.*, 130, 25–46.

Guisan A., Broennimann O., Engler R., Vust M., Yoccoz N.G., Lehmann A. & Zimmermann N.E. (2006). Using niche-based models to improve the sampling of rare species. *Conserv. Biol.*, 20, 501–511.

Guisan A., Graham C.H., Elith J. & Huettmann F. (2007a). Sensitivity of predictive species distribution models to change in grain size. *Divers. Distrib.*, 13, 332–340.

Guisan A., Zimmermann N.E., Elith J., Graham C.H., Phillips S. & Peterson A.T. (2007b). What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecol. Monogr.*, 77, 615–630.

Hanley J.A. & Mcneil B.J. (1982). The Meaning and Use of the Area under a Receiver Operating Characteristic (Roc) Curve. *Radiology*, 143, 29–36.

Heikkinen R.K., Luoto M., Virkkala R., Pearson R.G. & Korber J.H. (2007). Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecol. Biogeogr.*, 16, 754–763.

Hernandez P.A., Graham C.H., Master L.L. & Albert D.L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, 29, 773–785.

Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G. & Jarvis A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978.

Hijmans R.J. & Graham C.H. (2006). The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biol.*, 12, 2272–2281.

Hortal J., Lobo J.M. & Jimenez-Valverde A. (2007). Limitations of biodiversity databases: Case study on seed-plant diversity in Tenerife, Canary Islands. *Conserv. Biol.*, 21, 853–863.

Kadmon R., Farber O. & Danin A. (2004). Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecol. Appl.*, 14, 401–413.

Klink C.A. & Machado R.B. (2005). Conservation of the Brazilian Cerrado. *Conserv. Biol.*, 19, 707–713.

Lassueur T., Joost S. & Randin C.F. (2006). Very high resolution digital elevation models: Do they improve models of plant species distribution? *Ecol. Model.*, 198, 139–153.

Lobo J.M., Jiménez-Valverde A. & Real R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecol. Biogeogr.*, 17, 145–151.

Loiselle B.A., Jorgensen P.M., Consiglio T., Jimenez I., Blake J.G., Lohmann L.G. & Montiel O.M. (2008). Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *J. Biogeogr.*, 35, 105–116.

Lomolino M.V. (2004). Conservation Biogeography. In: *Frontiers of biogeography: new directions in the geography of nature* (eds. Lomolino MV & Heaney LR). Sinauer, pp. 293–296.

Machado R.B., Ramos Neto M.B., Pereira P.G.P., Caldas E.F., Gonçalves D.A., Santos N.S., Tabor K. & Steininger M. (2004). Estimativas de perda da área do Cerrado brasileiro. In. Conservation International Brasilia, DF.

Manel S., Williams H.C. & Ormerod S.J. (2001). Evaluating presence-absence models in ecology: the need to account for prevalence. *J. Appl. Ecol.*, 38, 921–931.

Mittermeier R.A., Gil P.R., Hoffman M., Pilgrim J., Brooks T., G. M.C., Lamoreux J. & Fonseca G.A.B. (2005). *Hotspots Revisited : Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions*. Second edn. Conservation International.

Myers N. (2003). Biodiversity hotspots revisited. *Bioscience*, 53, 916-917.

Myers N., Mittermeier R.A., Mittermeier C.G., da Fonseca G.A.B. & Kent J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403, 853–858.

Nogueira, C.C. & Rodrigues, M.T. (2006). The genus *Stenocercus* (Squamata: Tropiduridae) in extra-amazonian Brazil, with the description of two new species. *South American Journal of Herpetology*, 1, 149–165.

Nogueira C.C., Colli G.R. & Martins M. (In press). Local richness and distribution of the lizard fauna in natural habitat mosaics of Cerrado. *Austral Ecol.*

Parra J.L., Graham C.C. & Freile J.F. (2004). Evaluating alternative data sets for ecological niche models of birds in the Andes. *Ecography*, 27, 350–360.

Pawar S., Koo M.S., Kelley C., Ahmed M.F., Chaudhuri S. & Sarkar S. (2007). Conservation assessment and prioritization of areas in Northeast India: Priorities for amphibians and reptiles. *Biol. Conserv.*, 136, 346–361.

Pearson R.G., Raxworthy C.J., Nakamura M. & Peterson A.T. (2007). Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *J. Biogeogr.*, 34, 102–117.

Pearson R.G., Thuiller W., Araújo M.B., Martinez-Meyer E., Brotons L., McClean C., Miles L., Segurado P., Dawson T.P. & Lees D.C. (2006). Model-based uncertainty in species range prediction. *J. Biogeogr.*, 33, 1704–1711.

Peterson A.T. & Nakazawa Y. (2008). Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. *Global Ecol. Biogeogr.*, 17, 135–144.

Peterson A.T., Papes M. & Eaton M. (2007). Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography*, 30, 550–560.

Peterson A.T., Papes M. & Soberon J. (2008). Rethinking receiver operating characteristis analysis applications in ecological niche modeling. *Ecol. Model.*, 213, 63–72.

Peterson A.T. & Vieglais D.A. (2001). Predicting species invasions using ecological niche modeling: New approaches from bioinformatics attack a pressing problem. *Bioscience*, 51, 363–371.

Phillips S.J., Anderson R.P. & Schapire R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model.*, 190, 231–259.

Phillips S.J. & Dudik M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31, 161–175.

Phillips S.J., Dudik M. & Shapire R.E. (2004). A maximum entropy approach to species distribution modeling. In: *Proceedings of the 21st International Conference on Machine Learning* (eds. Greiner R & Schuurmans D). ACM press Banff, Canada, pp. 655–662.

Poyry J., Luoto M., Heikkinen R.K. & Saarinen K. (2008). Species traits are associated with the quality of bioclimatic models. *Global Ecol. Biogeogr.*, 17, 403–414.

Raes N. & ter Steege H. (2007). A null-model for significance testing of presence-only species distribution models. *Ecography*, 30, 727–736.

Raxworthy C.J., Martinez-Meyer E., Horning N., Nussbaum R.A., Schneider G.E., Ortega-Huerta M.A. & Peterson A.T. (2003). Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, 426, 837–841.

Rodrigues, M.T., Camacho, A., Nunes, P.M.S., Recoder, R.S., Teixeira, M., Valdujo, P.H., Ghellere, J.M.B., Mott, T. & Nogueira, C. (2008). A new species of the lizard genus *Bachia* (Squamata: Gymnophthalmidae) from the cerrados of central Brazil. *Zootaxa*, 1875, 39–50.

Rodrigues, M.T., Pavan, D. & Curcio, F.F. (2007). Two new species of lizards of the genus *Bachia* (Squamata, Gymnophthalmidae) from central Brazil. *J. Herpetol.*, 41, 545–553.

Segurado P. & Araújo M.B. (2004). An evaluation of methods for modelling species distributions. *J. Biogeogr.*, 31, 1555–1568.

Silveira A.L. (2007). Geographic Distribution: *Amphisbaena fuliginosa*. *Herpetological Review*, 38, 481–481.

Soberón J., Jiménez R., Golubov J. & Koleff P. (2007). Assessing completeness of biodiversity databases at different spatial scales. *Ecography*, 30, 152–160.

Soberón J. & Llorente J. (1993). The use of species accumulation functions for the prediction of species richness. *Conserv. Biol.*, 7, 480–488.

Soberón J.M., Llorente J.B. & Onate L. (2000). The use of specimen-label databases for conservation purposes: an example using Mexican Papilionid and Pierid butterflies. *Biodivers. Conserv.*, 9, 1441–1466.

Stockwell D.R.B. & Noble I.R. (1992). Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computers in Simulation*, 33, 385–390.

Stockwell D.R.B. & Peterson A.T. (2002). Effects of sample size on accuracy of species distribution models. *Ecol. Model.*, 148, 1–13.

Sullivan B.K. (1981). Distribution and relative abundance of snakes along a transect in California. *J. Herpetol.*, 15, 245–246.

Trivedi M.R., Berry P.M., Morecroft M.D. & Dawson T.P. (2008). Spatial scale affects bioclimate model projections of climate change impacts on mountain plants. *Global Change Biol.*, 14, 1089–1103.

Whittaker R.J., Araújo M.B., Paul J., Ladle R.J., Watson J.E.M. & Willis K.J. (2005). Conservation Biogeography: assessment and prospect. *Divers. Distrib.*, 11, 3–23.

Zweig M.H. & Campbell G. (1993). Receiver-operating characteristics (ROC) plots - A fundamental evaluation tool in clinical medicine. *Clin. Chem.*, 39, 561–577.

Table 1 – Species collected in the survey of PNCM. Number of points used to train the model, GARP and Maxent predictions (0 = did not predict, 1 = predict to occur at PCNM), and maximum value predicted in the PNCM. AUC values for each species for both methods are shown within parenthesis. GARP values were adjusted from 0-100 to a 0-1 scale.

Species	Number of points	GARP	Maxent	Maximum value
Ameiva ameiva	256	1 (0.57)	1 (0.82)	0.89/0.53
Amphisbaena alba	41	1 (0.79)	0 (0.9)	0.79/0.42
Anolis chrysolepis	39	1 (0.9)	1 (0.91)	0.7/0.65
Apostolepis cearensis	3	0 (0.5)	0 (0.99)	0/0.3
A. polylepis	3	0 (0.5)	0 (0.99)	0/0
Boa constrictor	111	1 (0.78)	1 (0.92)	0.99/0.54
Bothrops lutzi	15	0 (0.9)	1 (0.91)	0.4/0.64
B. moojeni	112	1 (0.86)	0 (0.95)	0.89/0.04
Chironius exoletus	21	1 (0.78)	1 (0.91)	0.29/0.77
C. flavolineatus	38	1 (0.83)	1 (0.9)	0.59/0.4
Cnemidophorus mumbuca	4	0 (0.5)	0 (0.98)	0/0.05
Colobosaura modesta	44	1 (0.9)	1 (0.91)	0.89/0.59
Corallus hortulanus	16	1 (0.89)	1 (0.93)	0.99/0.53
Drymarchon corais	55	1 (0.8)	0 (0.89)	0.89/0.25
Epicrates cenchria	103	1 (0.78)	1 (0.89)	0.79/0.52
Gymnodactylus carvalhoi	36	1 (0.91)	1 (0.96)	0.89/0.57

Hemidactylus mabouia	119	1 (0.64)	0 (0.86)	0.99/0.3
Hydrodynastes bicinctus	27	0 (0.88)	0 (0.93)	0.09/0.26
Iguana iguana	103	1 (0.64)	1 (0.83)	0.99/0.6
Imantodes cenchoa	24	0 (0.82)	0 (0.87)	0/0.3
Kentropyx calcarata	90	1 (0.74)	1 (0.86)	0.99/0.63
Leptotyphlops brasiliensis	5	0 (0.5)	1 (0.95)	0/0.64
Liophis almadensis	70	1 (0.84)	1 (0.92)	0.79/0.41
L. poecilogyrus	185	1 (0.81)	0 (0.94)	0.79/0.25
L. reginae	82	1 (0.82)	0 (0.91)	0.99/0.24
Liotyphlops ternetzii	9	0 (0.5)	1 (0.88)	0/0.33
Mabuya heathi	54	1 (0.9)	0 (0.93)	0.79/0.32
M. nigropunctata	132	1 (0.67)	1 (0.88)	0.89/0.48
Mastigodryas bifossatus	104	1 (0.82)	0 (0.91)	0.99/0.06
Micrablepharus maximiliani	50	1 (0.81)	1 (0.93)	0.99/0.66
Micrurus brasiliensis	9	0 (0.94)	0 (0.95)	0.29/0.23
Oxyrhopus trigeminus	111	1 (0.78)	1 (0.93)	0.99/0.75
Philodryas nattereri	94	1 (0.82)	1 (0.93)	0.99/0.4
P. olfersi	90	1 (0.77)	0 (0.9)	0.79/0.28
Phimophis guerini	38	1 (0.77)	0 (0.91)	0.59/0.27
P. iglesiasi	9	0 (0.89)	0 (0.94)	0/0.08
Pseudoboa neuwiedii	24	0 (0.82)	0 (0.96)	0.29/0.16
P. nigra	60	1 (0.81)	0 (0.93)	0.79/0.18
Psomophis joberti	46	1 (0.79)	1 (0.91)	0.89/0.67

Sibynomorphus mikanii	97	1 (0.92)	1 (0.93)	0.69/0.36
Spilotes pullatus	67	1 (0.85)	0 (0.9)	0.89/0.27
Tantilla melanocephala	37	1 (0.81)	1 (0.91)	0.99/0.48
Thamnodynastes hypoconia	21	0 (0.94)	0 (0.96)	0.2/0.38
Tropidurus oreadicus	50	1 (0.8)	1 (0.92)	0.99/0.75
Tupinambis merianae	66	1 (0.79)	0 (0.87)	0.99/0.31
T. teguixin	48	1 (0.71)	0 (0.82)	0.99/0.47
Typhlops brongersmianus	24	0 (0.84)	0 (0.93)	0.29/0.37
Waglerophis merremi	136	1 (0.84)	0 (0.93)	0.69/0.16

* *Amphisbaena* sp. was collected but not used in the analysis due to taxonomic uncertainties.

Table 2. Results of multiple logistic regression to model the effects of: Degree of dispersal on sampled points (Nearest Neighbor); Distance from the closest point (DNP); and Number of Points used in the modeling, on the ability of GARP and Maxent to successfully predict species occurrence in PNCM (0 = fail, 1 = success). Results after slash are from regression after eliminating species with known locality points lower that 15. Degree of freedom is equal to 1 in all cases, and sample sizes are 49 and 42. β are the individual regression coefficients, which are tested using the Wald's chi-square statistics. e^{β} is the odds ratio, which is the predicted change in odds for a unit increase in the corresponding independent variable. Odds ratios less than 1 correspond to decreases and odds ratios more than 1.0 correspond to increases in odds. Odds ratios close to 1.0 indicate that unit changes in that independent variable do not affect the dependent variable.

Predictor	β	SE β	Wald's χ^2	P	e^{β} (odds ratio)
			GARP		
Intercept	-9.90/-18.8	11.54/17.79	0.74/1.12	0.39/0.28	NA
Nearest Neighbor	4.33/19.2	8.07/19.8	0.29/0.94	0.59/0.33	75.6/73.8
DNP	-0.07/-0.09	0.07/0.08	1.03/1.44	0.31/0.23	0.93/0.91

Number of Points	0.72/0.98	0.68/0.81	1.12/1.46	0.29/0.23	2.06/2.65
			Maxent		
Intercept	3.21/3.23	1.75/2.10	3.35/2.37	0.06/0.12	NA
Nearest Neighbor	-0.86/-0.34	1.75/2.52	0.24/0.02	0.62/0.89	0.42/0.71
DNP	-0.01/-0.01	<0.01/<0.01	10.8/10.5	<0.01*/<0.01*	0.98/0.98
Number of Points	<-0.01/<-0.01	<0.01/<0.01	0.46/0.57	0.49/0.45	0.99/1.01

Figure Legends

Figure 1 – Map of the study area and sampling profile for the dataset for the Brazilian Cerrado Squamates. On the upper figure, the gray shading corresponds to the limits of the Cerrado Biome. The inset on the map of Brazil shows the state of Maranhão in the black square, where PNCM is located. On the lower figure, Kernel density function was applied using all sampling points to create a smooth tapered surface. Darker regions indicate higher density of sampling points (more specimens were collected from those regions). The star symbol represents the location of the field site in the Northern portion of the Cerrado.

Figure 2 – Diagrams representing possible scenarios for ENM when sampling was biased in different ways. . Circles represent known localities, dashed lines represent areas surveyed, and gray areas represent the predicted distribution of the species based on ENM. For details, see text.

