

ENVIRONMENTAL NICHE EQUIVALENCY VERSUS CONSERVATISM: QUANTITATIVE APPROACHES TO NICHE EVOLUTION

Dan L. Warren,^{1,2} Richard E. Glor,^{3,4} and Michael Turelli^{1,5}

¹*Department of Evolution and Ecology, University of California, Davis, California 95616*

²*E-mail: danwarren@ucdavis.edu*

³*Department of Biology, University of Rochester, Rochester, New York, 14627*

⁴*E-mail: rglor@mail.rochester.edu*

⁵*E-mail: mturelli@ucdavis.edu*

Received June 4, 2007

Accepted July 7, 2008

Environmental niche models, which are generated by combining species occurrence data with environmental GIS data layers, are increasingly used to answer fundamental questions about niche evolution, speciation, and the accumulation of ecological diversity within clades. The question of whether environmental niches are conserved over evolutionary time scales has attracted considerable attention, but often produced conflicting conclusions. This conflict, however, may result from differences in how niche similarity is measured and the specific null hypothesis being tested. We develop new methods for quantifying niche overlap that rely on a traditional ecological measure and a metric from mathematical statistics. We reexamine a classic study of niche conservatism between sister species in several groups of Mexican animals, and, for the first time, address alternative definitions of "niche conservatism" within a single framework using consistent methods. As expected, we find that environmental niches of sister species are more similar than expected under three distinct null hypotheses, but that they are rarely identical. We demonstrate how our measures can be used in phylogenetic comparative analyses by reexamining niche divergence in an adaptive radiation of Cuban anoles. Our results show that environmental niche overlap is closely tied to geographic overlap, but not to phylogenetic distances, suggesting that niche conservatism has not constrained local communities in this group to consist of closely related species. We suggest various randomization tests that may prove useful in other areas of ecology and evolutionary biology.

KEY WORDS: *Anolis*, climate, Cuba, ecology of speciation, ecological niche modeling, invasive species, Isthmus of Tehuantepec, niche conservatism.

The degree to which ecological niches are conserved (i.e., shared among closely related taxa) carries implications for a range of ecological and evolutionary phenomena, from the role of ecology in speciation (Graham et al. 2004; Kozak and Wiens 2006; Peterson and Nyári 2007) to the expected response by invasive species to novel habitats (Broennimann et al. 2007). The development of environmental niche models (ENMs), which are generated by combining species occurrence data with environmental GIS data layers, makes it possible to quantify niches at an unprece-

ented scale (reviewed in Kozak et al. 2008). Although much can be gained simply by plotting predicted distributions or examining the expected suitability of various environments, statistical tests are essential to interpreting the significance of these patterns. Unfortunately, existing statistical methods for assessing similarity of ENMs (among populations or species, for example) or testing specific hypotheses about niche conservation and its consequences tend to be ad hoc or conceptually ambiguous. This methodological problem has likely contributed to the absence of general

conclusions about niche conservation (Peterson et al. 1999; Peterson and Holt 2003; Graham et al. 2004; Knouft et al. 2006; Kozak et al. 2008; Pearman et al. 2008).

Consider two recent studies that use analyses of niche conservation to assess the prevalence of ecological divergence during speciation. After finding that sister taxa share similar environmental niches, Peterson et al. (1999) conclude that “. . .speciation takes place in geographic, not ecological, dimensions and that ecological differences evolve later.” Graham et al. (2004), meanwhile, find that closely related species of dendrobatid frogs have significantly different environmental niches and suggest that “. . .differential selection likely played an important role in species differentiation. . . .” Although such disagreement may reflect real differences in the nature of speciation across clades, it may also result from the strikingly different comparative methods these studies used to quantify and define niche conservation.

Peterson et al.’s (1999) conclusion that niches are conserved is based on a test of “niche similarity,” which asks whether ENMs from sister species predict one another’s known occurrences better than expected under the null hypothesis that they provide absolutely no information about one another’s ranges (see also Peterson and Nyári 2007; Kambhampati and Peterson 2007). Under this permissive definition, niche conservation is expected as the inevitable consequence of phylogenetic relationships and a finite rate of evolutionary divergence. Meanwhile, Graham et al. (2004) conduct a test of “niche equivalency” when they ask whether the niches of related species are effectively indistinguishable (see also Knouft et al. 2006; Pfenninger et al. 2007). If we view niche conservation as a continuum (ranging from niches that are identical to niches that are more similar than random), Graham et al.’s test of “niche equivalency” lies at the opposite end of the spectrum from Peterson et al.’s test of “niche similarity.” Because it is generally expected that the degree of niche conservation will fall somewhere between these extremes, it is not difficult to see how conflicting conclusions might result from testing only one extreme hypothesis or the other. Biologically meaningful conclusions about the prevalence of ecological divergence during speciation (and other hypotheses related to niche conservation) require standard metrics for describing similarity among ENMs and for testing well-defined and comparable null hypotheses across studies.

Here we propose new niche similarity metrics that may be used to test both of the extreme hypotheses considered above. We apply these metrics and tests to data from Peterson et al.’s (1999) groundbreaking analysis of sister species of Mexican animals separated by the Isthmus of Tehuantepec (a dry, scrubby area that is very different from the moist mountainous areas on either side (see Fig. 2 below)). Through this analysis, we are able to show that related species tend to occupy environmental niches that are similar (i.e., support for niche similarity), but rarely identical

(i.e., rejection of niche equivalency). In addition to challenging Peterson et al.’s (1999) assertion that speciation across the Isthmus of Tehuantepec is not associated with ecological differentiation, our results reinforce Wiens and Graham’s (2005) conclusion that “simply testing whether niches are conserved is not by itself particularly helpful or interesting and that a more useful focus is on the patterns that niche conservatism may (or may not) create.”

We also use our new niche similarity metrics and randomization tests to address the potential influence of niche conservatism on community structure from a phylogenetic perspective. Specifically, we ask whether the degree of environmental niche conservatism seen in a group constrains local community composition. Because the ENMs used in our analyses are constructed at a geographic (rather than microenvironmental) scale, they are more representative of a species’ β -niche (i.e., niches that differ among communities) than its α -niche (i.e., niches that differ within communities) (Silvertown et al. 2006; Ackerly and Cornwell 2007). Divergence between ENMs of two closely related species is therefore not expected to affect local community assembly by alleviating competition. It may, however, dictate whether these species are distributed among rather than within local communities characterized by shared β -niche characteristics. If environmental niches are strongly conserved, communities with shared β -niche characteristics may be comprised of particularly closely related species (Webb et al. 2002; Johnson and Stinchcombe 2007). We do not expect this pattern of “phylogenetic clustering,” however, if environmental niches are highly labile (Knouft et al. 2006). To test whether the degree of niche conservatism observed in a group restricts closely related species to communities that share similar β -niche characteristics, we integrate our niche divergence metrics with a molecular phylogenetic tree. We explore niche evolution in a clade of allopatrically, parapatrically, and sympatrically distributed species belonging to the adaptive radiation of West Indian *Anolis* lizards. Our results confirm the high degree of environmental niche lability observed in a previous study of this group (Knouft et al. 2006) and suggest it is unlikely that environmental niche conservation has restricted the most closely related anoles to occurring in a shared β -niche.

Materials and Methods

GENERATING ECOLOGICAL NICHE MODELS

Comparative studies of niche evolution using ENMs involve: (1) construction of ENMs and (2) comparison of resulting models. The methods we develop focus on the second step and may be used with any method of ENM construction that produces suitability measures in geographic space. We focus on ENMs constructed via maximum entropy using the program Maxent (Phillips et al. 2006) because this approach performed particularly well in a recent comparison of alternative ENM construction methods (Elith et al.

2006). This method produces continuously varying, nonnegative suitability scores for each cell in a specified geographic region (here, and in subsequent analyses, we refer to the “raw,” rather than the more commonly presented “cumulative,” Maxent scores).

In some cases, we evaluate the sensitivity of comparative conclusions derived from Maxent results by also using GARP (Stockwell and Peters 1999), an alternative method of ENM construction used by Peterson et al. (1999). Niche reconstruction using GARP was performed using the DesktopGARP program, which produces binary output by labeling each cell “suitable” or “unsuitable” (Stockwell and Peters 1999). GARP output is often given as a consensus of many independent runs, which may be reported either as a binary prediction (with some threshold consensus level considered a predicted presence, e.g., Pearson et al. 2007) or multivalued (number or proportion of models predicting presence in a cell, e.g., Elith et al. 2006). Likewise, continuous Maxent output can be converted to binary predictions by establishing a threshold suitability score to count as a predicted presence. Although many useful criteria for establishing such thresholds have been proposed (Fielding and Bell 1997; Liu et al. 2005; Jiménez-Valverde and Lobo 2007), measures of niche overlap can be sensitive to this decision (Supporting Appendix S1, also see Loiselle et al. 2003 for a similar phenomenon regarding predicted distributions). Indeed, using thresholds can obscure important biological detail and, under some conditions, be positively misleading about patterns of niche overlap. In our analyses, therefore, thresholds were used only for the chi-square tests, which require binary predictions.

