Traditionally, the goal of systematics has been to produce well-supported and biologically meaningful classifications (Mayr 1968). Many authors have therefore advocated the use of multiple lines of evidence such as information on organisms’ environmental requirements when delimiting species (Rader et al. 2006; De Queiroz 2007; Stockman and Bond 2007). One Major form of ecological divergence—changes in environmental requirements between taxa—can cause, or contribute to reproductive isolation (Nosil et al. 2003; Coyne and Orr 2004). As such, evidence of different environmental requirements can provide substantial support for delimiting species. Indeed, such data can have even broader uses in phylogenetics because many evolutionary hypotheses make predictions about how ecological requirements change over time (Ehrlich and Raven 1964; Schluter 2000; Graham et al. 2004; Losos 2008).

Accordingly, there is growing interest in using geographic distribution data (i.e., the locations where a species is present and/or absent) to make inferences about changes in environmental requirements between taxa (Peterson et al. 1999; Raxworthy et al. 2007; Rissler and Apodaca 2007; Kozak et al. 2008; Warren et al. 2008). Many such studies make inferences about ecological similarity between species by first estimating the probability that individual species will be present at locations across a study region (estimates of species distributions) and then comparing estimates between species. Other approaches are possible, notably multivariate analyses, which determine whether each species is present in a different set of environments. Comparisons of niche models offer a valuable complement to phylogenetic studies (Graham et al. 2004; Warren et al. 2008) and have several advantages over experimental measures of environmental requirements. First, ethical or practical considerations may make experimental data much more difficult to obtain than distribution records. Second, distribution data make it possible to sample from many more environments than we could hope to manipulate experimentally (Godsoe 2010). Third, distribution data reflect the consequences of many ecological interactions that may be missed in small-scale experimental studies (Soberon and Peterson 2005). Given the promise of these methods, it is essential to determine when they will correctly indicate that taxa are ecologically distinct.

In spite of this promise, there are considerable gaps in our knowledge of the relationship between environmental requirements and species distributions. Even taxa with identical requirements may disperse to different locations and so have different distributions. Several authors have hypothesized that this fact may lead existing methods to incorrectly infer changes in environmental requirements (Bond and Stockman 2008; Warren et al. 2008). However, we rarely know the actual level of ecological divergence in nature, so it has heretofore been impossible to determine if existing methods produce biased inferences.

To address this problem, I simulated the distributions of 2 species (hereafter Species 0 and Species 1) with niches that are completely exchangeable (i.e., an environment suitable to one species is just as suitable to the other; Templeton 1989), but which have allopatric distributions across a complex landscape. For the sake of simplicity, these taxa are referred to as species, though they could in fact represent other categories such as subspecies, or populations. The similarity of the environments to which each species can disperse was manipulated by simulating an environmental gradient across the landscape. I then tested whether a variety of methods would correctly infer that the environmental requirements of the 2 species were either identical or more similar than expected by chance.
THE MODEL

Simulated Distribution

There is debate on how exactly environmental variables affect the suitability of an environment for a focal species (Araujo and Guisan 2006). For this reason, I elected to model the probability that an environment was a part of the niche with a relatively simple logistic function:

\[
P(\text{environment} \in \text{niche}) = \frac{1}{1 + e^{0.1E_1^2 + 100(0.5 - E_2)^2}}. (1)
\]

This equation states that organisms are most likely to be present in environments with intermediate values for 2 environmental variables \((E_1, E_2)\). I assumed that both species had a niche defined by the same equation and hence were ecologically (demographically) exchangeable (Templeton 1989; Rader et al. 2006; Bond and Stockman 2008). Defining the probability that an environment is a part of the niche in this way is particularly useful because niche-modeling algorithms such as generalized linear models (GLMs) and boosted regression trees (BRTs) fit models of this form (McCullagh and Nelder 1989; Friedman et al. 2000). I selected parameter values for \(E_1\) and \(E_2\) that ensured the organisms were reasonably common (organisms were present in an average of 21% of locations) using trial and error. This was done to ensure that a large sample of presences were available for subsequent statistical analyses.