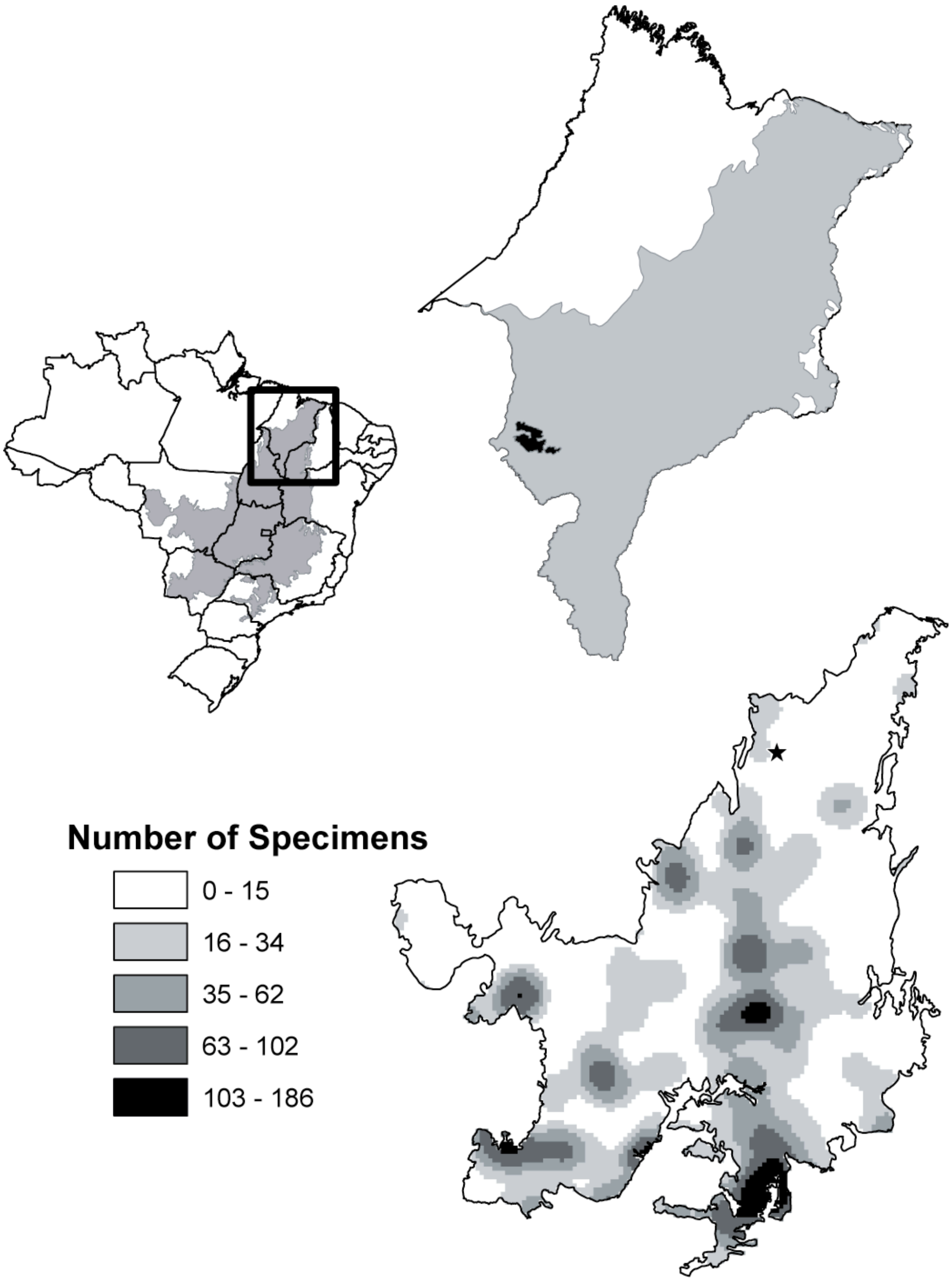
Figure 3 – Results of the accumulation curve analysis. Open circles represent mean values from 10,000 randomizations without replacement of the original matrix.

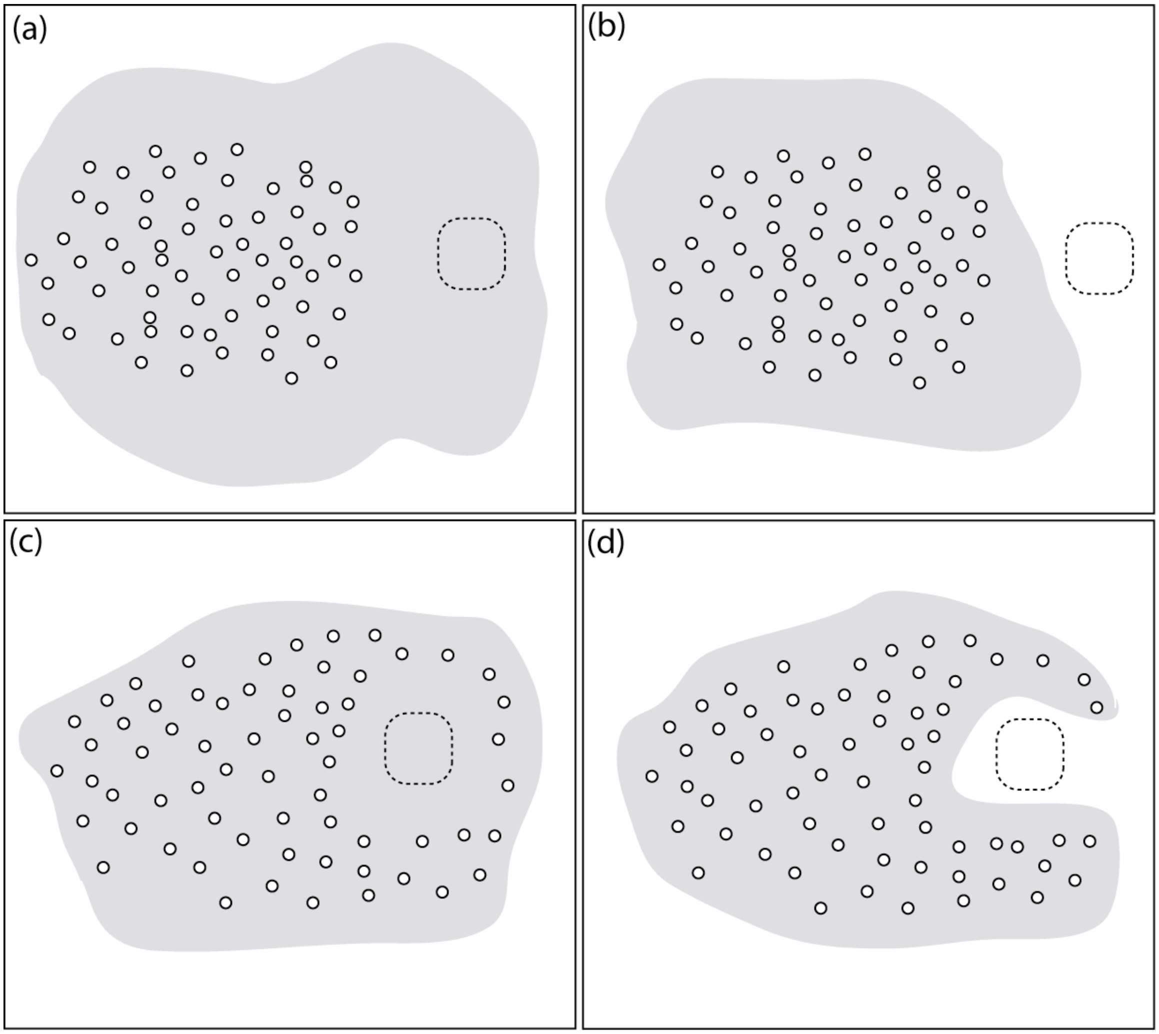
Figure 4 – Example of ENM results where GARP successfully predicted the presence of the species in PNCM, whereas Maxent failed to predict the presence. a-c, species follow the pattern described in Figure 2a; d-f, species follow the pattern described in Figure 2c (see text). Circles represent known localities, blue represents GARP predictions, green

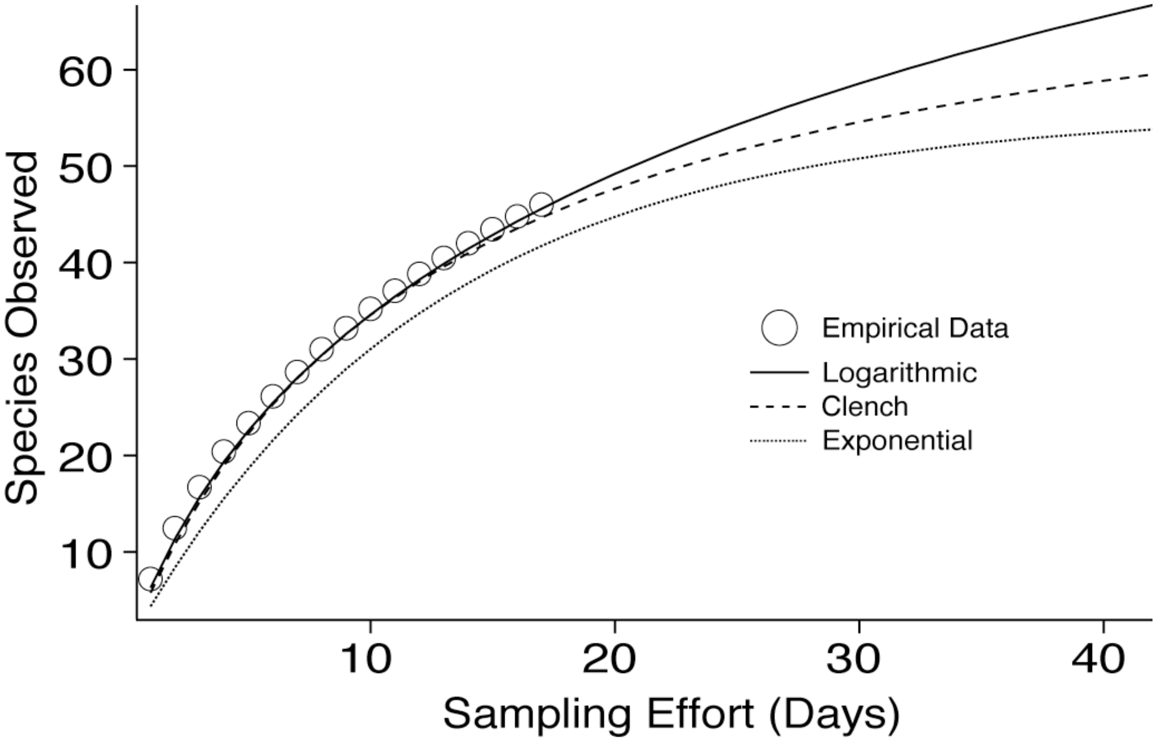
Maxent predictions, and red the coincidence of both methods. (a) – *Amphisbaena alba*, (b) – *Bothrops moojeni*, (c) – *Mastigodryas bifossatus*, (d) – *Drymarchon corais*, (e) – *Liophis poecilogyrus*, (f) – *Spilotes pullatus*.

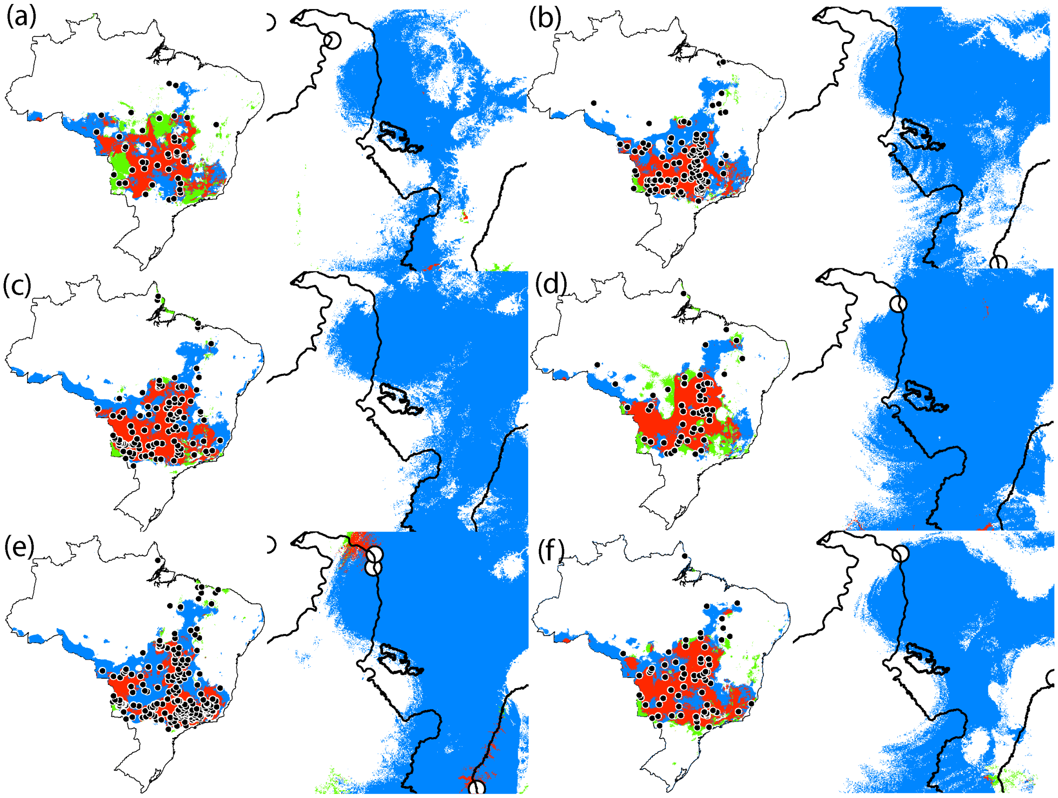
Figure 5 – Example of ENM results where Maxent successfully predict the presence of the species in PNCM, whereas GARP failed to predict the presence. (a) – *Bothrops lutzi*, (b) – *Leptotyphlops brasiliensis*. Both GARP and Maxent failed to predict species presence in some cases. (c) – *Apostolepis polylepis* (d) – *Thamnodynastes hypoconia*.

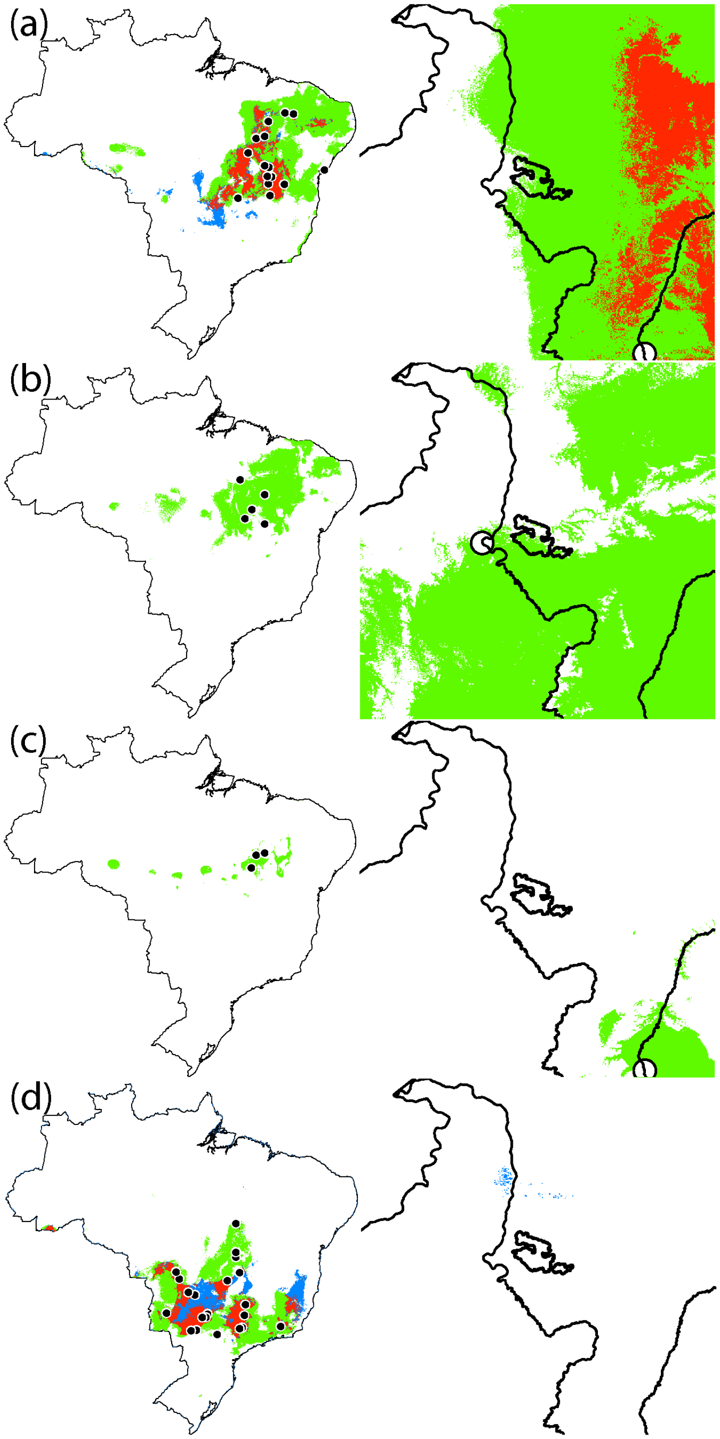
Figure 6 – Relationship between number of points used to train the models and largest predicted value in PNCM. Circles represent GARP models, and triangles represent Maxent models. Closed symbols are for species correctly predicted by the method, and open symbols for species collected in PCNM but not predicted by the method. Regression line is for GARP models ($F = 24.6$, $P < 0.01$, $r^2 = 0.35$).