Output from both Maxent and GARP, which assign nonnegative values to each cell in a study region, can be converted to probability distributions on geographic space by dividing the individual cell values by the sum of all values. For binary output, these values constitute a uniform probability distribution over all cells predicted to be suitable. For the continuous output from Maxent, we can obtain relative suitability scores for each cell by normalizing the raw scores. This step is necessary because raw scores are calculated only with respect to the sum of scores across the specific background datapoints used in model construction, rather than the entire geographic area in question.

NEW METRICS TO QUANTIFY NICHE SIMILARITY

A metric for quantifying similarity of ENMs is critical to testing patterns of niche evolution. Although several metrics have been proposed, they are subject to serious methodological shortcomings. Peterson et al. (1999), for example, used the magnitude of their chi-square statistic to assess degree of niche similarity between species; but this statistic is compromised by its dependence on sample sizes and predicted range sizes (as illustrated below), and by requiring binary predictions (presence vs. absence). Peterson and Holt (2003) proposed the “niche differentiation index,”

which asks how well an ENM from one species predicts the range(s) of other species relative to its own distribution; however, they did not suggest how this metric can be assessed statistically. Most recently, Stockman et al. (2008) introduced a simulation-based method that may be used to investigate patterns of niche overlap, but this method too is dependent on overly simplistic binary predictions.

Because we know of no a priori justification for any particular measure of environmental-niche differentiation or similarity, we consider two alternatives. Both assume probability distributions defined over geographic space, in which $p_{X,i}$ (or $p_{Y,i}$) denotes the probability assigned by the ENM for species X (or Y) to cell i . Our first similarity metric comes from the ecological literature; Schoener’s (1968) statistic for niche overlap,

$$D(p_X, p_Y) = 1 - \frac{1}{2} \sum_i |p_{X,i} - p_{Y,i}|, \quad (1)$$

ranges from 0 (niche models have no overlap) to 1 (niche models identical). This metric was chosen for its simplicity, long history of use, and the fact that it permits direct comparison to traditional measures of niche similarity that focus on microhabitat and/or diet. However, it may suggest a biological interpretation of $p_{X,i}$ that is unwarranted. Schoener’s D is typically applied with values of $p_{X,i}$ that reflect relative use of particular microhabitats and/or prey items. In contrast, for ENMs, we have no assurance that the $p_{X,i}$ are proportional to local species densities or any other measure of relative use. This motivates using a second metric that carries no biological assumptions concerning the meaning of the $p_{X,i}$, but simply treats p_X and p_Y as probability distributions. Many measures have been developed in mathematical statistics for comparing probability distributions. Here we consider Hellinger distance (van der Vaart 1998, p. 211), defined as

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

Hellinger distances lie between 0 and 2. They have previously been used in ecological studies, primarily in comparing community composition across sites (Legendre and Gallagher 2001). To compare Hellinger-based results to more conventional ecological measures of niche overlap, we propose the similarity statistic

$$I(p_X, p_Y) = 1 - \frac{1}{2} H(p_X, p_Y), \quad (3)$$

which ranges from 0 (no overlap) to 1 (niche models identical). The similarity between the calculations for I and D is such that we rarely see qualitative differences in the results obtained using these two metrics. That being the case, using D with ENMs is unlikely to be misleading despite the (possibly incorrect) implicit assumption about the meaning of $p_{X,i}$. Nevertheless, measurements of D from

ENMs may not be comparable to measurements of D from more traditional ecological data.

An alternative to the methods presented here is to consider overlap between models of habitat suitability in the space of environmental variables, rather than focusing on the projection of those models onto geographic space (e.g., Graham et al. 2004). This approach may facilitate understanding the evolution of ecologically relevant phenotypic characters. However, because not all combinations of environmental variables are represented in nature, there may be large regions of environmental space for which we have no presence or pseudoabsence data. As a result, models may over- or under-predict in these regions of environmental space without penalty, and these artifacts will be transferred to calculations of niche overlap (see Supporting Appendix S2 for an empirical example). By measuring overlap using the projection of the niche model onto the set of environmental variables actually present in nature, we restrict ourselves to regions of environmental niche space in which our models are most likely to accurately describe species' differences. These concerns apply to measurements of niche overlap using ENMs, but not to measurements of distance in environmental space (e.g., Broennimann et al. 2007). The randomization tests presented below may be applied to measures of distance in environmental space.

STATISTICAL TESTS OF NICHE CONSERVATISM

Our descriptive metrics can be used in statistical tests of the two alternative hypotheses addressed by the Peterson et al. (1999) and Graham et al. (2004) studies of the role of niche conservation in speciation: (1) the hypothesis of niche equivalency, rejected by Graham et al. (2004), predicts no statistically significant differences between alternative niche models, and (2) the hypothesis of niche similarity, supported by Peterson et al. (1999), predicts that niches are more similar than expected by chance under a specific null model. For niche similarity, we evaluate several null models, with different biological interpretations.

Are two niche models identical?

Tests of niche equivalency

Several previous studies of niche evolution test the hypothesis of niche equivalency via standard multivariate statistical analyses after extracting environmental data from known localities or predicted presence localities from an ENM (e.g., Graham et al. 2004). We address the same question via a randomization test that relies on the metrics D (eq. 1) and I (eq. 3). For a pair of species with n_X (n_Y) occurrences of species X (Y), we create pseudoreplicate datasets by randomly partitioning the pooled set of $n_X + n_Y$ occurrences into sets of size n_X and n_Y . Niche models are created from

each pseudoreplicate and these are compared using the similarity measures I and D . This process is repeated m times to create a null distribution of similarity values, denoted $\{\tilde{I}_i\}$ and $\{\tilde{D}_i\}$. (We use $m = 100$ in the applications below, because this typically suffices to reject the null hypothesis with high confidence.) The observed values of $I(p_X, p_Y)$ and $D(p_X, p_Y)$ are compared to the percentiles of these null distributions in a one-tailed test to evaluate the hypothesis that the niche models for X and Y are not statistically significantly different. This process was automated with a Perl script, available from the authors by request. The probabilities obtained would accurately reflect the probabilities under the null hypothesis if the sampling of each species was unbiased with respect to its environmental tolerances. These randomization procedures test niche conservatism in the strictest sense, the effective equivalency of the environmental niche between species.

Are two niche models similar?

Tests of niche similarity

Rather than testing niche-model equivalency between sister species, Peterson et al. (1999) and Martinez-Meyer and Peterson (2006) ask whether one species' niche model predicted the occurrences of a second species better than expected by chance. Peterson et al. (1999) and Martinez-Meyer and Peterson (2006) tested the ability of the model for species X to predict species Y by comparing the number of occurrences of Y correctly predicted from X, denoted $c_X(Y)$, to the number of correct predictions expected under a specific null hypothesis.

Because the details of this method have never been fully described, we provide a brief review. Let N_Y denote the number of cells in the study area considered for species Y, and let $n_X(Y)$ denote the number of those cells in which species Y is predicted to occur based on the ENM developed for species X. Let o_Y denote the number of cells in which species Y is known to occur. If the niche model for species X has no relation to the occurrences of species Y, the probability of occurrence of species Y in any cell would be simply $q_X(Y) = n_X(Y)/N_Y$, irrespective of the environmental parameters in that cell. This is equivalent to a uniform Bayesian prior in which the environmental conditions that determine the distribution of species X say nothing about the environmental conditions that determine the distribution of species Y. Under this extreme null hypothesis, the distribution of the number of correctly predicted cells, $c_X(Y)$, would be binomial with parameters N_Y and $q_X(Y)$. Martinez-Meyer and Peterson (2006) assessed niche conservation by using this binomial distribution, based on the "no information" null hypothesis, to calculate the probability that the number of correct predictions would be at least as great as the observed value, $c_X(Y)$, by chance. In contrast, Peterson et al. (1999) used a χ^2 test, with the expected number of correct predictions being $N_Y q_X(Y)$, as in the Martinez-Meyer

and Peterson (2006) “binomial test.” The Peterson et al. (1999) test statistic is

$$\chi^2(Y|X) = \frac{[c_X(Y) - o_Y q_X(Y)]^2}{o_Y q_X(Y)} = \frac{\left[c_X(Y) - o_Y \left(\frac{n_X(Y)}{N_Y} \right) \right]^2}{o_Y \left(\frac{n_X(Y)}{N_Y} \right)} \quad (4)$$

which is compared to a chi-square distribution with one degree of freedom. Because these binomial-based null distributions depend on binary predictions, they cannot be applied to ENMs that generate continuous output without establishing a threshold suitability value to count as a predicted presence (cf. Kozak and Wiens 2006).

Peterson et al. (1999) also used an alternative, and perhaps more reasonable, test to assess niche conservatism. They compared niche similarity (assessed by the magnitude of the chi-square statistic in eq. 4) between sister species to niche similarity between confamilial taxa. Following Peterson et al. (1999), we also assess niche conservatism by comparing the overlaps seen between each pair of sister species to their average niche overlap with a confamilial outgroup.

We propose a randomization test that generalizes the chi-square test of Peterson et al. (1999) and applies to ENMs that make either binary (presence vs. absence) or multivalued (e.g., relative suitability) predictions. Like the Peterson et al. (1999) test, our test asks whether one species' ENM predicts that of a second species better than expected by chance, where the null distribution is chosen using study-specific criteria. However, unlike their test, our “background similarity” test takes into account the level of local environmental heterogeneity by incorporating environmental differences between species occurrence locations and the surrounding space.