Species distributions were simulated on a landscape of 10,000 locations in a 100 \(\times\) 100 cell grid. Each location had its own spatial \((X, Y)\) coordinates. In turn, these coordinates were used to calculate the environmental attributes \((E_1\) and \(E_2))\) that determine the probability that an environment was suitable in Equation 1. The first environmental variable \((E_1)\) contained fine-scale environmental variation. The value of this variable was governed by normal distribution, with an independent value at each location, a mean of 0 and a variance of 1. The second environmental variable \((E_2)\) had a more complex spatial structure, with small-scale spatial autocorrelation and large-scale environmental gradients (Fig. 1a). I generated the small-scale autocorrelation by randomly adding 15 “peaks” and “valleys” to the landscape. Each peak had a bivariate normal distribution centered at a randomly selected \((X, Y)\) coordinate on the landscape, a standard deviation of 10% of the length of the landscape, and a correlation between the variables of 0. Valleys were created by simulating a similar normal distribution and multiplying the result by \(-1\). The value of environmental variable 2 at any location was calculated as the sum of the effects of all peaks and valleys (Fig. 1a). I then simulated a gradient by adding a value proportional to the \(X + Y\) coordinate to each point on the map, rescaling the height of the peaks such that the maximum value of this variable was 1 and the minimum value was 0 on each landscape. As the strength of the gradient increased, the amount of regional variation increased and the level of local variation decreased (Fig. 1b–d). One hundred replicated environments were simulated for each of 5 different levels of the environmental gradient (0, 0.5, 1, 4, and 16).

The landscape was divided into 2 regions along a diagonal line from the upper left corner to the lower right corner (Fig. 1a). Locations above this line are labeled Region 0 and locations below are labeled Region 1. A separate species occurred in each region (hereafter Species 0 and Species 1). Each species could disperse to any environment within its region and so would be present in any location within the region with a probability determined by Equation (1). The division between the 2 regions represents a barrier to dispersal strong enough to keep each species in its own region and so make it impossible for the species to encounter one another. The assumption of a strong boundary between regions is a valuable simplification because it circumvents the need to model interactions between species (Chase and Leibold 2003; Case et al. 2005). It also represents a proxy of classic biogeographic barriers such as Wallace’s line (Wallace 1860; Mayr 1944), the Isthmus
of Panama, or the Isthmus of Tehuantepec. Such barriers can prevent many taxa, even vagile species such as birds or marine organisms with planktonic larvae, from dispersing from one region to another (Knowlton et al. 1993; Peterson et al. 1999), but still allow organisms access to many environments within a region. A major consequence of dividing the organisms into 2 species with allopatric ranges in this way is that it ensures that the 2 species have identical fundamental niches (environmental requirements) but different realized niches (geographic distributions).

By altering the strength of the regional gradient, it is possible to alter the range environments to which each species can disperse. If there is no regional gradient, then approximately the same environments are present in each region (Fig. 1b). If there is a moderate gradient, then Species 0 can disperse to environments where \( E_2 \) is higher on average, whereas Species 1 disperses to environments with slightly lower \( E_2 \) values (Fig. 1c). If there is a strong environmental gradient, then Species 0 has access to environments with scores on \( E_2 \) that are larger than any value of \( E_2 \) occupied by Species 1 (Fig. 1d).

**Multivariate Analyses**

I tested whether both species lived in identical environments by randomly sampling 400 locations on the simulated landscape and comparing the environmental variables at locations where Species 1 was present with the scores of locations where Species 2 was present using a multivariate analysis of variance (MANOVA). This method is frequently reported in the literature and compares estimates of the mean for each environmental variable in each region, whereas other methods test the similarity of ecological niche models.