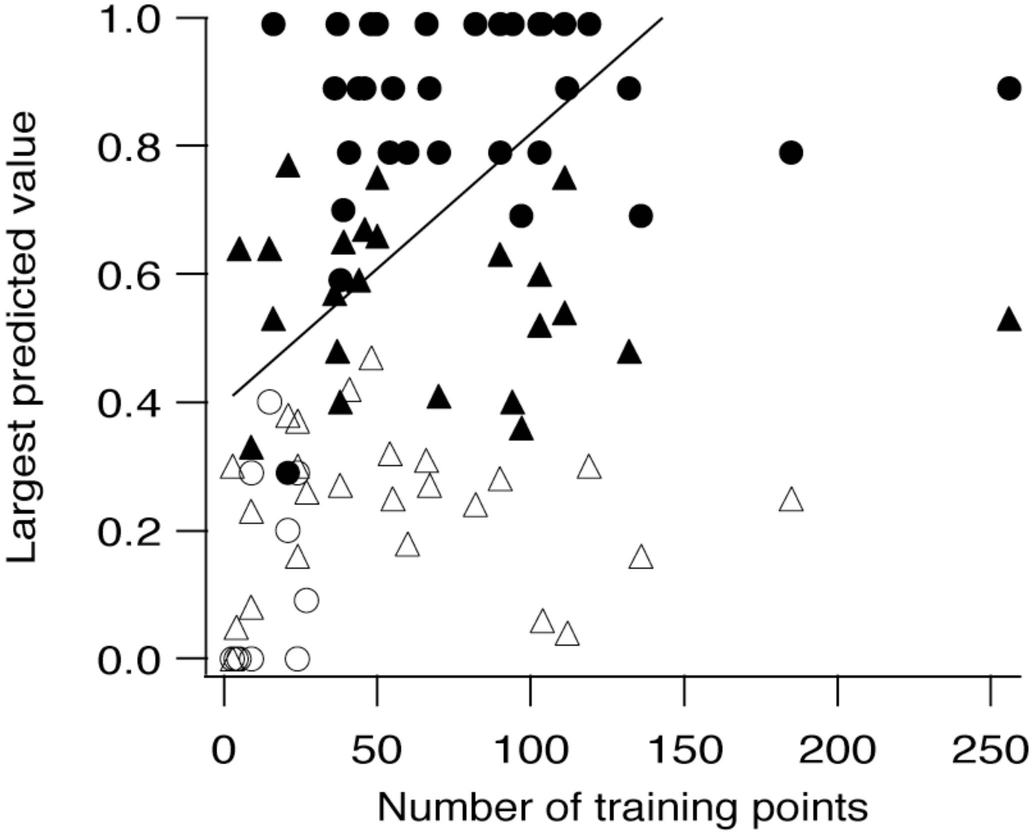

Number of Specimens
0 - 15
16 - 34
35 - 62
63 - 102
103 - 186


(a)
(b)
(c)
(d)


Species Observed
10
20
30
40
50
60
10
20
30
40
Sampling Effort (Days)
Empirical Data
Logarithmic
Clench
Exponential


(a)
(b)
(c)
(d)
(e)
(f)


(a)
(b)
(c)
(d)


Largest predicted value
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
Number of training points

Chapter III

Detecting the influence of climatic variables on species' distributions: a test using GIS niche-based models along a steep longitudinal environmental gradient

(formatted for *Journal of Biogeography*)

ABSTRACT

Aim: To investigate the influence of climatic variables in shaping species' distributions across a steep longitudinal environmental gradient.

Location: The state of Oklahoma, south-central United States.

Methods: We used Geographical Information Systems (GIS) niche-based models to predict the geographic distributions of six pairs of closely related amphibian and reptile species across a steep longitudinal environmental gradient. We compared results from modelling with actual distributions to determine whether species' distributions were primarily limited by environmental factors and to assess potential roles of competition and historical factors in influencing distributions.

Results: For all species pairs, GIS models predicted an overlap zone in which both species should occur even though in some cases this area was occupied by only one of the species. We found that environmental factors clearly influence distributions of most species pairs. We also found evidence that suggests competition and evolutionary history have a role in determining distributions of some species pairs.

Main conclusions: Niche-based GIS modelling is a useful tool to investigate species distribution patterns and factors affecting them. Our results showed that environmental factors strongly influenced species' distributions, and that competition and historical factors may also be involved in some cases. Further, results suggested additional lines of research, such as ecological comparisons among populations occurring inside and outside of predicted overlap zones, which may provide more direct insight into the roles of competitive interactions and historical factors in shaping species' distributions.

Keywords: Biogeography, competition, GARP, Oklahoma, niche modelling, niche overlap

INTRODUCTION

Factors that influence the geographic distributions of species include how organisms relate to their environment (i.e., niche requirements) and interspecific interactions such as competition, predation, and parasitism (MacArthur, 1984; Chesson, 2000; Chave *et al.*, 2002). Historical factors such as geographic barriers and/or lack of sufficient dispersal opportunities also influence species distributions (Brown *et al.*, 1996; Patterson, 1999). A species niche is defined as the set of environmental conditions required for the species to maintain a viable population in order to persist through time (Hutchinson, 1957; Chase & Leibold, 2003). A species is seldom able to occupy all of its geographic potential range, and the presence of other species as well as historical factors often reduce distributions to a smaller subset of the potential range (Hutchinson, 1957; Chase & Leibold, 2003).

Teasing apart which factors exert the most influence is a major challenge when investigating how ecological and/or historical factors shape the distribution of a species (Endler, 1982). Much of the difficulty stems from the lack of objective means for identifying regions of suitable habitat and integrating how ecological and/or historical factors shape the species' actual distribution. Recent advances in Geographical Information Systems (GIS) allow modelling of species' distributions based on attributes of the environment that should be correlated with niche requirements of species (Peterson, 2001; Guisan & Thuiller, 2005). These models identify previously unsampled locations where the species has a high probability of occurrence and have been applied successfully to predict the geographic distributions of several animal groups in a variety of ecosystems (Peterson, 2001; Luoto *et al*., 2002; Raxworthy *et al*., 2003). However, these techniques do not account for species interactions or historical factors (Araujo & Guisan 2006). By examining congruence or discordance between predicted distributions and actual distributions, researchers can evaluate the potential role of ecological and historical factors in determining a species geographic distribution (Anderson *et al.*, 2002a; Anderson *et al.*, 2002b).

GIS niche-based models that focus on closely related species occurring in adjoining or slightly overlapping areas along a well-sampled environmental gradient might be useful in exploring the roles of competitive interactions and/or environmental characteristics in limiting and shaping the distributions of species. Results of such analyses can reveal three possible patterns (Fig. 1), each of which can be interpreted to support a specific factor in determining species' distributions. First, if a GIS niche-based model shows that two species do not overlap at all in their predicted distributions, then the most likely explanation is that the distributions of both species are limited by unfavourable environmental factors that prevent further expansion along an environmental gradient (Fig. 1A). Second, if niche-based modelling predicts an overlap zone where both species are known to occur, then favourable environmental conditions exist that should allow both species to occur in the overlap zone (Fig. 1B). In this case, niche segregation at smaller spatial scales (e.g., microhabitat, activity period, diet, etc.) would be predicted to allow for coexistence of both species. In addition, if the species are sister taxa with different niche characteristics, niche lability or character displacement may be occurring to avoid competitive interactions (Losos *et al.*, 2003). Alternatively, if evidence for niche partitioning (e.g., microhabitat segregation) is lacking, then resources may not be limiting, and species can occur together without niche or character shifts.

Coexistence may also be facilitated by other ecological interactions such as predation, where a top predator can regulate population size in a way that relaxes competition among closely related species (Hanski, 1981). In the third possible pattern, only one species occupies the predicted overlap zone, indicating that favourable conditions may exist for both species, but either one species competitively excludes the other or some historical factor prevented one of the species from colonizing the area (Fig. 1C). In this case, if the species are sister taxa, theoretical and empirical evidence predict that niches are probably conserved (Peterson *et al.*, 1999; Wiens & Graham, 2005). If natural history data confirm that they have similar niches, then competition or historical factors might determine their distributions in the overlap zone. However, if the species have different niches, then the absence of one species in the predicted overlap zone may reflect the impact of historical factors.

Here we investigate potential factors that limit the distributions of species by building niche-based models of geographic distributions of closely related species across a steep longitudinal environmental gradient. With a sharp transition from eastern deciduous forest in the east to open habitats in the west, the central United States is ideal for testing models of this type. To evaluate generality of the models, we selected six species pairs differing in evolutionary history, bauplan, and overall ecology (two scincid lizards, two ranid frogs, two viperid snakes, two microhylid frogs, two scaphiopodid frogs, and two colubrid snakes).

MATERIAL AND METHODS

Study area

A distinct environmental gradient is ideal for investigating factors that influence the distributions of species. Overall, a multitude of climatic–environmental variables show an east to west transition in the United States (Fig.2), and this transition is steepest in the state of Oklahoma (Figs. 2 and 3). Oklahoma is located in the south-central United States, bordered by Missouri and Arkansas to the east, Texas to the south, New Mexico to the west, and Colorado and Kansas to the north (Fig. 4). Oklahoma can be divided into nine major physiognomic regions (see Caire, 1989 and references therein for details). The eastern region of the state lies within the western edge of the Interior Highlands (Ozark and Ouachita uplifts) and primarily consists of oak–hickory forest. In contrast, the western region of the state lies within areas of sandstone and gypsum hills in a flatter grassland/prairie physiognomy (Johnson & Duchon, 1994). The climate of Oklahoma exhibits a marked change along an east–west gradient, which in turn is mirrored by the state's vegetation (Figs. 3 and 4). For example, mean annual precipitation varies from 1270 mm in the southeast Ouachita Mountains to 600 mm in the high plains of the west (Fig. 3B). In addition, rainfall among years and different seasons (Fig. 3D) can be highly variable (Johnson & Duchon, 1994).

In summary, we took advantage of a very interesting combination of factors that makes our analysis particularly meaningful. First, Oklahoma is situated within a sharp longitudinal environmental gradient (Fig. 2). Second, the western/eastern geographical limits of the distributions of the species lie within the state. Finally, our conclusions require good sampling inside the overlap zone, which we have for the region in which the steep environmental gradient exists (i.e., Oklahoma). Although adding the overlap zone north (e.g., Kansas) and/or south (e.g., Texas) of Oklahoma could be interesting in itself, the additional variables (e.g., latitude, and their effects on species distributions and community structure, differences in steepness of the gradient, etc.) would divert the study from its primary focus (Why do ‘western species’ not occur further east, and why do ‘eastern species’ not occur further west?).

Species selection

Ectotherms such as amphibians and reptiles are good model organisms to study the effects of climatic–environmental characteristics on the distributions of species. Their thermal ecology, physiology, and behaviour are generally highly dependent upon environmental conditions, and in temperate-zone amphibians, which generally rely on aquatic habitats for reproduction, precipitation can be especially critical (Zug *et al.*, 2001). Thus, environmental characteristics are likely to have a strong influence on limiting species’ distributions in these groups.

The state of Oklahoma has a high diversity of amphibians and reptiles with 58 species of amphibians and 102 species of reptiles (Sievert & Sievert, 2005). Additionally, many species reach their eastern or western distribution limits within the state (Sievert & Sievert, 2005). For this study, we selected pairs of closely related amphibian and reptile species from a diversity of higher taxonomic groups whose distributions show that one species is associated with the western region of the state and the other species is associated with the eastern region. We chose sister species or closely related species based upon recent phylogenetic hypotheses (Cole & Hardy, 1981; Brandley *et al.*, 2005; Hillis & Wilcox, 2005). Six pairs of species met our criteria (three amphibian and three reptile species pairs): the microhylid frogs *Gastrophryne olivacea–G. carolinensis*; the ranid frogs *Rana blairi–R.sphenocephala*; the scaphiopodid frogs *Scaphiopus couchii–S. hurterii*; the scincid lizards *Eumeces obsoletus–E. fasciatus*; the viperid snakes *Sistrurus catenatus–S. miliarius*; and the colubrid snakes *Tantilla nigriceps–T. gracilis*.

Niche modelling

Locality data from the state of Oklahoma for all species were collected primarily from voucher specimens and databases housed in the Herpetology collection of the Sam Noble Oklahoma Museum of Natural History (OMNH). Reptiles and amphibians of Oklahoma are particularly well represented in the collection, at over 28,000 specimens. In addition, to ensure that we sampled the full range of environmental conditions in which each species occurs, we included locality points from throughout their entire North American distributions. For records outside Oklahoma, we gathered museum specimen locality data from online databases such as HerpNet (http://www.herpnet.org) and GBIF (http://www.herpnet.org/gbif/gbif.html). We selected one locality per county, where available, and either obtained geographic coordinates from collection databases or published studies, or georeferenced localities using locality descriptions (data available from G. Costa on request). For pygmy rattlesnakes (*Sistrurus*), both species are wide-ranging and may consist of multiple evolutionary lineages. Thus, we chose to limit our focus to only the subspecies occurring in Oklahoma, *Sistrurus catenatus tergeminus* and *S. miliarius streckeri* (Campbell *et al.*, 2004) because these taxa are most relevant to our study. Both subspecies are easily delimited by their distributions and morphological characteristics (Campbell *et al.*, 2004).

We used DesktopGARP® to create niche-based models of species' distributions. This software uses the GARP algorithm (Genetic Algorithm for Rule-Set Prediction), which includes several distinct algorithms in an iterative, artificial-intelligence approach based on species presence data points (Stockwell & Peters, 1999). DesktopGARP® generates predicted distributions of each species based on characteristics of environmental–climatic variables for localities in which a given species has been previously documented. A database of climatic variables is created and loaded into the software program. We created a database using 20 variables available in the Worldclim project (Hijmans *et al.*, 2005). Details, descriptions, and files for download are available free on-line at: http://www.worldclim.org/.

We used the following options while running DesktopGARP®: Optimization parameters – 100 runs, 0.001 convergence limit, and 1000 maximum interactions; Rule Types – Atomic Range, Negated Range, and Logistic regression; Best subset active, 5% omission error, 50% commission error, and 50% of points for training; Omission measure = extrinsic, and Omission threshold = Hard; 10 models under hard omission threshold.

The output of DesktopGARP® consists of Arc/Info grid maps with 'zeros' where the species does not occur and 'ones' where the species is predicted to occur. To generate a single distribution map for each species, we used the area covered by at least 5 out of the 10 models in the best subset selection. This procedure is called 'ensemble forecasting' and it has been discussed recently in detail by Araújo & New (2007). Using the area covered by of 5 out of 10 models is an arbitrary decision, but using less than 5 would result in more liberal predicted distributions, and using more than 5 would result in predictions that are too conservative. In addition, by using only the models in the best subset selection, we are optimizing our results with respect to omission/commission relationships (see Anderson et al., 2003).

RESULTS

The models produced had very low values of commission and omission errors (Table 1). Overall results of niche modelling for all six species pairs are presented in Fig. 5. None of the six species pairs exhibited the pattern described in Fig. 1A, because niche modelling always predicted some overlap in geographic distributions. As a general pattern, the overlap zone was always located in the central portion of the state (Fig. 5A–F), coincident with the centre of the environmental variable curves (Fig. 3). In some cases, the overlap zone was relatively narrow (e.g., *Scaphiopus couchii–S. hurterii* and *Sistrurus catenatus–S. miliarius*; Figs. 5D and 5E, respectively), whereas in other cases, the overlap zone was wide (e.g., all other species pairs, Figs. 5A–C and 5F).

Four species pairs showed a pattern similar to Fig. 1B, where both species occur in the predicted overlap zone (Figs. 5A, 5C, and 5E–F). In the *Eumeces obsoletus–E. fasciatus* comparison (Fig. 5A), *E. obsoletus* reaches areas at the eastern limit of the predicted overlap zone, and *E. fasciatus* occurs in one area at the western limit of the overlap zone (Fig. 5A). In *Rana blairi–R. sphenocephala* (Fig. 5C), both species occur throughout the large predicted overlap zone with several areas where the species are known to occur syntopically. *Sistrurus catenatus–S. miliarius* exhibited the narrowest overlap zone among all species pairs. The eastern species, *S. miliarius*, and the western species, *S. catenatus*, are both present in the central part of the overlap zone (Fig. 5E). Neither species, however, has yet been collected in most of the overlap zone. Finally, *Tantilla nigriceps–T. gracilis* (Fig. 5F) also exhibited the pattern in Fig. 1B. Both species occur together in the western portion of the overlap zone; however, the eastern species extends farther westward than the western species extends eastward (Fig. 5F).