As in the Peterson et al. (1999) test, let N_X and N_Y denote the number of cells in the study areas considered for species X and Y, and let o_X and o_Y denote the number of cells in which each species is known to occur. Our similarity test compares the actual similarity of the ENMs for X and Y (as assessed by I or D values) to the distribution of similarities obtained by comparing the ENM for X to an ENM obtained by choosing o_Y cells at random from among the N_Y cells in the study area for Y. We then repeat the test in the opposite direction, comparing I or D values between X and Y to the I or D values representing Y's overlap with an ENM derived from randomly chosen points from the study area for X. This analysis is repeated 100 times in each direction to construct an expected distribution of I or D values between an ENM generated using actual occurrence data and one generated from random datapoints. Because results outside of the null distribution in either direction are both biologically plausible and interesting, we treat this as a two-tailed test. The null hypothesis, that measured niche overlap between species is explained by regional similarities or differences in available habitat, is rejected if the actual similarity

between two species falls outside of the 95% confidence limits of the null distribution. This test is similar to the Peterson et al. (1999) chi-square test, but allows the use of multivalued ENMs. The spatial prediction made under the null hypothesis also differs in a subtle, but potentially significant, way; although the chi-square test models the expected predictive success of a random model, the test we present uses the predictive success of an actual ENM built from random points. Because of the spatial autocorrelation seen in most environmental variables, these ENMs are expected to produce predicted suitabilities that are spatially autocorrelated even when the data used to build them are drawn at random from the environmental background.

This test serves two complementary purposes. Rejection of the null hypothesis indicates that the niche models of two species are more similar (or different) than would be expected by chance, with a definition of “chance” that is less naive about the underlying distribution of environmental variables than either the equivalency or chi-square tests presented above. Rejection of the null also indicates that the observed niche differentiation between species is a function of habitat selection and/or suitability rather than simply an artifact of the underlying environmental differences between the suite of habitats available to the two species. This may be particularly important in studies like Peterson et al. (1999) that focus on multiple comparisons across the same geographic barrier. Failure to reject the null hypothesis does not necessarily imply no niche differentiation/conservatism, rather it may indicate that the sample size or distribution of habitat is such that there is insufficient power to make inferences regarding niche evolution.

PHYLOGENETIC ANALYSIS OF NICHE DIFFERENTIATION

Our metrics for niche comparison can be integrated with phylogenetic trees. Most traditional comparative methods are designed to analyze traits of individual tip taxa, not pairwise comparisons between tips. Although it is possible to choose comparisons that are phylogenetically independent either by restricting our analysis to sister taxa or by adopting the method proposed by Maddison (2000), both methods discard substantial amounts of data, and, in our analysis, severely limit our ability to infer patterns resulting from niche conservatism. We therefore illustrate two ways of integrating our niche similarity metric with phylogenies. The first uses a Mantel test (e.g., Knouft et al. 2006). The null hypothesis under this test is that the phylogenetic distance between two species is unrelated to the distance between these same two species' ENMs. We implement this test using R (<http://www.r-project.org/>) and the *ade4* package (Thioulouse et al. 1997) to generate 1000 replicates.

Our second approach, which we use to summarize our data, adapts the method Fitzpatrick and Turelli (2006) used to analyze phylogenetic patterns of range overlap as a function of divergence

time. This method produces topologically weighted averages for comparisons across internal nodes in a phylogeny. It neither attempts to reconstruct ancestral states nor requires complete taxon sampling. The proposed weighting is simply a heuristic attempt to weight the comparisons across each node to take into account shared ancestry, rather than averaging all comparisons equally, as done in Coyne and Orr (1989) (cf. Bolnick and Near [2005] for a similar method that incorporates branch lengths). The resulting pattern can be analyzed via resampling to detect phylogenetic signal (cf. Fitzpatrick and Turelli 2006).

Data and Analyses

POINT OCCURRENCE AND GIS DATA

Mexican animals

For our reanalysis of Peterson et al.'s (1999) data, occurrence points were obtained for 78 species representing sister pairs of birds (16), butterflies (5), and mammals (5) separated by the Isthmus of Tehuantepec in Mexico, and for one confamilial species for each sister pair. The occurrence data were generously provided by the authors of that study. These 26 sets of three species (see Appendix 1) constitute the subset of those used in the original analysis for which we could obtain at least one occurrence point for each species. Groups with small ($n < 15$) sample sizes were retained, both because some groups with comparable sampling were used in the original study and because we wanted to examine the behavior of alternative methods when data are limited and ENMs may be unreliable. The exclusion of these groups does not materially affect our conclusions regarding niche evolution. Because resampling of duplicate occurrences can lead to pseudoreplicate niche models that are more similar to each other than they should be (resulting in a bias toward detecting significant niche differentiation), we trimmed our occurrence points to a maximum of one per grid cell for each species. We also removed occurrence points for which we did not have data on one or more of the GIS layers representing an environmental variable (hereafter "climate layers").

Cuban anoles

For our analysis of niche evolution in Cuban anoles, occurrence points were obtained from a previous study (Knouft et al. 2006). ENMs were constructed using the methods discussed previously. From this study, we also obtain an ultrametric phylogeny based on a Bayesian analysis of ~ 1500 bp of mitochondrial DNA. For our Mantel tests, we obtain three matrices. The first is comprised of phylogenetic distances from the ultrametric phylogeny. The second consists of niche similarities (I or D) for each species pair. The third consists of measures of range overlap generated in ArcGIS from minimum convex polygons enclosing known occurrence points and trimmed to the boundaries of Cuba. We use

these matrices to test two predictions of the hypothesis that niche conservatism constrains community composition. First, we test the basic assumption that range overlap covaries with environmental niche similarity. After finding that it does, we ask whether similarity of ENMs is associated with phylogenetic relatedness. The phylogeny and number of occurrence points for each species is given in the Appendix 2.

GIS data

GIS layers used in ENM construction for Mexican taxa included climate data from Worldclim (Hijmans et al. 2005) and potential vegetation data from CONABIO (Rzedowski 1990). ENMs for this study were constructed using three sets of climate layers: (1) the four layers used by Peterson et al. (1999) (elevation, mean temperature, mean precipitation, and potential vegetation, hereafter referred to as the "PSS" layers); (2) the first 19 "Bioclim" layers, which represent a subset of the Worldclim global climate layers that are thought to be among the most biologically relevant, and are commonly used in ENM construction (see <http://cres.anu.edu.au/outputs/anuclim/doc/bioclim.html> for details), and (3) a set of layers derived from the Bioclim layers using Principal Components Analysis (PCA). The PCA layers were used to compensate for possible overparameterization of niche models due to the large number of variables traditionally used in Bioclim analyses. Following PCA in ArcGIS (ESRI 2006), we retained layers for the two axes with the largest eigenvalues. These axes represent 99.3% of the variation in the original 19 layers within Mexico, whereas axes with smaller eigenvalues represented less than 0.2% each. To facilitate comparison with the results obtained by Peterson et al. (1999), we restricted analyses to the political boundaries of Mexico. This may have some effect on the results obtained, as some of the species are known to occur outside of this area. Following Peterson et al. (1999), the PSS layers were analyzed at a lower spatial resolution than the PCA and Bioclim layers. This could induce bias toward detecting niche equivalency with the PSS layers, but should not affect the comparison of overlap between sister species to their overlap with an outgroup. Given our results, there is little reason to suspect that this difference in resolution affected our analyses.

For the Knouft et al. (2006) data, models were constructed using the first four principal components of the 19 Bioclim layers obtained from Worldclim (representing $> 99\%$ of the variation in the original dataset for Cuba), with the addition of normalized difference vegetation index (NDVI) layers from the Global Land-cover Facility (Hansen et al. 1998). NDVI layers are constructed from satellite imagery, and represent an estimate of the sparseness of vegetation in each grid cell. For this study, ENMs were constructed using Maxent, whereas those in the original study were constructed using WhyWhere (Stockwell 2006).

GENERATION OF ENVIRONMENTAL NICHE MODELS

Maxent

For the Mexican animals, construction of ENMs and randomization tests were performed for each sister-taxon pair using all three sets of GIS data layers. Apart from using raw suitability scores, Maxent runs used default values for all program settings. The models produced were generally good predictors of species occurrences as measured by the area under the receiver–operating characteristic curve (AUC), a commonly used, but possibly misleading (Peterson et al. 2007; Lobo et al. 2008) metric for goodness of fit for ENMs. Most had AUC scores greater than 0.80, with a few exceptions that appear to be related to small sample size (*Chlorostilbon salvini*) or a cosmopolitan distribution (*Colaptes cafer*, *Mimus polyglottos*). AUC scores for all Maxent ENMs are given in Supporting Appendix S3. Because no sampled absence data were available, these AUC scores represent the ability of the model to distinguish presence data from background data, rather than distinguishing presence from absence (Phillips et al. 2006). AUC scores for these models were calculated using training data, which may produce misleading results (Fielding and Bell 1997). Precise model evaluation is not integral to the present work, however, and the similarity of results between Maxent and GARP in the chi-square analyses, despite large differences in predicted distributional area (see below), suggests that even a fairly substantial lack of precision in model predictions is unlikely to alter our conclusions. This may not be generally true, however, and should not be construed as an indication that model evaluation is unimportant in measuring niche overlap. There is an ongoing debate over whether Maxent models tend to overfit the input data (Peterson et al. 2007; Phillips 2008). If this bias is real and is present in the Maxent models built for this study, the primary effect would be to reduce measures of niche overlap. This would produce a pattern consistent with that seen in our identity test, but our results are unlikely to be due to this potential source of bias for two reasons: first, the difference between observed overlap and expected overlap if niches were equivalent was in most cases quite extreme, so that the bias would have to be very strong to explain it; and second, the results of our background test for similarity are opposite to those expected if a strong bias toward overfitting was reducing estimates of overlap.