**Niche Modeling (Estimated Distribution)**

To create niche models, I sampled presence or absence points at random from each of the 2 regions in my landscape. For the sake of clarity, these models will be referred to as an estimate of a species' distribution, as opposed to the simulated distribution described above. Because there is no clear consensus on which niche-modeling algorithm to use to compare distributions (Peterson et al. 2007; Phillips 2008), 3 algorithms were implemented: GLMs, BRT, and maximum entropy (maxent). All 3 methods have strong theoretical and empirical support (Myers 1990; Friedman et al. 2000; Elith et al. 2006). GLM is a parametric method that requires the investigator to specify the relationship between the dependent and the independent variables. To accomplish this, I fit a model with linear and squared terms for the 2 environmental variables. BRT is a semiparametric method that determines the relationship between dependent and independent variables by combining inferences from a large number of decision trees. This method was implemented in R (R Development Core Team 2006; Ridgeway 2006; Elith et al. 2008) with a tree complexity of 2 (i.e., the model fit interactions between 2 variables), a relatively slow learning rate of 0.001, and a bagging fraction of 0.7. Maxent estimates the probability that a species will be present by constraining its predictions to resemble the empirical data and by minimizing the information contained in the residuals. This method was implemented in maxent version 3.2.19 (Phillips et al. 2006; Phillips and Dudík 2008) by modeling presences with default settings (regularization multiplier = 1, maximum iterations = 500, convergence threshold = 0.00001, maximum background points = 10,000, and output format = logistic). GLM and BRT niche models used 100 presences and absences from the appropriate region (Region 0 for Species 0 and Region 1 for Species 1). Maxent used 200 presences. This method does not require absences but characterizes the available environments by sampling background points. As with absences, these background points were only sampled from the appropriate region.

I tested the accuracy of each niche model using the receiver operating curve (AUC) statistic (Freeman 2007). This is a nonparametric estimate of a model's ability to distinguish between presence and absence points. It ranges from 0 to 1, with a score of 1 representing perfect discrimination and a score 0.5 representing a model that performs no better than random chance.

**Comparisons of Estimated Distributions**

To determine whether niche models were equally good at predicting the distribution of one species, 400 locations were sampled at random from one region. This sample was used to measure the ability of the niche model derived from the species to predict its own distribution. The same set of points was used to measure the ability of a niche model for the other species to predict the distribution of the first species. I tested whether the models were identical by determining whether the accuracy of a model from the correct species (say Species 0) was markedly better than the accuracy of a model from the incorrect species (say Species 1). An equivalent test was then applied to the other region to determine if niche models were equally good at predicting the distribution of the second species.

The \( \chi^2 \) test proposed by Peterson et al. (1999) was used to determine if the niche model of one species predicts the distribution of the other species more accurately than one would expect by chance. To accomplish this, 400 points were sampled from the range of each species. These points were used to create a niche model for the first species that was in turn used to extrapolate presences in the range of the second. By determining the proportion of samples from the second region in which organisms were present, and the proportion of samples that were classified correctly by a model of the first species, it is possible to determine if more presences were predicted correctly than would be expected by chance (Peterson et al. 1999; Warren et al. 2008). This test has been recently criticized for cases where one species is less common than another (Warren et al. 2008). This problem should have little effect on the
simulations presented here because both species were reasonably abundant across all replicates.

To evaluate the robustness of my conclusions, I performed additional sensitivity analyses for a subset of the statistical tests described above (MANOVAs and model comparisons with GLM and BRT). In these analyses, the extent of the region sampled to estimate a species’ distribution, 2 BRT parameters (bagging fraction and learning rate), the scale of local autocorrelation in $E_2$ and the coefficients associated with the 2 environmental variables in Equation (1) were manipulated (see online Appendix 1, available from http://www.sysbio.oxfordjournals.org/, for more details).

Permutation Tests of Niche Models

I implemented the permutation tests proposed by Warren et al. (2008) in R. To replicate the random overlap test of niche identity of Warren et al. (2008), I created an individual niche model for each species in maxent as described above, but with background points sampled from both Region 0 and Region 1. The distance between the predictions of these niche models was calculated using the $I$ statistic of Warren et al. (2008), a metric based on Hellinger distance that varies between 0 for nonoverlapping model predictions and 1 for identical model predictions. A null distribution of $I$ distances was calculated by creating 100 permutations of the original data set. In each permutation, 200 of the available presences were randomly assigned to Species 0 and 200 were randomly assigned to Species 1. I then calculated the $I$ statistic of Warren et al. for the estimated distribution of each simulated pair of species.