Two species pairs exhibited patterns similar to that shown in Fig. 1C. In the *Gastrophryne olivacea–G. carolinensis* comparison, most of the predicted overlap zone is dominated by the western species, *G. olivacea*. The eastern species, *G. carolinensis,* is restricted to the extreme eastern portion of the overlap zone, while the western species, *G. olivacea*, advances much farther east along the Arkansas River floodplain (Fig. 5B). In the *Scaphiopus couchii–S. hurterii* comparison, the predicted overlap zone is dominated by the western species, *S. couchii,* whereas the eastern species, *S. hurterii*, is restricted to the extreme eastern portion of the overlap zone (Fig. 5D).

DISCUSSION

The observed patterns are clearly a reflection of the steep east to west gradient of climatic conditions that characterizes the state of Oklahoma. However, a predicted distribution overlap zone was present in all species pairs, suggesting that regions exist that have favourable macroenvironmental conditions for both species to occur. To understand factors that limit the distributions of these species, it is necessary to examine each case in detail.

The two species of skinks (*Eumeces*) selected for this study are closely related, appearing as sister taxa in the most recent phylogeny (Brandley *et al.*, 2005), and *E. obsoletus* appears as the sister taxon to the clade containing *E. septentrionalis* + *E. fasciatus* in another study (Schmitz *et al.*, 2004). Although these two species are closely related, they clearly segregate in habitat use on a finer scale. *Eumeces obsoletus* is found in more xeric microhabitats in grasslands, prairies, and deserts (Fitch, 1955), whereas *E. fasciatus* is widely distributed throughout the deciduous hardwood forests of eastern North America (Conant, 1975). Therefore, microhabitat availability within the area of suitable macroenvironmental–climatic conditions may be the ultimate factor determining species presence in the overlap zone.

Five species of *Gastrophryne* are currently recognized (Frost, 2006); however, no phylogeny containing all species in the group is available. Previous work suggests that *G. olivacea* and *G. carolinensis* are sister species (Nelson, 1972), and hybridization between these species is known to occur (Blair, 1955). *Gastrophryne olivacea* occurs throughout the western three-fourths of the state whereas the distribution of *G. carolinensis* is limited to the eastern one-third. The predicted overlap zone is wide, but inhabited mainly by the western species. Areas of sympatry are known from the coastal plain of east Texas through northeast Oklahoma (Blair, 1955). In areas of sympatry, the two species have significant differences in their mating calls compared to conspecific populations in allopatry (Blair, 1955; Loftus-Hills & Littlejohn, 1992), which suggests a possible mechanism to avoid hybridization (i.e., reproductive character displacement). Because of these characteristics, this species pair provides an ideal scenario to look for competitive interactions. Differences in diet, microhabitat use, breeding period, or other niche aspects between sites where these species are sympatric versus allopatric would be indicative of strong competitive interactions within the area of sympatry. In order to coexist, one species (or both) should diverge along one or more niche axes (i.e., niche segregation). In Oklahoma, the fact that the overlap zone is largely dominated by the western species even though suitable habitat for the eastern species exists suggests that competition may play a role in determining the distributional limits of these species.

Although not sister taxa, *Rana blairi* and *R. sphenocephala* are closely related; *R. sphenocephala* is the sister taxon to the clade containing *R. blairi* (Hillis & Wilcox, 2005). The niche models predicted a wide overlap zone where both species should be found, which is corroborated by the large number of known locality points for both species in the overlap zone. In addition, areas of sympatry are widespread and hybridization is common throughout the ranges of these species (Parris, 2001). This pattern suggests that the distributions of these species are not greatly influenced by competition. Studies that investigate populations of both species at a smaller spatial scale in the overlap zone may provide insight into whether character displacement or niche segregation occurs to allow coexistence and avoid competitive interactions. If resources are not limiting, the determining factors for the eastern species to invade habitats farther west and vice-versa may simply be the climatic–environmental conditions required by each species.

Spadefoot toads (*Scaphiopus*) exhibited an interesting pattern because the western species, *S. couchii*, predominates in the predicted overlap zone. The presence of just one of the species in the overlap zone predicted by the GIS models (Fig. 1C) indicates that competitive exclusion or historical factors influence the distributions of these species, especially if available microhabitats for *S. hurterii* are present. Because the species are closely related (Garcia-Paris et al., 2003), they are predicted to have similar niche requirements (Peterson *et al.*, 1999). However, natural history data show that the species, in fact, occupy quite different microhabitats and thus, competitive exclusion is not expected. *Scaphiopus couchii* occurs in deserts and xeric regions for which they have morphological, behavioural, and physiological adaptations (Mayhew, 1965), whereas *S. hurterii* occurs predominantly in areas of sandy, gravelly, or soft, light soils in open woodland, savanna, and mesquite scrub (Mayhew, 1965; Bartlett & Bartlett, 1999). Therefore, following our predictions, the best explanation for the observed pattern is that historical factors limit the distribution of *S. hurterii* (see comments on Fig. 1C). The main portions of the distributions of these species lie more to the south and southeast for *S. hurterii*, and to the southwest for *S. couchii*. Although environmental conditions suitable for both species occur in the overlap zone, especially in the northern portion of the state, it appears that the species have not colonized these areas. It is possible that the species have not had sufficient dispersal opportunities. Pleistocene glaciations and their associated effects have been hypothesized to influence current distribution patterns of many eastern North American amphibians and reptiles (Hewitt, 1996, Howes *et al.*, 2006).

Considering pygmy rattlesnakes (*Sistrurus*), the eastern species, *S. miliarius*, is found in areas at the western limit of the overlap zone whereas the western species, *S. catenatus*, does not extend far eastward in the overlap zone. These two species are sister taxa (Knight *et al.*, 1993), but differ considerably in microhabitat. *Sistrurus catenatus*, as currently recognized, is a wide-ranging species and although it is usually found in moist habitats such as swamps, marshes, bogs, wet meadows, or seasonally moist grasslands in the northeastern U.S., the species occurs in river bottoms, dry grasslands, mesquite plains, and other dry areas in the west (Ernst, 1992). Throughout most of its range, the eastern species, *S. miliarius*, occurs mainly near water in mixed pine–hardwood forest, scrub pinewood, sandhills, and wiregrass flatwoods. In Texas and Oklahoma, the species is restricted to mesic grasslands (Ernst, 1992). In addition to habitat differences, dietary studies show that these species generally consume different prey (Werler & Dixon, 2000, Holycross & Mackessy, 2002). In this case, distributions of these two species are best explained by presence of the right kind of microhabitat and prey within the area of suitable macroenvironmental–climatic conditions. Support for our interpretation of results for *Sistrurus* suffers from lack of sampling points in most of the overlap zone. Additional sampling, especially in areas of predicted overlap, is needed to confirm our conclusions.

In black-headed snakes (*Tantilla*) used in this study, the predicted overlap zone is inhabited by both species. No comprehensive phylogeny for this genus is available, but a study with a limited number of taxa found that these species are not sister taxa, but closely related (Cole & Hardy, 1981). Natural history data show that these two species occupy similar habitats of rocky stretches, hillsides, rotten wood, and a surface of damp soil (Werler & Dixon, 2000). Following our predictions, segregation may occur in other aspects of their niches such as diet and/or daily activity. Not enough data are available on these ecological aspects; however, *T. nigriceps* can be up to 1.5 times larger than *T. gracilis* (Werler & Dixon, 2000), and body size differences are associated with dietary differences in snakes (Filippi *et al.*, 2005, Mushinsky *et al.*, 1982). The co-occurrence pattern is not evenly distributed; the eastern species, *T. gracilis*, occurs far westward into the overlap zone, whereas the western species, *T. nigriceps*, does not extend far eastward into the overlap zone. Further work should focus on investigating why the western species does not occupy areas farther east.

Environmental conditions clearly affected the distributions of species along the environmental gradient we studied, demonstrating the utility of using niche models to investigate distribution patterns and the factors affecting them. Nevertheless, we cannot assign an active role to competition and historical effects based on these data alone. By using niche models as representations of the potential geographic distributions of species, we are able to provide directional hypotheses that can be tested in future studies. Investigations at smaller spatial scales on the ecology of the species studied here, especially comparisons among populations occurring inside and outside the predicted overlap zone, will give more conclusive results about the roles of competitive interactions and historical factors in shaping the distributions of species.

ACKNOWLEDGEMENTS

We thank the University of Kansas Natural History Museum for providing locality points for *Sistrurus*. This work was submitted in partial fulfilment of GCC's Ph.D. degree at the University of Oklahoma. We thank T. Rashed, J. Kelly, and M. Kaspari for serving on GCC's doctoral committee and providing valuable comments on the manuscript. GCC was supported by a fellowship from CAPES/Fulbright and by a graduate research assistantship from the Sam Noble Oklahoma Museum of Natural History.

REFERENCES

Anderson, R.P., Gomez-Laverde, M. & Peterson, A.T. (2002a) Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Global Ecology and Biogeography*, **11**, 131-141.

Anderson, R.P., Peterson, A.T. & Gomez-Laverde, M. (2002b) Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos*, **98**, 3-16.

Anderson, R. P., Lew, D. & Peterson, A. T. (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling.* **162**, 211-232.

Araujo, M. B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography.* **33**, 1677-1688.

Araújo, M. B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution.* **22**, 42-47.

Bartlett, R.D. & Bartlett, P.P. (1999) *A Field Guide to Florida Reptiles and Amphibians*. Gulf Publishing Company, Houston, TX, USA.

Blair, W.F. (1955) Mating call and stage of speciation in the *Microhyla olivacea–M. carolinensis* complex. *Evolution*, **9**, 469-480.

Brandley, M.C., Schmitz, A. & Reeder, T.W. (2005) Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. *Systematic Biology*, **54**, 373-390.

Brown, J.H., Stevens, G.C. & Kaufman, D.M. (1996) The geographic range: Size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics*, **27**, 597-623.

Caire, W. (1989) Physiognomic regions of Oklahoma. *Mammals of Oklahoma* (ed. by W. Caire, J.D. Tyler, B.P. Glass and M.A. Mares), pp. 12-26. University of Oklahoma Press, Norman, Oklahoma.

Campbell, J. A., Lamar, W. W. & Brodie, E. D. (2004) *The Venomous Reptiles of the Western Hemisphere*. Comstock, Ithaca, New York

Chase, J.M. & Leibold, M.A. (2003) *Ecological Niches: Linking Classical and Contemporary Approaches*. University of Chicago Press, Chicago.

Chave, J., Muller-Landau, H.C. & Levin, S.A. (2002) Comparing classical community models: Theoretical consequences for patterns of diversity. *American Naturalist*, **159**, 1-23.

Chesson, P. (2000) Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics*, **31**, 343-358.

Cole, C. J. & Hardy, L. M. (1981) Systematics of North American colubrid snakes related to *Tantilla planiceps (blainville). Bulletin of the American Museum of Natural History.* **171**, 199-284.

Conant, R. (1975) *A Field Guide to Reptiles and Amphibians of Eastern and Central North America.* 2nd edn. Houghton Mifflin, Boston, Massachusettts, U.S.A.

Endler, J.A. (1982) Problems in distinguishing historical from ecological factors in biogeography. *American Zoologist*, **22**, 441-452.

Ernst, C.H. (1992) *Venomous Reptiles of North America*. Smithsonian Institution Press, Washington.

Fitch, H.S. (1955) Habits and adaptations of the Great Plains skink (*Eumeces obsoletus*). *Ecological Monographs*, **25**, 59-83.

Filippi, E., Rugiero, L., Capula, M., Capizzi, D. & Luiselli, L. (2005) Comparative food habits and body size of five populations of *Elaphe quatuorlineata*: the effects of habitat variation, and the consequences of intersexual body size dimorphism on diet divergence. *Copeia.* 517-525.

Frost, D.R. (2006) Amphibian Species of the World: an Online Reference. Version 4.0 (17 August, 2006). Electronic Database accessible at http://research.amnh.org/herpetology/amphibia/index.html. . In New York, USA.: American Museum of Natural History.

Garcia-Paris, M., Buchholz, D.R. & Parra-Olea, G. (2003) Phylogenetic relationships of Pelobatoidea re-examined using mtDNA. *Molecular Phylogenetics and Evolution*, **28**, 12-23.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.

Hanski, I. (1981) Coexistence of competitors in patchy environment with and without predation. *Oikos*, **37**, 306-312.

Hewitt, G. M. (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society.* **58**, 247-276.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Hillis, D.M. & Wilcox, T.P. (2005) Phylogeny of the New World true frogs (*Rana*). *Molecular Phylogenetics and Evolution*, **34**, 299-314.

Holycross, A. T. & Mackessy, S. P. (2002) Variation in the diet of *Sistrurus catenatus* (Massasauga), with emphasis on *Sistrurus catenatus edwardsii* (Desert Massasauga). *Journal of Herpetology.* **36**, 454-464.

Howes, B. J., Lindsay, B. & Lougheed, S. C. (2006) Range-wide phylogeography of a temperate lizard, the Five-lined Skink (*Eumeces fasciatus*). *Molecular Phylogenetics and Evolution.* **40**, 183-194.

Hutchinson, G.E. (1957) Population Studies - Animal ecology and demography - Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415-427.

Johnson, H.L. & Duchon, C.E. (1994) *Atlas of Oklahoma Climate*. University of Oklahoma Press, Norman, Oklahoma.

Knight, A., Styer, D., Pelikan, S., Campbell, J. A., Densmore, L. D. & Mindell, D. P. (1993) Choosing among hypotheses of Rattlesnake phylogeny - a best-fit rate test for DNA-sequence data. *Systematic Biology.* **42**, 356-367.

Loftus-Hills, J.J. & Littlejohn, M.J. (1992) Reinforcement and reproductive character displacement in *Gastrophryne carolinensis* and *G. olivacea* (Anura: Microhylidae): A reexamination. *Evolution*, **46**, 896-906.

Losos, J.B., Leal, M., Glor, R.E., de Queiroz, K., Hertz, P.E., Schettino, L.R., Lara, A.C., Jackman, T.R. & Larson, A. (2003) Niche lability in the evolution of a Caribbean lizard community. *Nature*, **424**, 542-545.

Luoto, M., Kuussaari, M. & Toivonen, T. (2002) Modelling butterfly distribution based on remote sensing data. *Journal of Biogeography*, **29**, 1027-1037.

MacArthur, R.H. (1984) *Geographical Ecology: Patterns in the Distribution of Species*. Princeton University Press, Princeton.

Mayhew, W.W. (1965) Adaptations of the amphibian, *Scaphiopus couchi*, to desert conditions. *American Midland Naturalist*, **74**, 95-109.

Mushinsky, H. R., Hebrard, J. J. & Vodopich, D. S. (1982) Ontogeny of water snake foraging ecology. *Ecology.* **63**, 1624-1629.

Nelson, C.E. (1972) Systematics studies of the North American microhylid genus *Gastrophryne*. *Journal of Herpetology*, **6**, 111-137.

Parris, M.J. (2001) Hybridization in leopard frogs (*Rana pipiens* complex): Variation in interspecific hybrid larval fitness components along a natural contact zone. *Evolutionary Ecology Research*, **3**, 91-105.

Patterson, B.D. (1999) Contingency and determinism in mammalian biogeography: The role of history. *Journal of Mammalogy*, **80**, 345-360.

Peterson, A.T. (2001) Predicting species' geographic distributions based on ecological niche modeling. *Condor*, **103**, 599-605.

Peterson, A.T., Soberón, J. & Sánchez-Cordero, V. (1999) Conservatism of ecological niches in evolutionary time. *Science*, **285**, 1265-1267.

Raxworthy, C.J., Martinez-Meyer, E., Horning, N., Nussbaum, R.A., Schneider, G.E., Ortega-Huerta, M.A. & Peterson, A.T. (2003) Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, **426**, 837-841.

Schmitz, A., Mausfeld, P. & Embert, D. (2004) Molecular studies on the genus *Eumeces* Weigmann, 1834: Phylogenetic relationships and taxonomic implications *Hamadryad*, **28**, 73-89.

Sievert, G. & Sievert, L. (2005) *A Field Guide to Oklahoma's Amphibians and Reptiles*. Oklahoma Department of Wildlife Conservation, Oklahoma City, Oklahoma.

Stockwell, D. & Peters, D. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, **13**, 143-158.

Werler, J. E. & Dixon, J. R. (2000) *Texas Snakes: Identification, Distribution, and Natural History*. University of Texas Press, Austin.