GARP

We constructed GARP distribution models using the four layers from the Peterson et al. (1999) study. GARP analyses were conducted using Desktop GARP version 1.1.3, available at <http://nhm.ku.edu/desktopgarp/index.html>. Execution details for the original study of Mexican fauna were not available, so our GARP analysis is modeled after the procedure used in Elith et al. (2006). For each species, we conducted 100 runs using a convergence limit of 0.01 and 1000 maximum iterations. We then selected the best subset of models based on a hard threshold of

20% omission with a commission threshold of 50%. All other settings were left at their default values. This procedure resulted in a set of 10 models for each species.

For the chi-square test, we chose individual threshold suitabilities for each species using the minimum suitability score that correctly predicted a known occurrence. The resulting binary ENMs produced by Maxent were considerably larger than those produced by GARP, with an average ratio of areas of 1.93. This pattern was statistically significant (*t* test, $P < 0.01$). This difference has the potential to affect the outcome of the chi-square tests, as the larger area predicted by the Maxent ENMs increases the fraction of known occurrences that must be correctly predicted to obtain statistical significance. However, as the overall results of the analyses using the two methods do not differ significantly, we do not feel that this difference is a cause for concern. The binary-transformed Maxent and GARP results were tested for cross-species prediction following the methods of Peterson et al. (1999). These calculations were performed using ArcGIS 9.2 (ESRI 2006).

Results

ARE TWO NICHE MODELS IDENTICAL? TESTS OF NICHE EQUIVALENCY

Using both the *I* and *D* metrics, we reject the hypothesis of niche equivalency between sister species in nearly every case (Table 1). A majority of the exceptions to this pattern are due to sample sizes that are too small to produce a significant result (i.e., when $n_i < 5$, minimum possible $P = 0.0625$). The only disagreement between the *I* and *D* results is for *Falco sparverius*–*Falco tropicalis* using the PSS layers, where statistically significant niche differentiation was detected using *D*, but the difference was not quite significant ($P < 0.06$) using *I*. Our conclusions about niche equivalency also appear insensitive to the layers used to generate ENMs.

ARE TWO NICHE MODELS SIMILAR? TESTS OF NICHE SIMILARITY

Peterson et al.'s chi-square test

To facilitate comparison with the results of Peterson et al. (1999), we begin by testing the hypothesis of niche similarity using their chi-square test statistic as applied to ENMs generated by both GARP and Maxent. This analysis suggests that most species pairs are significantly similar, regardless of whether GARP or Maxent is used (Table 2). We also tested similarity using the binomial distribution, with similar results.

Randomization test of background similarity

Tests for similarity between sister species based on the background similarity test reveal that sister species are in most cases more similar than expected by chance. Although some of the exceptions to this pattern may be due to small sample size, two comparisons with large sample sizes for both species showed that

Table 1. Tests of niche equivalency. A significant value denotes a pair of species that are ecologically distinct (*, $P \leq 0.05$, **, $P \leq 0.01$, ns, $P > 0.05$). Results for I and D metrics are given for three sets of layers (Bioclim, PCA, and the Peterson et al. [1999] layers, denoted PSS). Statistical significance for I and D are via randomization tests (see text). N_a and N_b values are sample sizes for the first and second species used in a comparison, respectively.

Species Pair	N_a, N_b	I			D		
		Bioclim	PCA	PSS	Bioclim	PCA	PSS
Butterflies							
<i>abderus-electryon</i>	55,16	0.39**	0.48**	0.57**	0.08**	0.21**	0.33**
<i>charops-nigricans</i>	39,16	0.39**	0.43**	0.59**	0.08**	0.14**	0.37**
<i>eunoe-chamula</i>	10,6	0.44**	0.58**	0.54**	0.14**	0.34**	0.26**
<i>leonilae-nigricans</i>	11,16	0.36**	0.40**	0.56**	0.08**	0.10**	0.36**
<i>nimbice-ochracea</i>	143,13	0.43**	0.58**	0.73ns	0.14**	0.38**	0.60ns
Birds							
<i>beryllina-devillei</i>	426,82	0.48**	0.60**	0.64**	0.21**	0.43**	0.44**
<i>heloisa-elliotti</i>	116,11	0.46**	0.60**	0.63**	0.18**	0.39**	0.41**
<i>canivetii-salvini</i>	209,1	0.53ns	0.61ns	0.68ns	0.25ns	0.37ns	0.54ns
<i>cafer-mexicanoides</i>	441,49	0.37**	0.42**	0.51**	0.07**	0.10**	0.23**
<i>coronata-ridgwayi</i>	139,65	0.45**	0.53**	0.61**	0.15**	0.23**	0.40**
<i>occidentalis-flavescens</i>	217,88	0.50**	0.58**	0.67**	0.23**	0.34**	0.55**
<i>ruber-versicolor</i>	208,28	0.46**	0.50**	0.59**	0.17**	0.19**	0.38**
<i>fulgens-viridiceps</i>	326,46	0.44**	0.50**	0.64**	0.14**	0.23**	0.46**
<i>sparverius-tropicalis</i>	538,6	0.59**	0.64**	0.72ns ($P < 0.06$)	0.39**	0.49**	0.53*
<i>pustulatus-sclateri</i>	512,109	0.51**	0.58**	0.62**	0.25**	0.36**	0.44**
<i>polyglottos-gilvus</i>	469,139	0.41**	0.50**	0.54**	0.09**	0.23**	0.30**
<i>chrysopeplus-aurantiacus</i>	111,46	0.41**	0.48**	0.51**	0.13**	0.21**	0.28**
<i>jardinii-sanctorum</i>	273,57	0.39**	0.47**	0.57**	0.09**	0.18**	0.34**
<i>obsoletus-neglectus</i>	237,6	0.34**	0.38**	0.37**	0.05**	0.08**	0.06**
<i>varia-fulvescens</i>	11,13	0.52ns	0.79ns	0.61**	0.26*	0.73ns	0.41**
<i>assimilis-leucachen</i>	450,61	0.51**	0.64**	0.68**	0.25**	0.49**	0.51**
Mammals							
<i>lepturus-lophurus</i>	12,10	0.59**	0.80ns	0.60**	0.33**	0.69ns	0.35**
<i>melanocarpus-zarhynchus</i>	388,23	0.36**	0.43**	0.43**	0.06**	0.15**	0.14**
<i>colliaei-yucatanensis</i>	116,40	0.35**	0.54**	0.49**	0.03**	0.24**	0.19**
<i>aztecus-minor</i>	20,3	0.56ns	0.58ns	0.62ns	0.32ns	0.39ns	0.39ns
<i>megalops-zarhynchus</i>	18,23	0.37**	0.58**	0.51**	0.08**	0.35**	0.24**
Proportion showing niche equivalency		0.12	0.15	0.15	0.08	0.15	0.12

the species distributed southeast of the isthmus had no better ability to predict the niche model of the species in the northwest than expected based on overall environmental similarity between the two regions. This implies that ecological differentiation between these species may reflect differences in the availability of habitat in their respective regions rather than an actual change in habitat preference. No sister pairs differed more than expected based on random sampling of regional differences in environmental variables.

Outgroup comparison

Comparison of overlap between sisters to their average overlap with a confamilial species yields varying results depending on which environmental coverages are used, but does not depend strongly on the choice of I or D to measure overlap. With strictly

abiotic coverages, we do not find significantly more similarity between sisters than between nonsisters (Table 3). When potential vegetation is included, however, we find that sisters are more similar than nonsisters for most comparisons, and that this is a statistically significant pattern regardless of which metric we use ($P < 0.001$ using Student's t , $P < 0.005$ using binomial). We conducted a similar test in which we compared overlap between sister species to the maximum (rather than the mean) overlap between one of the sister species and the reference confamilial species. No statistically significant patterns were seen in this analysis, even when the potential vegetation layer was included.

PHYLOGENETIC ANALYSIS OF NICHE EVOLUTION

Our integrated analysis of ENMs and a molecular phylogeny for Cuban anoles is summarized in Figure 1. Because a Mantel test

Table 2. Tests of niche similarity. Comparison of observed similarity to similarity expected by chance using the chi-square test of Peterson et al. (1999), with both GARP and binary-transformed Maxent ENMs, and our background similarity test using both *I* and *D*. Results are for PSS layers only. Results for chi-square tests are chi-square values for “south predicting north, north predicting south,” each followed by an assessment of statistical significance (*, $P \leq 0.05$, **, $P \leq 0.01$, ns, $P > 0.05$). Results for background similarity tests are the measured overlap between the pair followed by assessments of significance, given as “south predicting north, north predicting south.” For the summary in the final line, species are considered more similar than expected under the null hypothesis if the comparison is statistically significant in either direction.