To implement the random background test of ecological similarity, maxent estimates of species distributions were created for each species by sampling 200 presences and background points from across the entire study area (Region 0 + Region 1). The predictions of the estimated distribution of Species 0 were then compared with the predictions of the model for Species 1. The null distribution consisted of 100 permutations of the original data. In each permutation, a niche model was created from a sample of 200 presences from Region 1 and background points from Region 1 and Region 1. I then calculated the $I$ statistic for the distance between predictions of these simulated species. Following Warren et al. (2008), these data were tested against the 2-sided alternative hypotheses that the distributions are either more or less similar than expected by chance.

When comparing the results of each of the tests presented here, it is useful to consider a distinction highlighted by Warren et al. (2008): rejection of the null hypothesis may have opposite meanings for different tests. In tests of ecological identity, such as MANOVAs, comparisons of AUC scores, and the random overlap tests, a rejection of the null indicates that the taxa are significantly different. In tests of ecological similarity, such as the $\chi^2$ test and the random background test, rejection of the null indicates that the taxa are more similar than expected by chance.

**RESULTS**

Most of the methods tested accurately inferred that the simulated species were identical in the presence of a modest environmental gradient. All the methods may erroneously infer that the species had different environmental requirements when each dispersed to a different set of environments. The severity of this problem varied with the algorithm used to estimate the distribution of a species and with the statistical test employed.

In the absence of regional gradients, MANOVA would determine that species had identical requirements, but even the smallest environmental gradient resulted in the test inferring that species were significantly different (Fig. 2a). The accuracy of extrapolations based on niche models depended on both the strength of regional variation and the niche-modeling algorithm used. In the absence of strong environmental gradients, there was little difference between a model from the correct species and extrapolations based on the other species (Fig. 2b–d). Extrapolations based on GLM were the most reliable, but this method occasionally produced inaccurate models in the presence of a strong environmental gradient (Fig. 2b). In the presence of strong regional gradients, extrapolations based on BRT and maxent were far less accurate than the true model (Fig. 2c,d). Likewise, maxent- and BRT-based niche models produced predictions no better than chance when there was an environmental gradient (see online Appendix 2). In spite of this poor behavior, each algorithm produced acceptable models in their home range (mean ± standard deviation for AUC scores; GLM 0.799 ± 0.064; BRT 0.796 ± 0.067; and maxent 0.755 ± 0.096).

In the presence of an environmental gradient, the random overlap test invariably rejected the hypothesis that the species were identical. Even in the absence of a gradient, this test frequently inferred that the species were different (Table 1). The random background test typically inferred that species were more similar than expected by chance, even in the presence of a strong environmental gradient (Table 2). However, in both of these tests, the distance between estimated distributions of the original species increased markedly in the presence of a strong environmental gradient.

The general conclusions of these analyses were consistent across most of the parameter combinations evaluated in the sensitivity analyses. Specifically, environmental gradients exaggerate ecological differences and that this problem was more likely to affect MANOVAs- and BRT-based analyses.

**DISCUSSION**

When using distribution data to make inferences about evolutionary change, it is important to recognize that the ecological niche is only one of the forces shaping distributions. Systematists should be particularly concerned about the role of dispersal limitation, as this agent can shape the process of speciation as much as changes in ecological requirements. Two strongly
supported ideas in evolution are that speciation is more likely when dispersal limitation reduces gene flow between populations so that in turn sister species frequently have allopatric or parapatric distributions (Jordan 1905; Mayr 1942; Endler 1977; Coyne and Orr 2004; Gavrilets 2004). Evolutionary theory and extensive observations thus predict that even sister species with identical environmental requirements will often occur in different regions. One of the best documented facts in ecology is that different regions contain different environments (Darwin 1859; MacArthur 1972; Udvardy 1975; Bailey 1995), and so we should expect many sister taxa to occur in different environments, even in the absence of changes in environmental requirements.