Wiens, J.J. & Graham, C.H. (2005) Niche conservatism: intregating evolution, ecology, and conservation biology. *Annual Review of Ecology and Systematics*, **36**, 519-539.

Zug, G.R., Vitt, L.J. & Caldwell, J.P. (2001) *Herpetology: An Introductory Biology of Amphibians and Reptiles*, 2 ed. Academic Press, San Diego.

Table 1: Statistical parameters for the 10 models in the best subset selection. N = Total number of points used in the modelling exercise. χ^2 = Chi-square statistics. P = Probability that a random prediction has the same number of correct predicted points as the one generated by DesktopGARP®. Commission = Percentage of the predicted area that exceeds the recorded occurrence. Omission (Int) = Intrinsic omission, the percentage of training points that are predicted absent but are presence records. Omission (Ext) = Extrinsic omission, the percentage of test points that are predicted absent, but are presence records. Values represent the average of the 10 models in the best subset selection.

Species	N	(χ^2)	P	Commission	Omission (int)	Omission (ext)
Eumeces fasciatus	323	795.52	<0.01	15.39	3.90	4.47
E. obsoletus	200	799.7	<0.01	10.70	0.32	1.61
Gastrophryne olivacea	260	928.7	<0.01	11.23	1.94	4.19
G. carolinensis	427	1953.7	<0.01	9.25	0.09	3.05
Rana blairi	343	1350.6	<0.01	10.61	0.00	2.92
R. sphenocephala	415	1353.6	<0.01	12.64	1.45	2.42
Scaphiopus couchii	108	596.2	<0.01	7.98	0.00	1.86
S. hurterii	50	543.8	<0.01	3.99	0.00	4.80
Tantilla gracilis	204	924.8	<0.01	9.20	0.05	3.82
T. nigriceps	160	671.4	<0.01	10.04	0.00	2.64
Sistrurus miliaris	109	549.25	<0.01	8.53	0.67	2.40
S. catenatus	94	1184.2	<0.01	3.43	0.00	2.91

Figure Legends

Figure 1. Predicted results of using niche modelling to investigate factors that may affect distributions of closely related species along an environmental gradient. (A) Species do not overlap in their predicted distributions. (B) Both species occur in the predicted overlap zone. (C) Only one species occupies the predicted overlap zone.

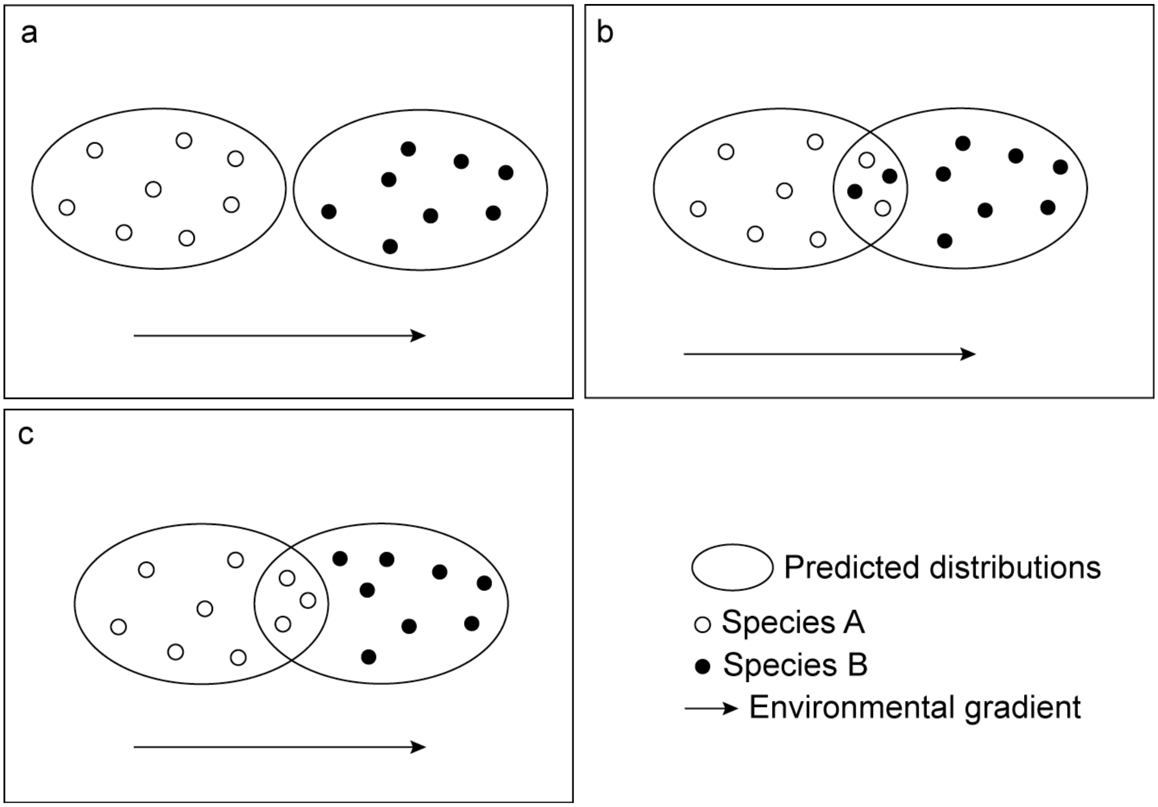
Figure 2. Relationships of six of the 20 climatic–environmental variables used in the niche modelling along a longitudinal axis across the continental United States (this subset of variables was chosen specifically to illustrate the longitudinal gradient). A – Altitude. B – Mean annual precipitation. C – Precipitation of the driest quarter. D – Precipitation seasonality. E – Minimum temperature of the coldest month. F – Temperature seasonality. Data from Worldclim (http://www.worldclim.org/). In the graphs, each point represents the mean value for eight latitudes (ranging from 33ºN to 40ºN) within the same longitude; X -axis is longitude and Y-axes are environmental variables. Temperature variables are in degrees Celsius x 10, precipitation variables are in mm, and altitude is in meters above sea level. Dashed lines represent the longitudes in which the state of Oklahoma lies.

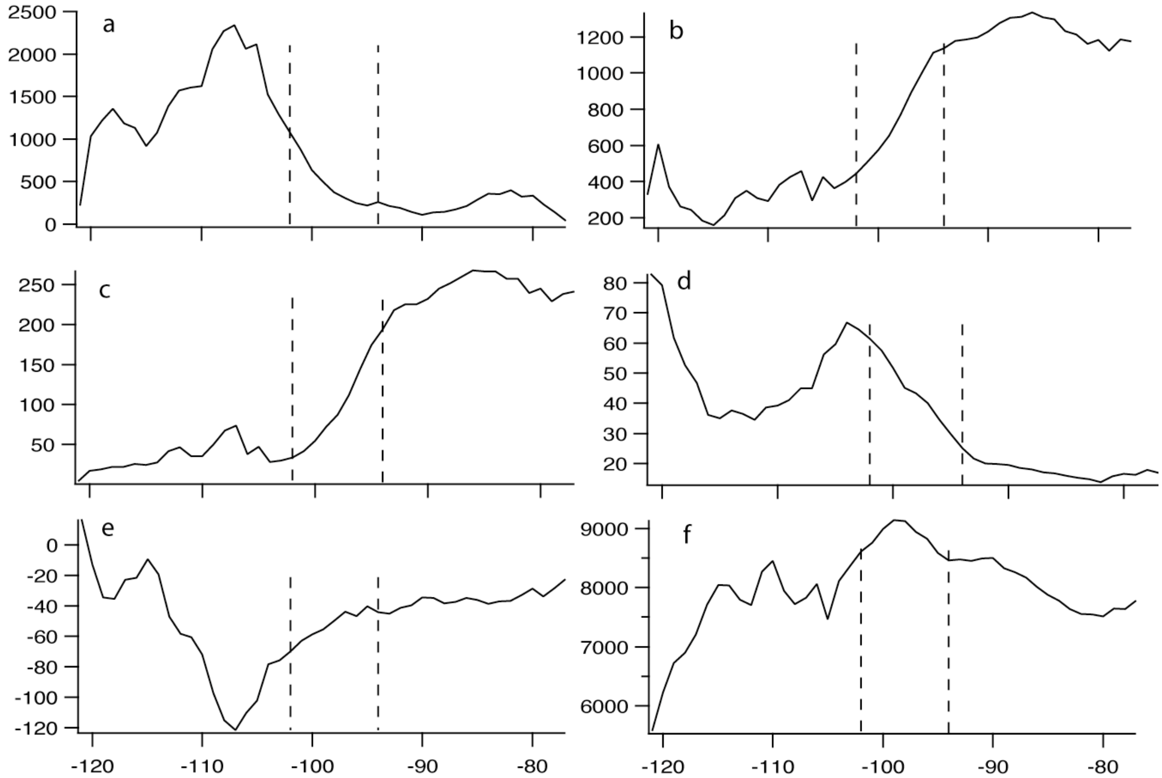
Figure 3. Relationships of six of the 20 climatic–environmental variables used in this study along a longitudinal axis within the state of Oklahoma, illustrating the east-west gradient. A – Altitude. B – Mean annual precipitation. C – Precipitation of the driest quarter. D – Precipitation seasonality. E – Minimum temperature of the coldest month. F – Temperature seasonality. Data from Worldclim (http://www.worldclim.org/). In the

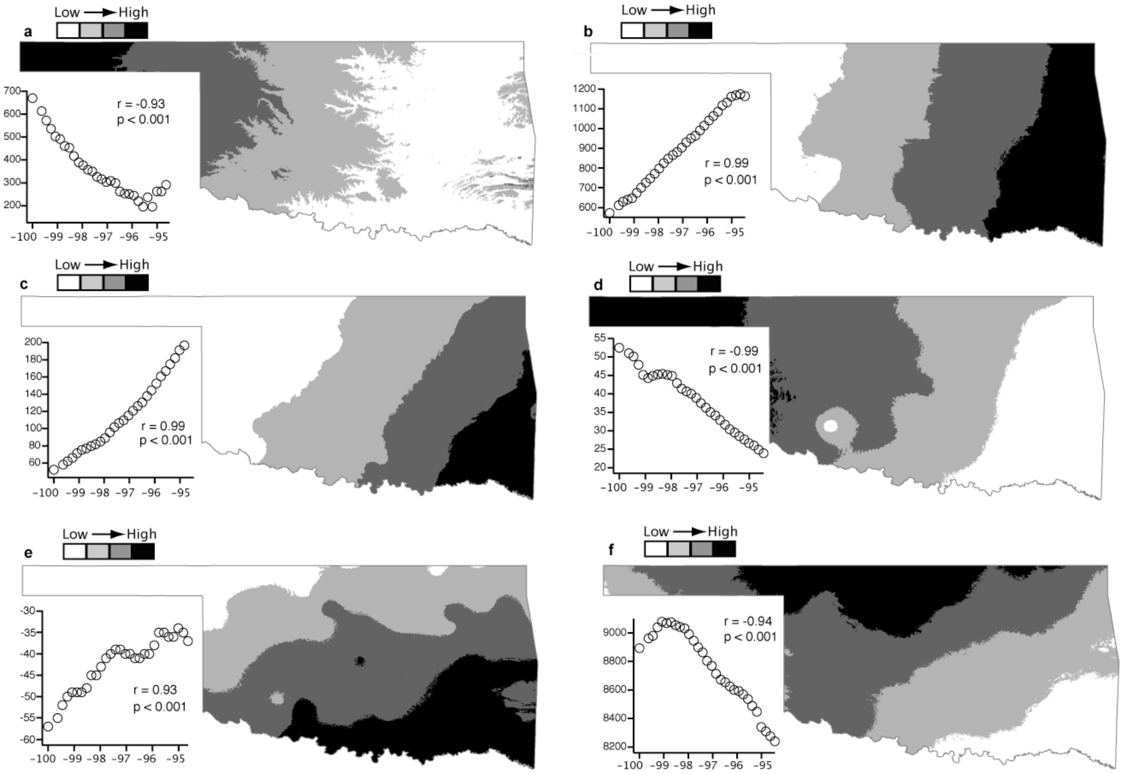
graphs, each point represents the mean value for eight latitudes within Oklahoma (ranging from 34ºN to 37ºN) within the same longitude; X-axis is longitude and Y-axes are environmental variables. Temperature variables are in degree Celsius x 10, precipitation variables are in mm, and altitude is in meters above sea level.

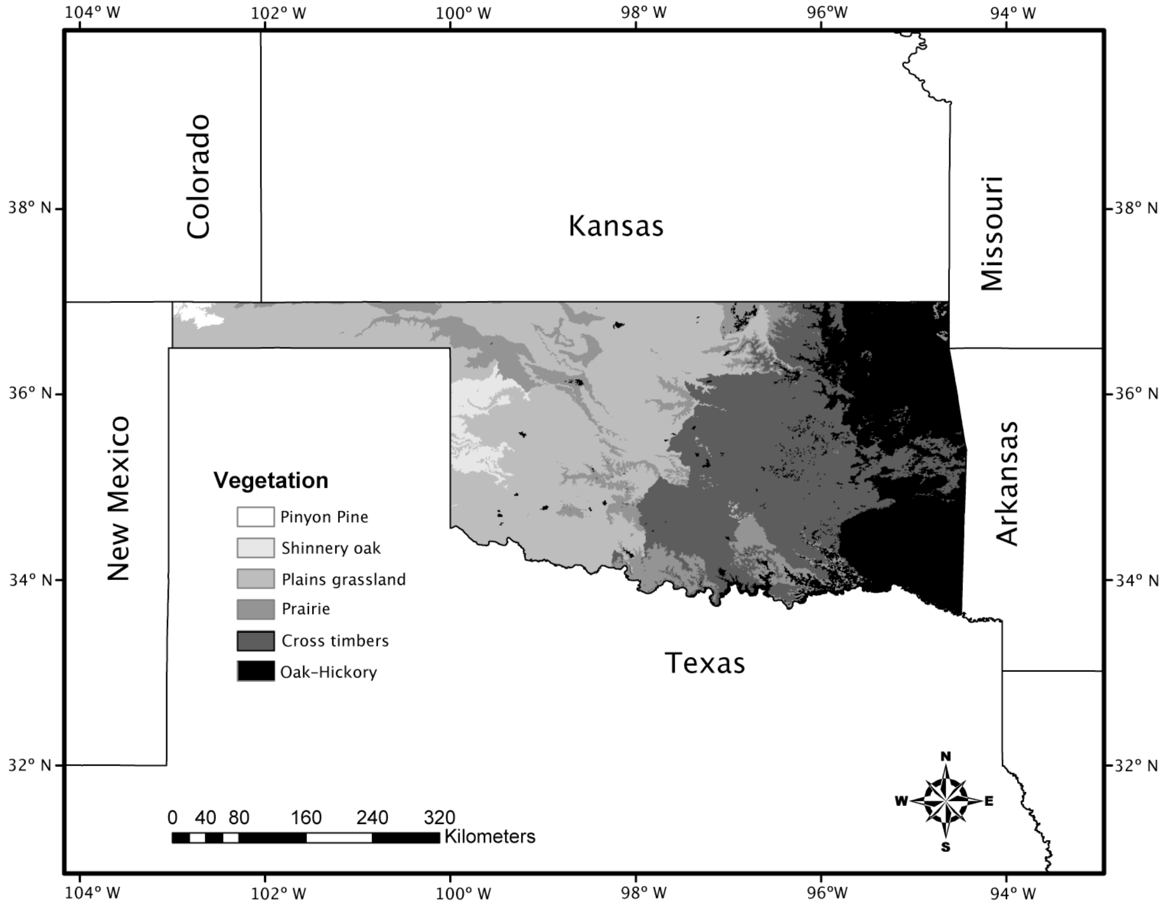
Figure 4. Map of the south–central United States showing the state of Oklahoma, and the major vegetation types of the region.