Species Pair	N_a, N_b	χ^2 test		Background test	
		GARP	Maxent	<i>I</i>	<i>D</i>
Butterflies					
<i>abderus-electryon</i>	55,16	0.7ns,58.93**	0.01ns,91.89**	0.57ns,**	0.33**,**
<i>charops-nigricans</i>	39,16	8.48**,.66.27**	2.76ns,69.09**	0.59**,**	0.37**,**
<i>eunoe-chamula</i>	10,6	12.58**,.35.11**	1.39ns,10.3**	0.54ns,ns	0.26ns,ns
<i>leonilae-nigricans</i>	11,16	7.48**,.20**	3.29ns,15.12**	0.56**,**	0.36**,**
<i>nimbice-ochracea</i>	143,13	1.47ns,63.05**	0.03ns,63.97**	0.73**,**	0.6**,**
Birds					
<i>beryllina-devillei</i>	426,82	9.05**,.406.51**	0.06ns,418.41**	0.64**,**	0.44**,**
<i>heloisa-elliotti</i>	116,11	1.88ns,230.56**	0.02ns,121.19**	0.63**,**	0.41**,**
<i>canivetii-salvini</i>	209,1	0ns,0ns	0ns,233.13**	0.68**,**	0.54**,**
<i>cafer-mexicanoides</i>	441,49	137.29**,.92.93**	0.64ns,100.67**	0.51**,**	0.23**,**
<i>coronata-ridgwayi</i>	139,65	117.02**,.189.37**	53.27**,.88.81**	0.61**,**	0.4**,**
<i>occidentalis-flavescens</i>	217,88	68.49**,.94.96**	2.91ns,169.36**	0.67**,**	0.55**,**
<i>ruber-versicolor</i>	208,28	51.93**,.277.66**	14.18**,.112.45**	0.59**,**	0.38**,**
<i>fulgens-viridiceps</i>	326,46	39.28**,.217.61**	4.4*,.270.09**	0.64**,**	0.46**,**
<i>sparverius-tropicalis</i>	538,6	2.52ns,12.03**	0ns,11.46**	0.72ns,**	0.53ns,**
<i>pustulatus-sclateri</i>	512,109	13.41**,.488.92**	0.37ns,438.11**	0.62**,**	0.44**,**
<i>polyglottos-gilvus</i>	469,139	17.01**,.18.56**	0.39ns,9.07**	0.54**,**	0.3**,**
<i>chrysopeplus-aurantiacus</i>	111,46	2.39ns,59.97**	0.4ns,57.11**	0.51**,**	0.28**,**
<i>jardinii-sanctorum</i>	273,57	101.19**,.104.3**	39.59**,.101.72**	0.57**,**	0.34**,**
<i>obsoletus-neglectus</i>	237,6	41.44**,.031ns	0.89ns,1.08ns	0.37ns,ns	0.06ns,ns
<i>varia-fulvescens</i>	11,13	49.33**,.24.34**	50.72**,.25.76**	0.61**,**	0.41**,**
<i>assimilis-leucachen</i>	450,61	0.09ns,462.27**	0.04ns,555.74**	0.68**,**	0.51**,**
Mammals					
<i>lepturus-lophurus</i>	12,10	3.23ns,48.26**	7.87**,.15.92**	0.6**,*	0.35**,**
<i>melanocarpus-zarhynchus</i>	388,23	45.3**,.1.14ns	1.18ns,2.01ns	0.43ns,**	0.14ns,**
<i>colliaei-yucatanensis</i>	116,40	3.52ns,25.18**	0.02ns,24.48**	0.49ns,*	0.19ns,**
<i>aztecus-minor</i>	20,3	3.81ns,2.72ns	0.92ns,29.75**	0.62*,ns	0.39*,ns
<i>megalops-zarhynchus</i>	18,23	46.28**,.46.61**	5.34*,.39.53**	0.51**,**	0.24**,**
Proportion showing significant similarity		1.0	0.92	0.77, 0.88	0.81, 0.88

Table 3. Results for tests of niche conservatism using outgroup comparisons with the same sets of climate layers as in Table 1. Tests were performed using *I* and *D* metrics on all three sets of climate layers. Results are given as the proportion of 26 triplets (sister pair plus confamilial outgroup) in which the sisters are more similar to each other than either is on average to the outgroup.

Layers	<i>I</i>			<i>D</i>		
	Bioclim	PCA	PSS	Bioclim	PCA	PSS
Proportion of comparisons showing niche conservatism	0.50	0.35	0.77	0.46	0.35	0.77
<i>t</i> -test significant	No	No	$P < 0.001$	No	No	$P < 0.001$
Binomial test significant	No	No	$P < 0.005$	No	No	$P < 0.005$

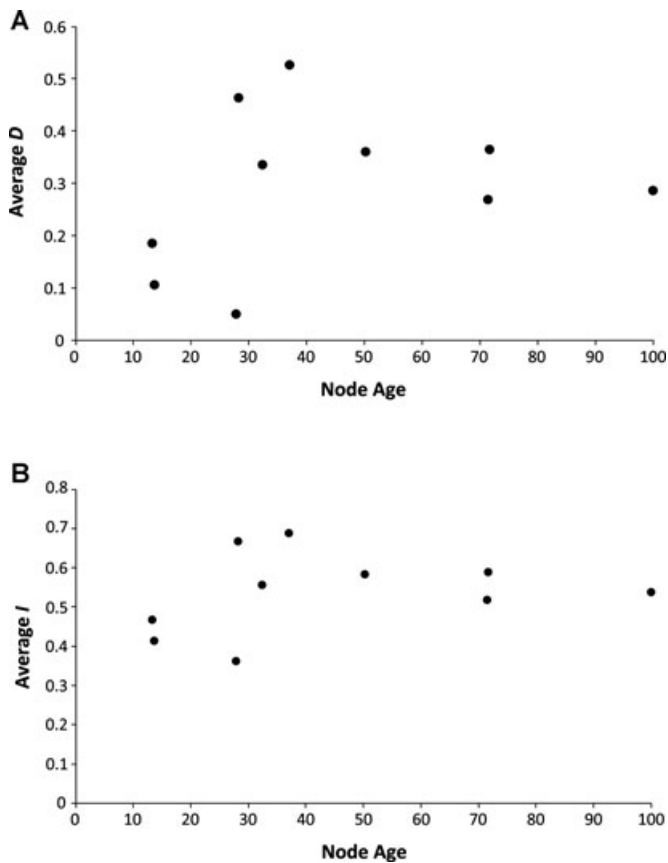


Figure 1. Average niche overlap as a function of node age using Fitzpatrick and Turelli's (2006) method for phylogenetically corrected average contrasts. Node ages are given as percentage of total tree length. Results are shown for Cuban Anolis, using both *D* (Panel A) and *I* (Panel B).

indicates a strong association between environmental niche similarity and the degree to which species range's overlap we expect local sympatry among species with similar environmental niches. To the degree that this is true, we expect a pattern of conserved environmental niche evolution will be associated with sympatry of related forms in local communities. The Mantel test, however, indicates the absence of a significant relationship between phylogenetic distance and environmental niche divergence ($P < 0.77$). This absence of evidence for strong environmental niche conservatism suggests that related forms are not particularly likely to share similar environmental niche characters, and unlikely to show a tendency to occur sympatrically as a result.

Discussion

IS ECOLOGICAL DIVERGENCE ASSOCIATED WITH SPECIATION?

Our reanalysis of Peterson et al.'s (1999) data supports the hypothesis that conflict about the prevalence of ecological divergence during speciation stems from testing dramatically different

null hypotheses. By conducting analyses similar to those in the original study, we support the hypothesis of niche similarity by finding that sister species tend to be more similar than expected based on comparison to random predictions or environmental similarities between the regions they inhabit (Table 2). Related tests asking whether the ENMs of sister species are more similar than pairs of confamilial taxa were only able to support this hypothesis when potential vegetation was included in the analysis. This result underscores the sensitivity of conclusions about niche evolution to choice of environmental variables, and may indicate that tolerance for different vegetation types is more phylogenetically constrained than tolerance for different climatic conditions. When an even more stringent niche-conservation criterion—that sister species' ENMs are more similar than either is to an outgroup's ENM—is tested, our data do not indicate significant environmental niche conservatism, even when potential vegetation is considered. These results suggest that ENMs from sister species are nearly always more similar to one another than random predictions, but may often be no more similar than ENMs sampled from more distantly related taxa.

Peterson et al.'s (1999) assertion that speciation across the Isthmus of Tehuantepec “takes place in geographic, not ecological, dimensions” is further challenged by the rejection of the hypothesis of niche equivalency. This hypothesis is rejected among sister species in nearly every case, indicating that significant environmental niche differentiation occurs in association with most speciation events (Table 1).

In addition to highlighting how methodological differences may explain discordant conclusions, our results strongly support the conventional wisdom that the niches of related species tend to be similar, but rarely identical. Indeed, both our more stringent comparisons of Mexican sister taxa to confamilial outgroups and our analysis of the Cuban anoles show that environmental niches can be quite labile. Thus, future comparative studies must treat niche conservation as a continuum, as done with analyses of phenotypic evolution (Felsenstein 1985) or range overlap (Barraclough and Vogler 2000; Fitzpatrick and Turelli 2006).

DOES NICHE CONSERVATION CONSTRAIN COMMUNITY COMPOSITION?

Our analysis of Cuban anoles shows that environmental niche overlap is closely tied to geographic overlap, but not to phylogenetic distances. Together, these results suggest a degree of niche conservation that is unlikely to phylogenetically bias composition of local communities (Fig. 1). If niche similarity were strongly associated with phylogenetic relatedness, we would have predicted that studies of community composition would recover significant phylogenetic clustering. As noted above, this conclusion may only hold with respect to the variables used in this particular analysis; it is entirely possible that different conclusions would be drawn if

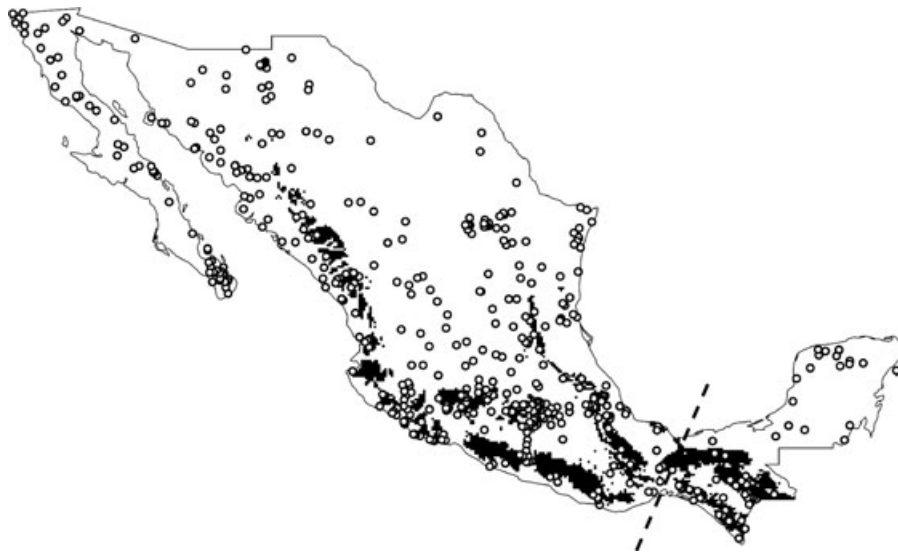


Figure 2. Predicting a broadly distributed species distribution using the niche model of a more geographically restricted species. Points represent actual occurrence of *Falco sparverius*, whereas the underlying landscape represents habitat suitability modeled using the distribution of *Falco tropicalis* (dark gray, predicted presence; white, predicted absence). The location of the Isthmus of Tehuantepec is represented by a dashed line. Despite correctly predicting less than 1/5 of the occurrence points for *sparverius*, this model shows a highly significant fit ($\chi^2 = 40.887$, $P < 0.0005$) due to the geographically restrictive niche model for *tropicalis*. This comparison is made over the whole of Mexico for the purposes of illustration only, and uses a threshold of 50% of the maximum GARP suitability score. The comparisons made in Table 2 are made only using the portion of Mexico on the opposite side of the Isthmus of Tehuantepec from the sister species to match the methodology used by Peterson et al. (1999) (A. T. Peterson, pers. comm.).

microhabitat or diet data were analyzed in the same fashion. It is important to note that this analysis has treated niche conservatism as analogous to phylogenetic signal (i.e., the tendency for related species to share similar traits). Several recent analyses have suggested that niche conservatism be defined as ecological similarity between species over and above the similarity that is expected under simple models of Brownian motion (Pearman et al. 2008; Losos, in press). From this perspective, the presence of phylogenetic signal is necessary, but not sufficient, for the diagnosis of phylogenetic niche conservatism. Although this may be a worthwhile distinction, it remains to be seen whether these patterns can be distinguished in practice.

ROBUSTNESS OF RESULTS TO ENM CONSTRUCTION METHODS

Both the ENM construction methods and the statistical tests used to evaluate their similarity can contribute error and bias results. There are many choices that must be made in constructing ENMs from occurrence data, each of which will introduce bias if they produce ENMs that are either overly permissive or overly restrictive (see Elith et al. 2006). Discussion and guidance are available for many of the issues that arise in ENM construction, including choice of algorithm (e.g., Elith et al. 2006; Guisan and Zimmerman 2000), choice of environmental variables (e.g., Parra et al. 2004; Peterson and Nakazawa 2008), spatial resolution (e.g., Engler et al. 2004; Guisan et al. 2007a,b), and sample size (e.g.,

Guisan et al. 2007b; Papes and Gaubert 2007). In addition to the concerns that accompany any attempts to model species' tolerances, comparative studies of ENMs must consider the effects these biases might have on measures of ecological similarity. Methods that routinely overpredict ranges, for example, will tend to detect more similarity between species' niches, perhaps leading to the spurious inference of niche conservatism or equivalency. An overly restrictive method, however, will tend to result in the opposite, potentially recovering niche differentiation where no actual ecological differences exist.

EVALUATION OF OUR METHODS RELATIVE TO THEIR PREDECESSORS

In addition to permitting quantitative assessment of niche similarity and statistical tests at both extremes of the niche conservatism continuum, we believe our methods are superior to the one other statistical test—Peterson et al.'s (1999) chi-square test—that has been proposed for testing one end of this continuum. The chi-square test of niche similarity is particularly problematic when comparisons are made between broadly and narrowly distributed forms. An example can be seen in Figure 2, which compares *F. sparverius* to *F. tropicalis*. In such cases, the null prediction of the fraction of occurrences that should be accurately predicted "at random" becomes small enough that even a fairly poor prediction of the sister species' distribution becomes statistically significant. As a result, the chi-square test infers niche conservatism in this

case in spite of the fact that the niche model for *tropicalis* omits more than 75% of the occurrence points for *sparverius*. In systems where a generalist species even partially overlaps the niche model of a specialist, this test will be very likely to detect a significant signal of niche conservatism even though this conclusion is difficult to reconcile with their ecology.

The chi-square and binomial null distributions employed by Peterson et al. (1999) and Martinez-Meyer and Peterson (2006) may also be problematic when sample sizes from the two species are unbalanced. Indeed, statistical significance depends only on having an adequate sample size for one of the species; in the most extreme case a significant signal of niche conservatism may be detected if the sample for one species includes only a single locality as long as the sample from the other species is adequate. An example can be seen in Table 1, in the comparison between *Chlorostilbon canivetii* and *C. salvini* using Maxent. Although only one occurrence record was available for *salvini*, this record may still be used to build an ENM. Due to the necessarily restricted environmental space spanned by a single occurrence, this model predicts a very small portion of the total geographic area being examined. For the reasons discussed above, even a very poor fit to the occurrence data for *canivetii* (in this case correctly predicting 2 out of 63 occurrences) is statistically significant against the null hypothesis of “no information.” This leads to a conclusion of niche conservatism from a single datapoint. Such artifacts do not arise with our methods. Resampling-based tests are considerably less likely to reject the null hypothesis when sample sizes are low, leading to more conservative statistical conclusions.

Conclusions

Comparative studies of environmental niche evolution have suffered from methodological shortcomings that have contributed to conflicting interpretations of data that may be more apparent than real. Focusing on whether ecological divergence occurs in association with speciation, we develop new methods for comparing ENMs. Specifically, we have proposed two metrics for ENM similarity, including a traditional ecological measure of niche overlap, and shown how they can be used for both hypothesis testing and comparative phylogenetic analyses. Our tests of four hypotheses concerning niche evolution demonstrate that sister species' niches are rarely identical, but they tend to be more similar than non-sisters from the same family, more similar than expected based on random predictions, and more similar than expected based on the broad environmental background against which they occur. In addition to explaining apparent conflict about the prevalence of ecological divergence during speciation, these results reinforce the conclusions of Wiens and Graham (2005) that binary conclusions regarding niche conservatism are dependent on the scale, both temporal and ecological, at which they are analyzed.

Our results suggest that inferences about niche similarity or niche equivalency alone should not be used to infer the relative contribution of geographic or ecological processes to speciation. This is not to say that environmental niche modeling and hypothesis testing cannot inform our understanding of speciation. Consider, for example, Kozak and Wiens' (2006) analysis of the role of niche conservatism in speciation of salamanders in eastern North America. They show that the environments in the disjunct species ranges are more similar to one another than to the environments from intervening areas in which both are absent. Based on their analyses of intervening absence localities, Kozak and Wiens (2006) argued that the inability of species to colonize these regions contributed to the divergence of allopatric populations.

Using Schoener's *D* opens the possibility of examining niche evolution at multiple levels by producing commensurate results from ENMs and more traditional ecological data such as micro-habitat and/or diet overlap, provided that ENMs can be produced so that suitability scores are proportional to potential species density. Although we focus on only two measures of niche overlap, there are many alternatives (e.g., Schoener 1968; Petraitis 1979; Slobodchikoff and Schulz 1980; Smith and Zaret 1982) and some may be better suited for particular studies. In addition, measures of niche breadth may also be applied to ENMs (Levins 1968; Colwell and Futuyma 1971; Petraitis 1979). These hold special appeal for those wanting to study niche evolution in a phylogenetic framework, as they are traits of single tip taxa and therefore can be analyzed by standard comparative methods (e.g., Felsenstein 1985; Blomberg et al. 2003) that do not apply to niche overlaps.