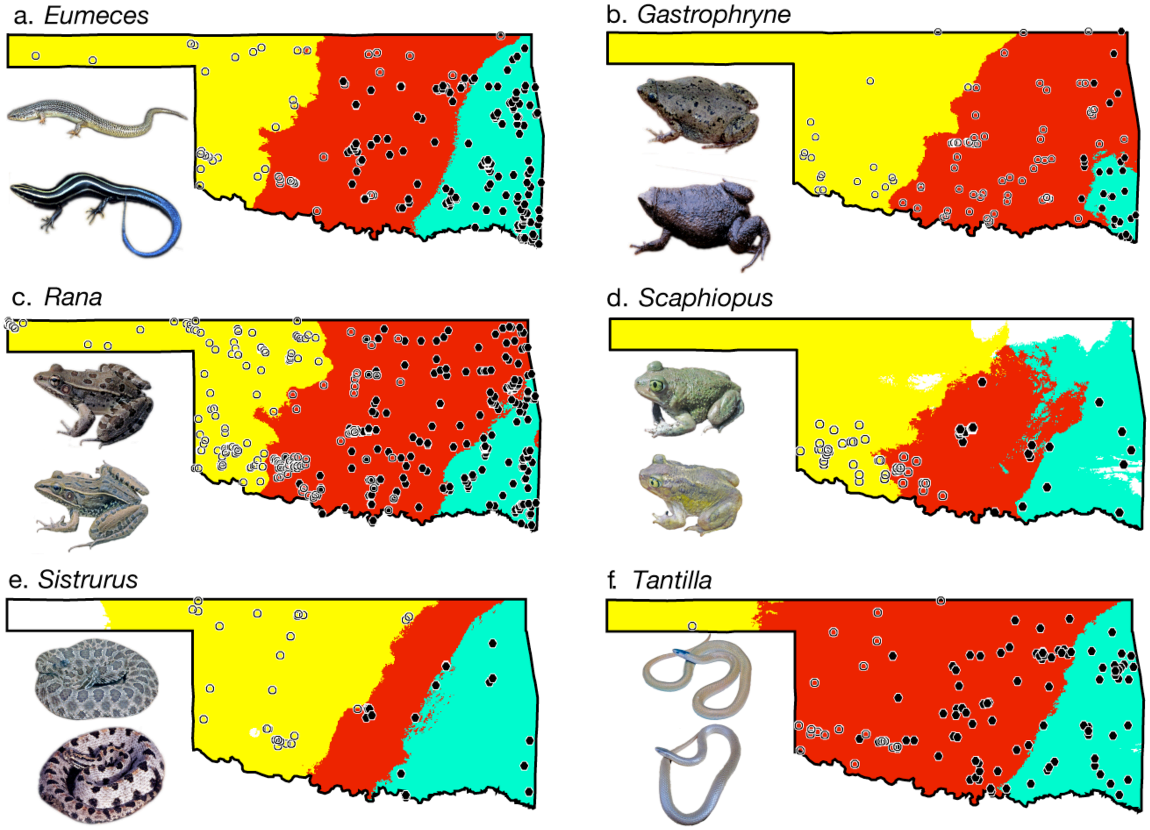
Figure 5. Niche modelling maps for the six species pairs. For each species pair, open circles on each map represent known localities for the species generally restricted to the western part of the state, whereas closed circles represent localities for the eastern species. Yellow represents the predicted distribution of the western species, blue represents the predicted distribution of the eastern species, and red is the predicted overlap zone based on the models. Photographs next to each map correspond to the western species (above) and the eastern species (below). (A) *Eumeces obsoletus–E. fasciatus*. (B) *Gastrophryne olivacea–G. carolinensis*. (C) *Rana blairi–R.sphenocephala*. (D) *Scaphiopus couchii–S. hurterii.* (E) *Sistrurus catenatus–S. miliarius*. (F) *Tantilla nigriceps–T. gracilis*.


a
b
c
Predicted distributions
Species A
Species B
Environmental gradient


a
2500
2000
1500
1000
500
0
b
1200
1000
800
600
400
200
c
250
200
150
100
50
d
80
70
60
50
40
30
20
e
0
-20
-40
-60
-80
-100
-120
f
9000
8000
7000
6000
-120
-110
-100
-90
-80
-120
-110
-100
-90
-80


a
Low → High
r = -0.93
p < 0.001
700
600
500
400
300
200
-100
-99
-98
-97
-96
-95
b
Low → High
r = 0.99
p < 0.001
1200
1100
1000
900
800
700
600
-100
-99
-98
-97
-96
-95
c
Low → High
r = 0.99
p < 0.001
200
180
160
140
120
100
80
60
-100
-99
-98
-97
-96
-95
d
Low → High
r = -0.99
p < 0.001
55
50
45
40
35
30
25
20
-100
-99
-98
-97
-96
-95
e
Low → High
r = 0.93
p < 0.001
-30
-35
-40
-45
-50
-55
-60
-100
-99
-98
-97
-96
-95
f
Low → High
r = -0.94
p < 0.001
9000
8800
8600
8400
8200
-100
-99
-98
-97
-96
-95


104° W
102° W
100° W
98° W
96°W
94° W
38° N
36° N
34° N
32° N
Colorado
Kansas
Missouri
New Mexico
Arkansas
Texas
Vegetation
Pinyon Pine
Shinnery oak
Plains grassland
Prairie
Cross timbers
Oak-Hickory
0 40 80 160 240 320
Kilometers
N
W
E
S


a. *Eumeces*
b. *Gastrophryne*
c. *Rana*
d. *Scaphiopus*
e. *Sistrurus*
f. *Tantilla*

Chapter IV

Biogeography of the Amazon molly: ecological niche and range limits of an asexual hybrid species

(formatted for *Proceedings of the Royal Society B*)

Species range limits are complex biological phenomena where many factors interact to determine a species distribution. In this study we apply ecological niche modelling to understand relative contributions of environmental factors, dispersal limitations, and biotic interactions in limiting the distribution of the Amazon molly (*Poecilia formosa*), a sperm-dependent unisexual species of hybrid origin. We also used a recently developed metric to calculate the degree of niche overlap between the hybrid and its parental species. ENM produced highly significant models (AUC > 0.99). Annual mean temperature and minimum temperature of the coldest month were the variables that best explained the distribution of the *P. formosa*. Two different processes are acting to limit the distribution of *P. formosa*. At the northern limit of the range, suitable environmental conditions are absent, even though the host species is present further north. At the southern limit, the host species is present and areas with appropriate environmental conditions are present further south, suggesting that dispersal ability is the limiting factor. The niche overlap analysis showed that *P. formosa* is not more similar than expected by chance to either parental species, therefore having its own niche identity.

Keywords: Amazon molly, biogeography, hybrids, niche modelling, niche overlap, range limits

1. INTRODUCTION

One of the greatest challenges in biogeography is to understand factors that shape the distributions of species. What key elements allow a species to colonize certain areas and not others? How do competition, dispersal limitation, and/or environmental requirements interact to determine a species distribution? Understanding ecological and evolutionary dynamics at species range limits can provide valuable insights into a wide range of biological phenomena such as biological invasions (e.g. Peterson 2003), organisms' responses to large-scale environmental fluctuations in the past (Graham et al. 1996) and/or predicted responses in the future due to climate change (Pearson & Dawson 2003).

Species distributions are complex and many factors can interact to determine the limit of a species range (Holt 2003; Holt & Keitt 2005). A widely accepted notion in ecology is that physiological environmental tolerance will be correlated with a species distribution (i.e. a species niche requirements; see Soberón 2007 for a review on niche concepts and their relation to geographic distributions). However, characterizing a species' niche may not be sufficient to explain its range limits. Species can occur in habitats outside their niche because recurrent migration sustains 'sink' populations (Holt 1997; Pulliam 2000) or a species may be absent from habitats within its niche because of dispersal barriers, or the presence of competitors. Also, a species' niche is not necessarily fixed, and species may be able to adapt to different environments, resulting in a change in distribution. Populations can become adapted in response to localized selection regimes, which has been demonstrated during biological invasions where niche shifts allowed a species to spread into areas outside its native niche envelope (Broennimann et al. 2007).

Another scenario that can influence species range limits is when a strong biotic interaction is present such as occurs in parasite/host relationships or in specialized herbivores. In these cases, the distribution of the species will obligatorily depend on the distribution of its host, and strong coevolved patterns in phylogeography and diversification may emerge (Attwood et al. 2007; Toon & Hughes 2008; Whiteman et al. 2007). We use ecological niche modelling on three species of fish that are linked through an unusual biotic relationship: one species, the unisexual hybrid Amazon molly (*Poecilia formosa*) relies on sperm from one of the parental species (either Sailfin molly – *P. latipinna* or Atlantic molly – *P. mexicana*) for triggering embryogenesis. We explore the relative importance of presence of a host for *P. formosa*, abiotic factors, and dispersal limitations to determine the species range limits. We also investigate how similar the niche requirements of the hybrid are to those of the parental species.. In addition, we evaluate a more recent range expansion of *P. formosa* that can be traced to the first introduction of *P. latipinna* as a host in central Texas, followed by the introduction of *P. formosa* about two decades later. We evaluate whether abiotic factors can predict this successful introduction, or whether niche shifts may have occurred to allow the colonization of a novel habitat.

2. MATERIAL AND METHODS

(a) ***Biological system***

Poecilia formosa is a unisexual species of hybrid origin. It reproduces by sperm-dependent parthenogenesis, termed gynogenesis, in which the sperm serves only to stimulate the development of unreduced eggs but normally makes no genetic contribution. The maternal ancestor of this species was *P. mexicana*, and the paternal ancestor was *P. latipinna* (Avise et al. 1991; Schartl et al. 1995; Tiedemann et al. 2005; Lampert & Schartl 2008). Both *P. mexicana* and *P. latipinna* are widely distributed species and occupy different microhabitats. *P. formosa* occupies only a fraction of the areas occupied by its hosts. Understanding why the Amazons do not occupy the entire range of their hosts has remained a puzzle. Previous work has highlighted the importance of the presence of suitable hosts, but the role of abiotic factors was not prominently considered (Schlupp et al. 2002).

In addition, *P. latipinna* was introduced into the San Marcos River in central Texas in the 1930s where it established a successful population (Brown 1953). In the 1950s, a few specimens of *P. formosa* were introduced to the same area and were able to use the *P. latipinna* already present as hosts to establish a new mating complex (Hubbs et al. 1991; Hubbs et al. 1953).

(b) ***Ecological Niche Modelling***

Ecological niche modelling (ENM) uses abiotic and/or biotic variables to model potential distributions of species based on known areas of occurrence (Peterson 2001; Peterson et al. 2002). Many methods are available to perform ENM, all of which apply essentially the same principles. First, georeferenced known-occurrence points of the species under study are linked with variables at each locality. Next, an algorithm uses this information to determine the probability that the species will be found at any point along each variable axis. The information is then combined for all variable axes to generate a 'niche' envelope in which suitable conditions for the species to occur are predicted to exist. This prediction is then projected onto a map of the geographic region of interest (see Elith et al. 2006 for a review on the algorithms). This approach makes it possible to map areas that are environmentally suitable for a species. ENM has become an important tool for studies in ecology, evolution, and conservation biology (Guisan & Thuiller 2005; Kozak & Wiens 2006; Wiens et al. 2006). ENM can also be a powerful tool in biogeography. By examining congruence and/or discordance between potential and actual distributions, researchers may be able to disentangle the roles of ecological and historical factors in shaping species distributions (Anderson et al. 2002; Costa et al. 2008; Swenson 2006). In this study, by using only abiotic variables in the ENM we attempt to contrast the relative importance of environmental variables and biotic interactions (presence of the host).

To generate a map of suitable conditions for *P. formosa* and its parental species, we used the Maxent algorithm, which requires only presence (not absence) species records and has been shown to perform well in comparison with other approaches (Elith et al. 2006). The algorithm works by fitting a probability distribution for species occurrence to the set of pixels across the region of interest. The algorithm is based on the mathematical principle that, given the appropriate constraints, the best explanation to unknown phenomena will maximize the entropy of the probability distribution. In ecological niche modelling, these constraints consist of the values of the environmental variables in the pixels at which the species has been documented. More details on how Maxent works are provided by Phillips et al. (2006) and Phillips et al. (2004). In our modelling exercise, we used the default parameters for Maxent v.3.2.1, which were based on maximizing the results in a diverse set of modelling situations. Details are described in a recent publication (Phillips 2008).

We obtained known locality points for *P. formosa*, *P. mexicana*, and *P. latipinna* from collection databases (http://www.gbif.org/), our own fieldwork, and the literature (Darnell & Abramoff 1968; Miller et al. 2005). When available, we used geographic coordinates directly from museum databases or published studies, and in other instances, we approximated point localities from locality descriptions using georeferencing techniques. We used the environmental variables from the Worldclim project (Hijmans et al. 2005), which are available for download at http://www.worldclim.org. Next, we built a correlation matrix among all variables and selected only variables that were not highly correlated ($r > 0.9$) for the modelling (Table 1). This approach allows us to better interpret the output of the models in terms of the individual contribution of each variable in the model. Although, fish distributions in a smaller spatial scale are likely to be more affected by variables that are related do water properties, at a larger spatial scale such as in this study, macroclimatic variables are likely to be informative and have been successfully applied in niche modelling of fish species (Chen et al. 2007; Domínguez-Domínguez et al. 2006).

To statistically evaluate model performance we used the area under the curve (AUC) on receiver operating characteristic (ROC). ROC analysis is a method designed to evaluate the specificity (absence of commission error) and sensitivity (absence of omission error) of a diagnostic test (Fielding & Bell 1997; Zweig & Campbell 1993). The AUC provides a threshold-independent measure of model performance as compared with that of null expectations (Fielding & Bell 1997), and it is the most commonly used statistic to evaluate ENM performance (Elith et al. 2006). When the AUC is 0.50, the model is performing no better than random. Higher AUC values indicates better models;

and a perfect prediction would have the value 1.0 (Hanley & Mcneil 1982). Although some studies have been pointing out problems with the AUC approach in ENM (Lobo et al. 2008; Peterson et al. 2008; Raes & ter Steege 2007), most of the problems concern the use of AUC in comparing among different ENM methods, which is not the case in the present exercise.

(c) ***Statistical Analysis***

To assess whether the environmental conditions in the native range of *P. formosa*, and *P. latipinna* are different from the conditions in the introduced area in the San Marcos River, we conducted a principal components analysis (PCA) using climatic data extracted from the localities of both species, and from 25 random points extracted within a 50-km buffer from the introduced locality. We also extracted climate data for 25 points in the Florida Lower Peninsula, where the models predicted *P. formosa*, but the species does not occur (see results). The PCA used the correlation matrix of the 16 bioclimatic variables included in the climatic niche models described above. For each 'group' (*P. formosa*, *P. latipinna*, Texas, and Florida) we calculated mean and standard deviation of PC scores along the two first axes (80% of total variation). We tested for differences in the means among groups along the two first PC axes using ANOVAs with post hoc Tukey's HSD tests. Statistical analyses were performed in SAS v.9.1.

To determine whether *P. formosa* has its own niche identity and how similar/dissimilar the niche use of the hybrid and parentals is, we used a new metric developed by Warren et al. (2008) which is calculated in the software ENMTools (available from the authors in Warren et al. 2008). This new metric is based on the Hellingar distance (*H*) (Van der Vaart 1998) and is defined as

$$I(p_X, p_Y) = 1 - \frac{1}{2} H(p_X, p_Y)$$

$$\textit{where}$$

$$H(p_X, p_Y) = \sqrt{\sum_i (\sqrt{p_{X,i}} - \sqrt{p_{Y,i}})^2}$$

where *I* is a similarity index of niche space use for a given species pair. *I* can vary from 0 (no overlap) to 1 (niche models are identical). $p_{X,i}$ (or $\mathrm{p}_{\mathrm{Y,i}}$) denotes the probability assigned by the ENM (output of Maxent in this case) for species *X* (or *Y*) to cell *i*. Due to the nature of ENM models, this new metric has advantages over traditional niche overlap measures of ecological data (e.g. Schoener 1968) because it carries no biological assumptions concerning the meaning of $p_{X,i}$ (or $\mathrm{p}_{\mathrm{Y,i}}$), treating p_X and p_{Y} as probability distributions. For additional information and details on *I* see Warren et al. (2008).

After calculating the values of *I* for the species pairs *P. formosa – P. mexicana,* and *P. formosa – P. latipinna*, we used two different randomization procedures to determine whether or not the ENMs produced for the parental and the hybrids are statistically different. The first procedure randomized the identity of occurrence points while keeping sample sizes constant. For a given species pair comparison *X* ,*Y* with n_X, n_Y occurrences, a set of pseudoreplicate datasets is randomly created by partitioning the pooled set of $n_X + n_Y$ occurrences into new sets of size n_X and n_Y. Then, niche models for each pseudoreplicate dataset are created and the similarity measure *I* is calculated. This process is repeated *m* times to create a null distribution of *I* values. We used $m = 100$ because this is enough to reject the null hypothesis with high confidence while keeping computer processing time feasible. The observed values of *I* are compared to the generated null distribution to evaluate the hypothesis that the niche models for *X* and *Y* are not statistically different. If the observed value falls below 95% of the distribution of simulated values, then null hypothesis is rejected. A significantly lower observed *I* value in the hybrid – parental comparisons would suggest that the hybrid has their own niche identity rather than having equivalent niche requirements to the parental species. This procedure is referred by Warren et al. (2008) as a test for niche equivalency, and it is also implemented in the software ENMTools for details see Warren et al. (2008).