The methods presented here are intended to complement, rather than replace, other methods used to analyze patterns of ecological divergence. Methods from community phylogenetics (Webb et al. 2002; Johnson and Stinchcombe 2007; Silvertown et al. 2006) may be more appropriate for answering questions about specific ecological factors favoring co-occurrence of species, whereas our methods may be more useful for describing broad patterns of niche divergence across multiple environmental axes within a clade. Our hypothesis-testing methods do not require phylogenies.

Although our study is focused on niche differentiation over evolutionary time scales, our analyses may also be used to study niche divergence over ecological time scales. For instance, whether introduced species change their use of environmental space when compared to their native ranges is of great interest to ecologists (e.g., Broennimann et al. 2007), and our methods provide an alternative (or supplementary) statistical framework for testing hypotheses relating to this phenomenon. Our hypothesis-testing methods can also be adapted to examine the ecological significance of phylogeographic breaks, the effects of spatial error on measures of niche breadth and overlap, and the importance

of niche conservatism in maintaining allopatry (similar to the analysis performed by Kozak and Wiens (2006)).

The methods presented here open new avenues for studying niche evolution, but the fusion of ecological and evolutionary analyses is still in its infancy. Although the recent use of ENMs as characters in evolutionary studies has been illuminating and valuable, there is much to be gained in bringing the methodology in this field more into line with the methods developed over decades of both ecological and evolutionary comparative studies. With respect to phylogenetic studies of niche overlap or distance between species in environmental space, it remains a challenge to develop null models for pairwise comparisons that can play the role of the canonical Brownian motion model for divergence of tips (Felsenstein 1985). Future studies involving phylogenetic niche analyses might compare relative divergence in micro- versus macrohabitat over time, or combine niche metrics with phenotypic data to explore the relationship between ecological and morphological divergence, ecological divergence and reproductive isolation (as in Funk et al. 2006), or ecological divergence as a function of range overlap.

Although the meaning and significance of a measure of overlap between species may be informed by hypothesis testing, quantifying rates and patterns of phylogenetic divergence is likely to tell us more about niche evolution than testing hypotheses regarding niche conservatism. At some level, the existence of niche conservatism is almost trivially true (there is some degree of ecological similarity between closely related species due to shared ancestry); and at another level, it is almost trivially false (closely related species are unlikely to be ecologically equivalent). To paraphrase Harvey and Pagel's discussion of niche conservatism (1991, p. 38), birds do not become insects, but that does not mean that we cannot tell a cockatiel from a cormorant.

ACKNOWLEDGMENTS

We thank the following for their assistance and comments on earlier drafts: A. Townsend Peterson, J. Soberon, V. Sanchez-Cordero, S. Phillips, J. Losos, J. Wiens, J. Bastow, P. Wainwright, B. Shaffer, D. Schemske, T. Schoener, C. Moritz, R. Anderson, B. Monahan, A. Wright, R. Carlson, S. Veloz, S. Cameron, K. Moore, D. Grossenbacher, P. Ryan, S. Elmen-dorf, M. Wu, M. Brandley, T. Iglesias, and L. Moyle. We also thank the reviewers of this manuscript, in particular one anonymous reviewer whose extensive comments resulted in substantial improvements.

LITERATURE CITED

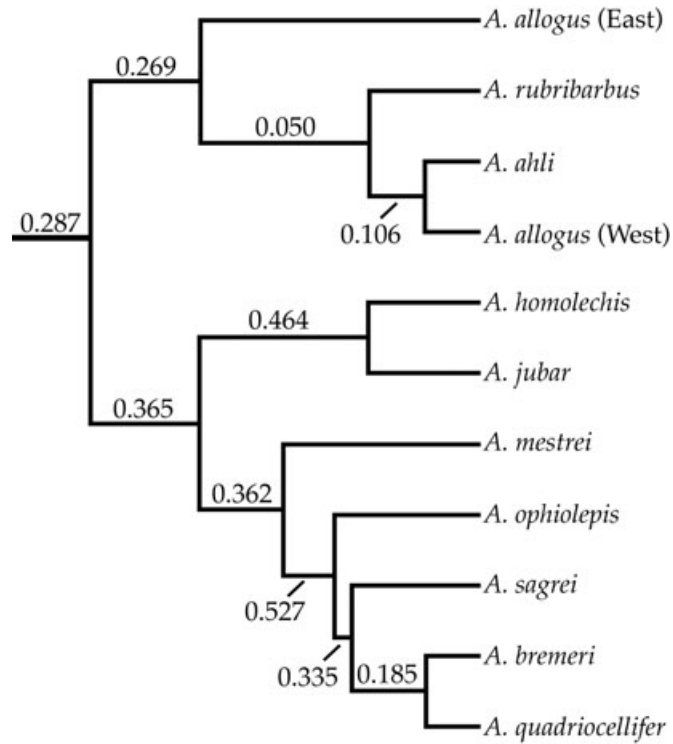
- Ackerly, D. D., and W. K. Cornwell. 2007. A trait-based approach to community assembly: partitioning of species trait values into within- and among-community components. *Ecol. Lett.* 10:135–145.
- Barracough, T. G., and A. P. Vogler. 2000. Detecting the geographical pattern of speciation from species-level phylogenies. *Am. Nat.* 155:419–434.
- Blomberg, S. P., T. Garland, Jr., and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717–745.
- Bolnick, D. I., and T. J. Near. 2005. Tempo of hybrid inviability in centrarchid fishes (Teleostei: Centrarchidae). *Evolution* 59:1754–1767.
- Broennimann, G., U. A. Treiner, H. Muller-Scharer, W. Thuiller, A. T. Peterson and A. Guisan. 2007. Evidence of climatic niche shift during biological invasion. *Ecol. Lett.* 10:701–709.
- Colwell, R. K., and D. Futuyma. 1971. On the measurement of niche breadth and overlap. *Ecology* 52:567–576.
- Coyne, J. A., and H. A. Orr. 1989. Patterns of speciation in *Drosophila*. *Evolution* 43:362–381.
- Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151.
- Engler, R., A. Guisan, and L. Rechsteiner. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *J. Appl. Ecol.* 41:263–274.
- ESRI. 2006. ArcGIS 9.2. Environmental Systems Research Institute, Redlands, CA.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15.
- Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24:38–49.
- Fitzpatrick, B. M., and M. Turelli. 2006. The geography of mammalian speciation: mixed signals from phylogenies and range maps. *Evolution* 60:601–615.
- Funk, D. J., P. Nosil, and W. Etges. 2006. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proc. Natl. Acad. Sci. USA* 103:3209–3213.
- Graham, C. H., S. R. Ron, J. C. Santos, C. J. Schneider, and C. Moritz. 2004. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution* 58:1781–1793.
- Guisan, A., and N. E. Zimmerman. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147–186.
- Guisan, A., C. H. Graham, J. Elith, and F. Huettmann. 2007a. Sensitivity of predictive species distribution models to change in grain size. *Divers. Distrib.* 3:332–340.
- Guisan, A., N. E. Zimmermann, J. Elith, C. H. Graham, S. Phillips, and A. T. Peterson. 2007b. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecol. Monogr.* 77:615–630.
- Hansen, M., R. DeFries, J. R. G. Townshend, and R. Sohlberg. 1998. UMD Global Land Cover Classification, 1 kilometer, version 1.0. Department of Geography, Univ. of Maryland, College Park, MD.
- Harvey, P. H., and M. D. Pagel. 1991. *The comparative method in evolutionary biology*. Oxford Univ. Press, Oxford, U.K.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25:1965–1978.
- Jiménez-Valverde, A., and J. M. Lobo. 2007. Threshold criteria for conversion of probability of species presence to either—or presence-absence. *Acta Oecol.* 31:361–369.
- Johnson, M. T. J., and J. R. Stinchcombe. 2007. An emerging synthesis between community ecology and evolutionary biology. *Trends Ecol. Evol.* 22:250–257.
- Kambhampati, S., and A. T. Peterson. 2007. Ecological niche conservation and differentiation in the wood-feeding cockroaches, *Cryptocercus*, in the United States. *Biol. J. Linn. Soc.* 90:457–466.
- Knouft, J. H., J. B. Losos, R. E. Glor, and J. J. Kolbe. 2006. Phylogenetic analysis of the evolution of the niche in lizards of the *Anolis sagrei* group. *Ecology* 87:S29–S38.

- Kozak, K. H., and J. J. Wiens. 2006. Does niche conservatism promote speciation? A case study in North American salamanders. *Evolution* 60:2604–2621.
- Kozak, K. H., C. H. Graham, and J. J. Wiens. 2008. Integrating GIS-based environmental data into evolutionary biology. *Trends Ecol. Evol.* 23:141–148.
- Legendre, P., and E. D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129:271–280.
- Levins, R. 1968. *Evolution in changing environments*. Monographs in population biology volume 2. Princeton Univ. Press, Princeton, NJ.
- Liu, C., P. M. Berry, T. P. Dawson, and R. G. Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28:385–393.
- Lobo, J. M., A. Jiménez-Valverde, and R. Real. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecol. Biogeogr.* 17:145–151.
- Loiselle, B. A., C. A. Howell, C. H. Graham, J. M. Goerck, T. Brooks, K. G. Smith, and P. H. Williams. 2003. Avoiding pitfalls of using species distributions models in conservation planning. *Conserv. Biol.* 17:1592–1600.
- Losos, J. B. In Press. Phylogenetic niche conservatism, phylogenetic signal, and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol. Lett.*
- Maddison, W. P. 2000. Testing character correlation using pairwise comparisons on a phylogeny. *J. Theor. Biol.* 202:195–204.
- Martinez-Meyer, E., and A. T. Peterson. 2006. Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition. *J. Biogeogr.* 33:1779–1789.
- Papes, M., and P. Gaubert. 2007. Modeling ecological niches from low numbers of occurrences: assessment of the conservation status of poorly known viverrids (Mammalia, Carnivora) across two continents. *Divers. Distrib.* 13:890–902.
- Parra, J. L., C. C. Graham, and J. F. Freile. 2004. Evaluating alternative data sets for ecological niche models of birds in the Andes. *Ecography* 27:350–360.
- Pearman, P. B., A. Guisan, O. Broennimann, and C. F. Randin. 2008. Niche dynamics in space and time. *Trends Ecol. Evol.* 23:149–158.
- Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. T. Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *J. Biogeogr.* 34:102–117.
- Peterson, A. T., and R. D. Holt. 2003. Niche differentiation in Mexican birds: using point occurrences to detect ecological innovation. *Ecol. Lett.* 6:774–782.
- Peterson, A. T., and Y. Nakazawa. 2008. Environmental data sets matter in ecological niche modeling: an example with *Solenopsis invicta* and *Solenopsis richteri*. *Global Ecol. Biogeogr.* 17:135–144.
- Peterson, A. T. and A. S. Nyári. 2007. Ecological niche conservatism and Pleistocene refugia in the thrush-like mourner, *Schiffornis* sp., in the neotropics. *Evolution* 62:173–183.
- Peterson, A. T., J. Soberon, and V. Sanchez-Cordero. 1999. Conservatism of ecological niches in evolutionary time. *Science* 285:1265–1267.
- Peterson, A. T., M. Papes, and M. Eaton. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* 30:550–560.
- Petraitis, P. S. 1979. Likelihood measures of niche breadth and overlap. *Ecology* 60:703–710.
- Pfenninger, M., C. Nowak, and F. Magnin. 2007. Intraspecific range dynamics and niche evolution in *Candidula* land snail species. *Biol. J. Linn. Soc.* 90:303–317.
- 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., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190:231–259.
- Rzedowski, J. 1990. *Vegetación Potencial*. Atlas Nacional de México. Instituto de Geografía, UNAM, Mexico.
- Schoener, T. W. 1968. *Anolis* lizards of Bimini: resource partitioning in a complex fauna. *Ecology* 49:704–726.
- Silvertown, J., K. McConway, D. Gowing, M. Dodd, M. F. Fay, J. A. Joseph, and K. Dolphin. 2006. Absence of phylogenetic signal in the niche structure of meadow plant communities. *Proc. R. Soc. Lond. B* 273:39–44.
- Slobodchikoff, C. N., and W. C. Shulz. 1980. Measures of niche overlap. *Ecology* 61:1051–1055.
- Smith, E. P., and T. M. Zaret. 1982. Bias in estimating niche overlap. *Ecology* 63:1248–1253.
- Stockman, A. K., R. M. Danell, and J. E. Bond. 2008. D-NOVL: a program to simulate overlap between two niche-based distribution models. *Mol. Ecol. Resources* 8:290–294.
- Stockwell, D. B. 2006. Improving ecological niche models by data mining large environmental datasets for surrogate models. *Ecol. Model.* 192:188–196.
- Stockwell, D. B., and D. Peters. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *Int. J. Geogr. Inform. Syst.* 13:143–158.
- Thioulouse, J., D. Chessel, S. Dolédec, J.-M. Olivier. 1997. ADE-4: A multivariate analysis and graphical display software. *Stat. Comput.* 7:75–83.
- Van der Vaart, A. W. 1998. *Asymptotic statistics*. Cambridge Univ. Press, Cambridge, U.K.
- Webb, C. O., D. D. Ackerly, M. A. McPeck, and M. J. Donoghue. 2002. Phylogenies and community ecology. *Annu. Rev. Ecol. Systemat.* 33:475–505.
- Wiens, J. J., and C. H. Graham. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. *Annu. Rev. Ecol. Evol. Systemat.* 36:519–539.

Associate Editor: D. Funk

Appendix 1. Species used in comparisons across the Isthmus of Tehuantepec. The occurrence data were obtained from the authors of Peterson et al. (1999). Some species from that study were omitted due to lack of data. Sample sizes are given in tables in the main body of this study.

Northern species	Southern species	Outgroup
Butterflies		
<i>Pyrrhosticta abderus abderus</i>	<i>Pyrrhosticta abderus electryon</i>	<i>Battus eracon</i>
<i>Pereute charops charops</i>	<i>Pereute charops nigricans</i>	<i>Pyrisitia proterpia</i>
<i>Dismorphia eunoe eunoe</i>	<i>Dismorphia eunoe chamula</i>	<i>Dismorphia amphione</i>
<i>Pereute charops leonilae</i>	<i>Pereute charops nigricans</i>	<i>Ascia monuste</i>
<i>Catasticta nimbice nimbice</i>	<i>Catasticta nimbice ochracea</i>	<i>Leptophobia aripa</i>
Birds		
<i>Amazilia beryllina</i>	<i>Amazilia devillei</i>	<i>Helimaster longirostris</i>
<i>Amazilia heloisa</i>	<i>Amazilia ellioti</i>	<i>Hylocharis eliciae</i>
<i>Chlorostilbon canivetii</i>	<i>Chlorostilbon salvini</i>	<i>Tilmatura dupontii</i>
<i>Colaptes cafer</i>	<i>Colaptes mexicanoides</i>	<i>Melanerpes uropygialis</i>
<i>Cyanocitta coronata</i>	<i>Cyanocitta ridgwayi</i>	<i>Cyanocorax beecheii</i>
<i>Empidonax occidentalis</i>	<i>Empidonax flavescens</i>	<i>Camptostoma imberbe</i>
<i>Ergaticus ruber</i>	<i>Ergaticus versicolor</i>	<i>Pipilo chlorurus</i>
<i>Eugenes fulgens</i>	<i>Eugenes viridiceps</i>	<i>Campylopterus hemileucurus</i>
<i>Falco sparverius</i>	<i>Falco tropicalis</i>	<i>Falco mexicanus</i>
<i>Icterus pustulatus</i>	<i>Icterus sclateri</i>	<i>Pheucticus melanocephalus</i>
<i>Mimus polyglottos</i>	<i>Mimus gilvus</i>	<i>Toxostoma bendirei</i>
<i>Pheucticus chrysopleus</i>	<i>Pheucticus aurantiacus</i>	<i>Lanio aurantius</i>
<i>Picooides jardiinii</i>	<i>Picooides sanctorum</i>	<i>Picooides scalaris</i>
<i>Salpinctes obsoletus</i>	<i>Salpinctes neglectus</i>	<i>Thryothorus maculipectus</i>
<i>Strix varia</i>	<i>Strix fulvescens</i>	<i>Glaucidium gnoma</i>
<i>Turdus assimilis</i>	<i>Turdus leucachen</i>	<i>Turdus infuscatus</i>
Mammals		
<i>Habromys lepturus</i>	<i>Habromys lophurus</i>	<i>Neotoma goldmani</i>
<i>Peromyscus melanocarpus</i>	<i>Peromyscus zarhynchus</i>	<i>Tylomys nudicaudus</i>
<i>Sciurus coliaei</i>	<i>Sciurus yucatanensis</i>	<i>Ammospermophilus leucurus</i>
<i>Artibeus aztecas aztecas</i>	<i>Artibeus aztecas minor</i>	<i>Chrotopterus auritus</i>
<i>Peromyscus megalops</i>	<i>Peromyscus zarhynchus</i>	<i>Peromyscus merriami</i>



Appendix 2. Phylogeny for Cuban *Anolis* species. Phylogeny for Cuban *Anolis* species from Knouft et al. (2006) estimated from a Bayesian analysis of ~1500 bp of mitochondrial DNA. Branch lengths are proportional to time. Numbers at nodes indicate average niche overlap across that node using phylogenetically corrected averages of Schoener's (1968) *D* (see text for details).

Supporting Information

The following supporting information is available for this article:

Appendix S1. Potential problems with applying a threshold to continuous data when making interspecific comparisons.

Appendix S2. Potential problems with measuring overlap between models in environmental space.

Appendix S3. Fit of Maxent ENMS to data.

This material is available as part of the online article from:

<http://www.blackwell-synergy.com/doi/abs/10.1111/j.1558-5646.2008.00482.x>

(This link will take you to the article abstract).

Please note: Wiley Blackwell Publishing is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.