We also used a second randomization procedure that takes into account differences in the environmental background to determine whether hybrid and parentals are more or less similar than expected by chance. In this case, instead of using the pooled dataset of $n_X + n_Y$ occurrences to draw sets of size n_X and n_Y. Sets of n_X occurrences are drawn randomly from the whole region where *Y* may occur and sets of n_Y occurrences are drawn randomly from the region where *X* may occur. We selected the regions where *X* and *Y* may occur by applying an absence/presence threshold to the Maxent output for species *X* and *Y*. Threshold selection in ENM is a research topic in itself (Jiménez-Valverde & Lobo 2007; Liu et al. 2005). We used threshold values where sensitivity (proportion of true positive predictions vs. the number of actual positive sites) is equal to specificity (the proportion of true negative predictions vs. the number of actual negative sites). This approach maximizes agreement between observed and modelled distributions balancing the cost arising from an incorrect prediction against the benefit gained from a correct prediction (Manel et al. 2001; Pearson et al. 2006). In addition, to evaluate whether or not the results of this analysis are influenced by the selection of a specific threshold, we repeated the analysis using different threshold values. Because there was no difference among the different threshold values used, we present only the result of the first analysis.

The null hypothesis that measured niche overlap between hybrid and parentals is explained by regional similarities or differences in available habitat is rejected if the empirically based similarity between hybrid and parentals falls outside the 95% confidence limits of the null distribution. Rejection of the null hypothesis indicates that the niche models of hybrid and parentals are more similar or different (depending on which side of the distribution the observed value falls) than would be expected by chance. Rejection of the null hypothesis also indicates that the observed niche differentiation between hybrid and parental species may be a function of habitat selection and/or suitability rather than simply an artifact of the underlying environmental differences between habitats available to the species. Failure to reject the null hypothesis does not necessarily imply no niche differentiation or niche similarity, rather it may indicate that the sample size or distribution of habitat is such that there is insufficient power to make inferences regarding niche differences between hybrid and parental species. This procedure is referred to by Warren et al. (2008) as a test for niche similarity, and it is also implemented in the software ENMTools (for details see Warren et al. 2008).

3. RESULTS

All models produced were highly significant based on the ROC analysis (all AUC > 0.99, Table 1). For *P. formosa* annual mean temperature (34.5%) and minimum temperature of the coldest month (20.2%) were the variables that contributed the most to the model. For *P. latipinna*, altitude (63.3%) had the greatest individual contribution, whereas in *P.*

mexicana, minimum temperature of the coldest month (69.7%) was the most important variable in the model (Table 1).

The model for *P. formosa* shows that, with the exception of south Florida, not many areas beyond the known distribution of the species contain environmental conditions required by the species (Fig 1A). In the model for *P. latipinna* this scenario was even more extreme, with virtually no areas outside the known distribution of the species having conditions within the species niche envelope. This species seems to be restricted to low altitude areas along the gulf coast from central Mexico to Florida and up the Atlantic coast the southeast United States (Fig 1B). The model for *P. mexicana* also predicted suitable environmental conditions for the species in south Florida, beyond that it mainly described the known range of the species (Fig 1C).

The PCA showed that conditions in the area where *P. latipinna* and *P. formosa* were introduced in Texas are significantly different from conditions present in their native ranges (Fig. 2). The ANOVA on PC1 was significant ($F = 112.6$, $P < 0.01$, with all pairwise Tukey's HSD comparisons $P < 0.05$ except *P. formosa* – Florida and *P. latipinna* – Florida). On PC2, the ANOVA was also significant ($F =46.4$, $P <0.01$ with all pairwise Tukey's HSD comparisons $P < 0.05$ except P. *latipinna* – Florida). Therefore, these analyses demonstrated that despite being geographically more distant, the environmental conditions in Florida are more similar to those in the native range of *P. formosa*, than to environmental conditions in Texas.

The measured niche overlap (*I*) between *P. formosa* and *P. latipinna* was 0.46, and between *P. formosa* and *P. mexicana* was 0.51. Both *I* values were significantly lower ($P < 0.01$) from a null distribution based on the randomization of the identity points

(Fig. 3A–B). These results indicate that the hybrid species has its own niche identity, rather than having niche requirements equivalent to either of the parental species. These values, however, were not statistically lower or higher from what would be expected based on comparison to random predictions taking into account environmental similarities between the regions they occur (Fig. 3C–D).

4. DISCUSSION

Ecological niche modelling usually ignores biotic interactions (e.g., competition) and emphasizes the role of abiotic factors. With respect to the distribution of Amazon mollies, we initially questioned whether that ENM would be able to predict its natural distribution because of the importance of the presence of a suitable host species, which was not accounted for in modelling. Based only on the presence of a suitable host, the Amazon mollies should have a much larger distribution because the distribution of *P. latipinna* extends northward to the distribution of *P. formosa*, and *P. mexicana* occurs continuously in areas hundreds of kilometres further south than the known range of *P. formosa*. Other sperm dependent hybrid species are known to have distributions that more closely match one of their hosts (e.g. *Poeciliopsis* Wetherington et al. 1989, *Ambystoma* Conant 1986, *Rana* Kuzmin 1995). Therefore, it is clear that other factors are acting to restrict the distribution of the Amazon molly. Previous work has suggested that dispersal limitation is the main factor explaining this pattern, highlighting the role of natural barriers such as near coast marine currents and the Sierra Madre Oriental (Schlupp et al. 2002). However, to fully assess the dispersal limitation hypothesis, it is necessary to examine environmental suitability of the areas beyond the known range of *P.*

formosa. If those areas in fact possess the necessary conditions for the species to occur, then the dispersal limitation hypothesis would be more strongly supported. On the other hand, if environmental conditions in those areas are not within the species niche requirements, then dispersal limitation is not what is restricting the species distribution.

Our results showed that in general, *P. formosa* appear to have colonized almost all areas where suitable environmental conditions for the species exist. Only an area in South Florida, which is separated from the continuous natural range by several thousand kilometres, would provide additional suitable habitat. Interestingly it seems that two different processes are acting to limit the distribution of *P. formosa*. No suitable environmental conditions exist for *P. formosa* at the northern limit, even though the host species occurs further north. In contrast, the host species is present at the southern limit, and areas that are predicted to have suitable conditions for *P. formosa* are currently unoccupied by the species. This pattern provides strong evidence for the dispersal limitation hypothesis at the southern limit of *P. formosa* range. *P. formosa* is hypothesized to have originated *ca* 100,000 generations ago in the region near its current southern range limit (Schartl et al. 1995; Schlupp et al. 2002). Therefore, the fact that the species has spread northward considerably more than southward, also corroborates the dispersal limitation hypothesis at the southern limit.

The ENMs do not predict the occurrence of *P. latipinna* or *P. formosa* in the area they were introduced in central Texas. However, these populations appear to be well-established, having co-occurred in the area for more than 50 years (Hubbs et al. 1991; Hubbs et al. 1953). The area in Texas is different climatically from the native ranges of both species, as shown by our PCA and ANOVAs. Therefore, this result does not seem to be a failure of the ENM to identify regions of suitable habitat for the species, but highlights the fact that the species are maintaining viable populations in an area well outside the climate characteristics of their native ranges. Many hypotheses may explain this pattern including phenotypic plasticity and/or adaptation to local environments. Previous work has shown that species may occur outside their native climate range after being introduced in a novel habitat (Broennimann et al. 2007). Another possible explanation is that because of ecological or historical reasons, both *P. formosa* and *P. latipinna* are not occurring in all areas that they could physiologically tolerate. Future laboratory experiments may help to address this question.

Interestingly, our niche overlap analysis showed that the hybrid species overlaps in a similar way with both parental species while still having its own niche identity. However, the niche similarity test showed that the hybrid is not more or less similar to the parental species than would be expected based on comparison to random predictions taking into account environmental similarities between the regions they occur. The analysis of niche overlap of ENMs was conceived and so far applied only to address niche conservatism; that is whether closely related species are more similar in their niche requirements than expected by chance (Warren et al. 2008). In their study, Warren et al. (2008) used several pairs of sister taxa and found that they were more similar than expected by chance. The failure to reject the null hypothesis in the niche similarity test does not necessarily imply a lack of niche differentiation or niche similarity (see methods) between hybrid and parental species. However, this result makes biological sense given models of allopatric speciation by niche conservatism (Wiens 2004; Wiens & Graham 2005). Under these models, speciation will occur in the geographical space without considerable ecological divergence (Kozak & Wiens 2006; Peterson et al. 1999), resulting in sister species being more similar than expected by chance. Hybrids originate by fusion of genomes of two often ecologically divergent species (degree of difference can vary depending on the hybrid). In this case, a certain degree of similarity is expected between hybrids and parental species, but because the hybrid is inheriting characteristics of both parental species, the hybrid may appear to have its own identity while also not being more similar to either one of its parental species than expected by chance.

In summary, we have shown that by applying ENM techniques we can explore factors influencing species range limits and have a better understanding of the relative contribution of different factors such as biological interactions, abiotic factors, and dispersal limitations. Our results regarding the region where the species were introduced have in fact raised more questions about physiological tolerance and niche requirements, and how that translates into predicting geographic range limits. Our study highlights the complexity of studying species range limits, but reiterates that an integrative approach is necessary to understand this phenomenon.

ACKNOWLEDGEMENTS

We thank Don Shepard and Laurie Vitt for comments on the manuscript, and Dan Warren for help in the use of ENM tools software and comments on the manuscript. This work was submitted in partial fulfilment of GCC's Ph.D. degree at the University of Oklahoma. We thank J. Caldwell, T. Rashed, J. Kelly, and M. Kaspari for serving on GCC's doctoral committee and providing valuable comments on the manuscript. GCC is supported by a Fulbright/CAPES PhD fellowship (#15053155–2018/04-7) and a

Graduate Research Assistantship from the Sam Noble Oklahoma Museum of Natural History. We thank the University of Oklahoma for financial support.

REFERENCES

Anderson, R. P., Peterson, A. T. & Gómez-Laverde, M. 2002 Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos* **98**, 3–16.

Attwood, S. W., Fatih, F. A., Mondal, M. M. H., Alim, M. A., Fadjar, S., Rajapakse, R. P. V. J. & Rollinson, D. 2007 A DNA sequence-based study of the *Schistosoma indicum* (Trematoda : Digenea) group: population phylogeny, taxonomy and historical biogeography. *Parasitology* **134**, 2009–2020.

Avise, J. C., Trexler, J. C., Travis, J. & Nelson, W. S. 1991 *Poecilia mexicana* is the recent female parent of the unisexual fish *Poecilia formosa*. *Evolution* **45**, 1530–1533.

Broennimann, O., Treier, U. A., Muller-Scharer, H., Thuiller, W., Peterson, A. T. & Guisan, A. 2007 Evidence of climatic niche shift during biological invasion. *Ecol. Lett.* **10**, 701–709.

Brown, W. H. 1953 Introduced fish species of the Guadalupe River basin. *Tex. J. Sci.* **2**, 245–251.

Chen, P. F., Wiley, E. O. & Mcnyset, K. M. 2007 Ecological niche modeling as a predictive tool: silver and bighead carps in North America. *Biol. Invasions* **9**, 43–51.

Conant, R. 1986 *A field guide to Reptiles and Amphibians (Eastern and Central North America).* Boston, Massachusettts, U.S.A.: Houghton Mifflin.

Costa, G. C., Wolfe, C., Shepard, D. B., Caldwell, J. P. & Vitt, L. J. 2008 Detecting the influence of climatic variables on species' distributions: a test using GIS niche-based models along a steep longitudinal environmental gradient. *J. Biogeogr.* **35**, 637–646.

Darnell, R. M. & Abramoff, P. 1968 Distribution of gynogenetic fish *Poecilia formosa* with remarks on evolution of species. *Copeia*, 354–361.

Domínguez-Domínguez, O., Martinez-Meyer, E., Zambrano, L. & De Leon, G. P. P. 2006 Using ecological-niche Modeling as a conservation tool for freshwater species: Live-bearing fishes in central Mexico. *Conserv. Biol.* **20**, 1730-1739.

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M., Peterson, A. T., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S. & Zimmermann, N. E. 2006 Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151.

Fielding, A. H. & Bell, J. F. 1997 A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* **24**, 38–49.

Graham, R. W., Lundelius, E. L., Graham, M. A., Schroeder, E. K., Toomey, R. S., Anderson, E., Barnosky, A. D., Burns, J. A., Churcher, C. S., Grayson, D. K., Guthrie, R. D., Harington, C. R., Jefferson, G. T., Martin, L. D., McDonald, H. G., Morlan, R. E., Semken, H. A., Webb, S. D., Werdelin, L. & Wilson, M. C. 1996 Spatial response of mammals to late quaternary environmental fluctuations. *Science* **272**, 1601–1606.

Guisan, A. & Thuiller, W. 2005 Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* **8**, 993–1009.

Hanley, J. A. & Mcneil, B. J. 1982 The Meaning and Use of the Area under a Receiver Operating Characteristic (Roc) Curve. *Radiology* **143**, 29–36.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. 2005 Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978.

Holt, R. D. 1997 On the evolutionary stability of sink populations. *Evol. Ecol.* **11**, 723–731.

Holt, R. D. 2003 On the evolutionary ecology of species' ranges. *Evol. Ecol. Res.* **5**, 159–178.

Holt, R. D. & Keitt, T. H. 2005 Species' borders: a unifying theme in ecology. *Oikos* **108**, 3–6.

Hubbs, C., Edwards, R. J. & Garrett, G. P. 1991 An annotated checklist of the freshwater fishes of Texas, with keys to identification of species. *Tex. J. Sci.***43**, 1–56.

Hubbs, C., Kuehne, R. A. & Ball, J. C. 1953 The fishes of the upper Guadelupe River, Texas. *Tex. J. Sci.* **2**, 216–244.

Jiménez-Valverde, A. & Lobo, J. M. 2007 Threshold criteria for conversion of probability of species precense to either-or presence-absence. *Acta Oecol.* **31**, 361–369.

Kozak, K. H. & Wiens, J. J. 2006 Does niche conservatism promote speciation? A case study in North American salamanders. *Evolution* **60**, 2604–2621.

Kuzmin, S. L. 1995 *Die Amphibien Russlands und angrenzender Gebiete*. Magdeburg: Westarp Wissenschaften.

Lampert, K. P. & Schartl, M. 2008 The origin and evolution of a unisexual hybrid: *Poecilia formosa*. *Philos. T. R. Soc. B.* **363**, 2901–2909.

Liu, C., Berry, P. M., Dawson, T. P. & Pearson, R. G. 2005 Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* **28**, 385–393.

Lobo, J. M., Jiménez-Valverde, A. & Real, R. 2008 AUC: a misleading measure of the performance of predictive distribution models. *Global Ecol. Biogeogr.* **17**, 145–151.

Manel, S., Williams, H. C. & Ormerod, S. J. 2001 Evaluating presence-absence models in ecology: the need to account for prevalence. *J. Appl. Ecol.* **38**, 921–931.

Miller, R. R., Minckley, W. L. & Norris, S. M. 2005 *Freshwater fishes of Mexivo*. Chicago, IL: University of Chicago Press.

Pearson, R. G. & Dawson, T. P. 2003 Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecol. Biogeogr.* **12**, 361–371.

Pearson, R. G., Thuiller, W., Araújo, M. B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T. P. & Lees, D. C. 2006 Model-based uncertainty in species range prediction. *J. Biogeogr.* **33**, 1704–1711.

Peterson, A. T. 2001 Predicting species' geographic distributions based on ecological niche modeling. *Condor* **103**, 599–605.

Peterson, A. T. 2003 Predicting the geography of species' invasions via ecological niche modeling. *Q. Rev. Biol.* **78**, 419–433.

Peterson, A. T., Ball, L. G. & Cohoon, K. P. 2002 Predicting distributions of Mexican birds using ecological niche modelling methods. *Ibis* **144**, E27–E32.

Peterson, A. T., Papes, M. & Soberon, J. 2008 Rethinking receiver operating characteristis analysis applications in ecological niche modeling. *Ecol. Model.* **213**, 63–72.

Peterson, A. T., Soberón, J. & Sánchez-Cordero, V. 1999 Conservatism of ecological niches in evolutionary time. *Science* **285**, 1265–1267.

Phillips, S. J. 2008 Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007). *Ecography* **31**, 272–278.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006 Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**, 231–259.

Phillips, S. J. & Dudik, M. 2008 Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**, 161–175.

Phillips, S. J., Dudik, M. & Shapire, R. E. 2004 A maximum entropy approach to species distribution modeling. In *Proceedings of the 21st International Conference on Machine Learning* (ed. R. Greiner & D. Schuurmans), pp. 655–662. Banff, Canada: ACM press.

Pulliam, H. R. 2000 On the relationship between niche and distribution. *Ecol. Lett.* **3**, 349–361.

Raes, N. & ter Steege, H. 2007 A null-model for significance testing of presence-only species distribution models. *Ecography* **30**, 727–736.

Schartl, M., Wilde, B., Schlupp, I. & Parzefall, J. 1995 Evolutionary origin of a parthenoform, the Amazon molly, *P. formosa*, on the basis of a molecular genealogy. *Evolution* **49**, 827–835.

Schlupp, I., Parzefall, J. & Schartl, M. 2002 Biogeography of the Amazon molly, Poecilia formosa. *J. Biog.* **29**, 1–6.

Schoener, T. W. 1968 Anolis lizards of Bimini - Resource partitioning in a complex fauna. *Ecology* **49**, 704–726.

Soberón, J. 2007 Grinnellian and Eltonian niches and geographic distributions of species. *Ecol. Lett.* **10**, 1115–1211.

Swenson, N. G. 2006 Gis-based niche models reveal unifying climatic mechanisms that maintain the location of avian hybrid zones in a North American suture zone. *J. Evol. Biol.* **19**, 717–725.

Tiedemann, R., Moll, K., Paulus K. B. & Schlupp, I. 2005 New microsatellite loci confirm hybrid origin, parthenogenetic inheritance, and mitotic gene conversion in the gynogenetic Amazon molly (*Poecilia formosa*). *Mol. Ecol. Notes* **5**, 586–589.

Toon, A. & Hughes, J. 2008 Are lice good proxies for host history? A comparative analysis of the Australian magpie, Gymnorhina tibicen, and two species of feather louse. *Heredity* **101**, 127–135.

Van der Vaart, A. W. 1998 *Asymptotic Statistics*. Cambridge, U.K.: Cambridge University Press.

Warren, D. L., Glor, R. E. & Turelli, M. 2008 Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution*, 2868–2883.

Wetherington, J. D., Weeks, S. C., Kotorak, E. & Vrijenhoek, R. C. 1989 Genotypic and environmental components of variation in growth and reproduction of fish hemiclones (*Poeciliopsis*: Poeciliidae). *Evolution* **43**, 635–645.

Whiteman, N. K., Kimball, R. T. & Parker, P. G. 2007 Co-phylogeography and comparative population genetics of the threatened Galapagos hawk and three ectoparasite species: ecology shapes population histories within parasite communities. *Mol. Ecol.* **16**, 4759–4773.

Wiens, J. J. 2004 Speciation and ecology revisited: Phylogenetic niche conservatism and the origin of species. *Evolution* **58**, 193–197.

Wiens, J. J. & Graham, C. H. 2005 Niche conservatism: intregating evolution, ecology, and conservation biology. *Annu. Rev. Ecol. Syst.* **36**, 519–539.

Wiens, J. J., Graham, C. H., Moen, D. S., Smith, S. A. & Reeder, T. W. 2006 Evolutionary and ecological causes of the latitudinal diversity gradient in hylid frogs: Treefrog trees unearth the roots of high tropical diversity. *Am. Nat.* **168**, 579–596.

Zweig, M. H. & Campbell, G. 1993 Receiver-operating characteristics (ROC) plots - A fundamental evaluation tool in clinical medicine. *Clin. Chem.* **39**, 561–577.

Table 1 – Variables used in our modelling exercise, their Bioclim code, and a heuristic estimate of the relative contributions for each environmental variable to the Maxent model of each species. Estimation is based on the increase in regularized gain that is added to the contribution of the corresponding variable. Within parenthesis are the AUC values of the model for each species.

Variable	Code	Species		
		P. formosa (0.997)	*P. latipinna* (0.994)	*P. mexicana* (0.992)
Annual mean temperature	Bio 1	34.5*	0.1	0.4
Mean diurnal range	Bio 2	0.2	2.0	1.9
Isothermality	Bio 3	0.3	0.2	1.9
Temperature seasonality	Bio 4	15.1	5.8	12.5
Max. temperature of warmest month	Bio 5	0.0	0.1	0.0
Min. temperature of coldest month	Bio 6	20.2*	9.2	69.7*
Mean temperature of wettest quarter	Bio 8	2.7	7.0	0.3
Mean temperature of driest quarter	Bio 9	0.6	4.4	0.2

Mean temperature of warmest quarter	Bio 10	0.0	0.0	0.8
Annual precipitation	Bio 12	0.4	0.0	0.5
Precipitation seasonality	Bio 15	3.8	1.3	3.1
Precipitation of wettest quarter	Bio 16	0.3	0.1	0.1
Precipitation of driest quarter	Bio 17	11.9	2.7	6.8
Precipitation of warmest quarter	Bio 18	0.0	3.8	1.2
Precipitation of coldest quarter	Bio 19	5.8	0.0	0.4
Altitude	Alt	4.2	63.3*	0.1

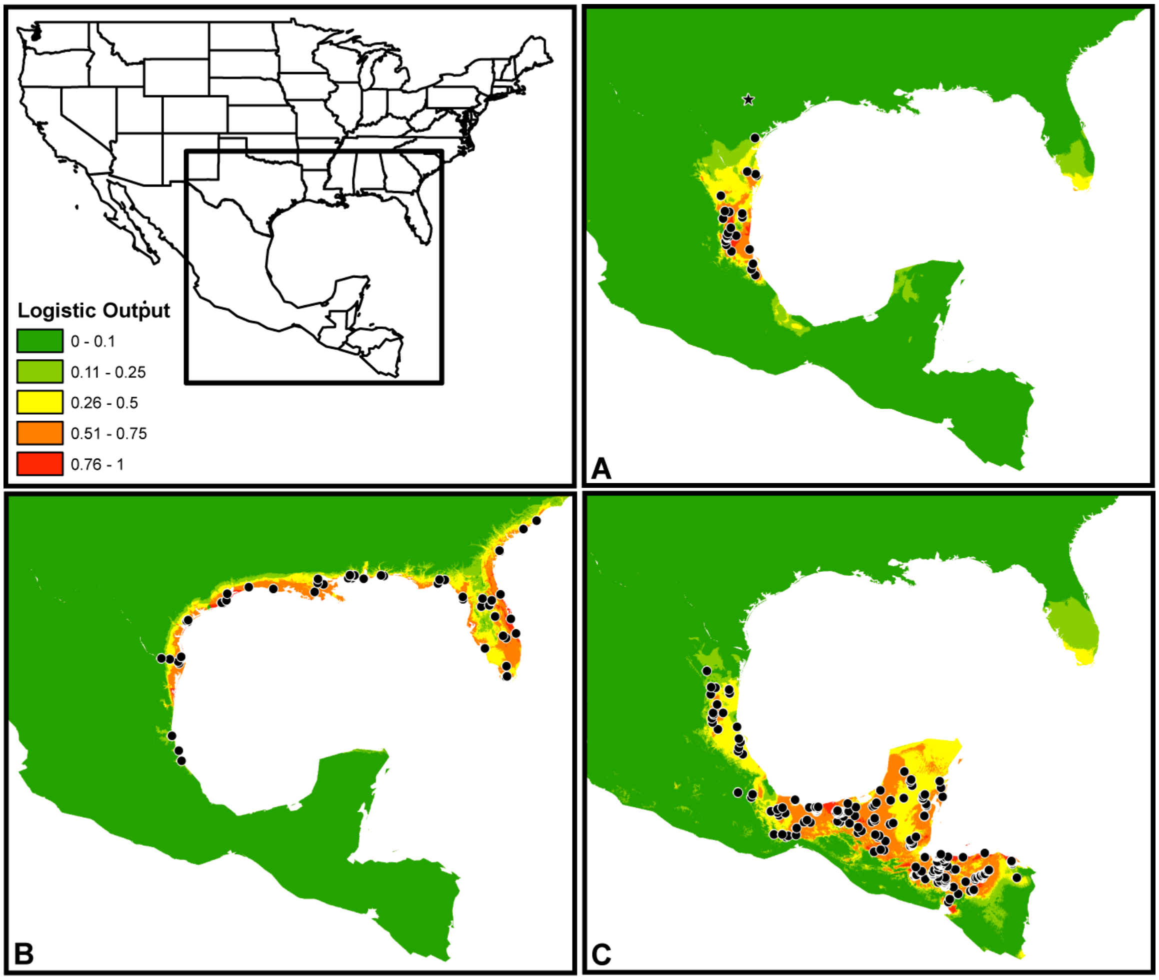
* Variables with the most important relative contribution to the model for the species.

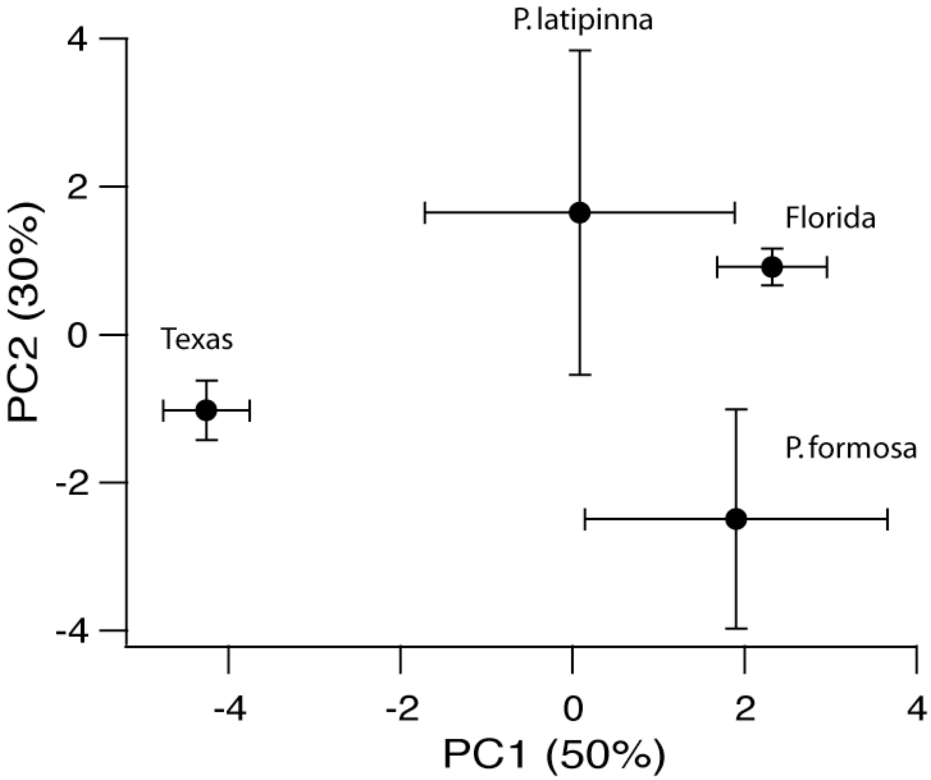
Figure Legend

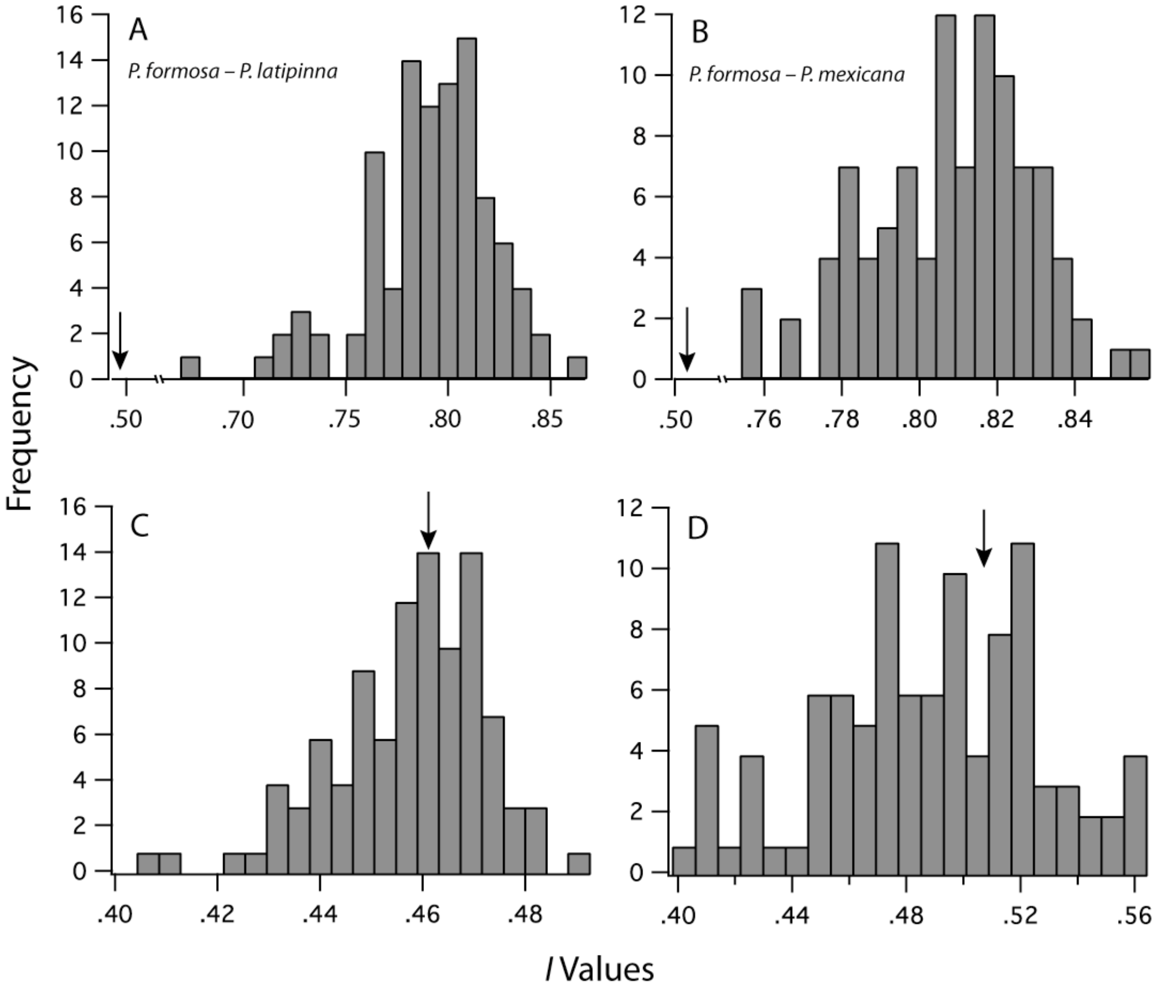
Figure 1 – Results of ENM for (A) *P. formosa*; (B) *P. latipinna*; and (C) *P. mexicana*. Black dots are the locality points used in the modelling. Maxent default output format is called logistic and depicts probability values ranging from 0-1, where higher values indicates higher environmental suitability and therefore higher probability of species occurrence (Phillips & Dudik 2008).

Figure 2 – Principal component analyses (PCAs) of the climate niche space occupied by *P. formosa*, *P. latipinna*, areas in Florida where *P. formosa* is absent but predicted, and areas in central Texas where both *P. formosa* and *P. latipinna* where introduced. The mean principal component scores and standard deviations are shown. First principal component explains 50% of the variation and was mainly influenced by minimum temperature of the coldest month, and temperature seasonality. Second principal component explains 30% of the variation and was mainly influenced by annual precipitation, and maximum temperature of the warmest month.

Figure 3 – Results of comparisons of observed values of *I* (Indicated by arrows) to null distributions generated by 100 pseudoreplicates. A – Results for the niche equivalency test (see methods) comparing observed values of *I* in *P.formosa – P. latipinna* ($P < 0.01$), and B – *P.formosa – P. mexicana* ($P < 0.01$) to the distribution of null values of *I*, C – Results of niche similarity test (see methods) comparing observed values of *I* in *P.formosa – P. latipinna* ($P = 0.72$), and D – *P.formosa – P. mexicana* ($P = 0.79$) to the distribution of null values of *I*.


Logistic Output
0 - 0.1
0.11 - 0.25
0.26 - 0.5
0.51 - 0.75
0.76 - 1
A
B
C


P. latipinna
Florida
Texas
P. formosa
PC2 (30%)
PC1 (50%)
4
2
0
-2
-4
-4
-2
0
2
4


Frequency
A
P. formosa – P. latipinna
16
14
12
10
8
6
4
2
0
.50
.70
.75
.80
.85
B
P. formosa – P. mexicana
12
10
8
6
4
2
0
.50
.76
.78
.80
.82
.84
C
16
14
12
10
8
6
4
2
0
.40
.42
.44
.46
.48
D
12
10
8
6
4
2
0
.40
.44
.48
.52
.56
I Values