A useful way to think about the relationship between the similarity of the environmental requirements of 2 species and the similarity of their geographic distributions is to consider the relationship between interfertility and the rate of hybridization. The fact that organisms of 2 populations can interbreed does not guarantee that they will do so under natural conditions (Coyne and Orr 2004). It only means that when they encounter each other, there is a nonzero probability of hybridization (Lepais et al. 2009). Likewise, the fact that 2 populations are ecologically exchangeable does not imply that they will occur in identical environments. It simply means that they will have similar performance when they encounter similar environments. Just as the current rate of hybridization is only a useful tool for taxonomy if taxa encounter each other, the similarity of distributions may only be informative when both taxa encounter similar environments.

My simulations indicate that the methods tested confound changes in environmental requirements with changes in the environments available to each taxa but that these methods vary in their susceptibility to this problem. MANOVA was particularly sensitive to this problem. This observation bears further attention because many previous studies have used this or similar methods (Graham et al. 2004; Kozak et al. 2006;
Recent studies have criticized the use of MANOVA in this context (Stockman and Bond 2007; Warren et al. 2008). The results presented here allow for an elaboration on these critiques. Bond and Stockman (2007) hypothesize that MANOVA may exaggerate the evidence for divergence in environmental requirements because it detects differences that are biologically insignificant or because the assumptions underlying MANOVA are not typically met by real data. In the simulations presented here, there was no difference in ecological requirements (biologically significant or otherwise), and so there are clearly other problems with applying this method to determine if taxa have identical environmental requirements. It is possible that a violation of the assumptions underlying MANOVA is partially responsible for the high rate of false positives observed in this study, but the problem runs deeper. A MANOVA uses a sample of observations to test the hypothesis that the mean of all environmental variables is identical for presence points in both species. In the simulations presented here, it is possible to determine how much the means of these variables actually differ between environments occupied by each species. In the presence of a steep environmental gradient, the mean of $E_2$ was approximately 0.1 units (or 10% of the range of $E_2$) higher for Species 0 than for Species 1. The MANOVAs presented here correctly infer that the mean of some environmental variables are different, and it is just that this difference does not bear on the question of ecological exchangeability.

The random overlap test was also prone to inferring that the taxa were significantly different. This may be because the randomization procedure used in this test is akin to assuming that one species can access all the environments to which the other species can disperse. When 2 taxa have allopatric or even parapatric distributions, this assumption is unlikely to hold. MANOVA and the random overlap test may represent a sensible method to test if species have identical distributions (realized niches) but seem to be poor choices for testing if they have identical environmental requirements.

In contrast, many of the methods based on niche models performed reasonably well, though there was substantial variation between each approach. BRT and maxent produced reasonable inferences in the presence of moderate environmental gradients, whereas GLM frequently inferred that species were identical, even with a strong environmental gradient. These results were consistent for comparisons of AUC scores and $\chi^2$ tests. The background similarity test correctly inferred that species were similar in most conditions. Maxent and BRT performed much better than GLM in a series of comparisons of available methods (Elith et al. 2006), and so it is somewhat puzzling that in this case, BRT and maxent produced inferior predictions. A plausible explanation for this pattern is that GLM incorporates more information on the relationship between the environmental variables and the probability that a species is present. GLM requires the user to specify how individual variables shape the probability that a species will be present (specifying, e.g., that there are linear and squared terms). Neither BRT nor Maxent requires this much detail and so have less information to extrapolate to new environments. The pattern of errors is consistent with this hypothesis. In the absence of regional gradients, all methods accurately reproduced the hump-shaped probability distribution. In the face of environmental variation, BRT and maxent extrapolated presences for nearly all observations in the environments that were unavailable (Fig. 3). Elith and Graham (2009) report similar problems with maxent and BRT. These results suggest that carefully constructed niche models can provide reasonable metrics for comparing environmental requirements, though a strong environmental gradient may lead to erroneous inferences.

However, these conclusions on model performance must be interpreted with caution because the simulated environments were much simpler than those found in the real world. With so few variables, it is difficult to produce a model that is not close to the truth, whereas in real life, it can be exceedingly difficult to select a model that truly reflects the underlying biology of a system (Burnham and Anderson 2002). Thus, the simulations ignore a major source of error, and it is possible that GLM’s strong performance may not represent strong performance in nature. We need further tests of this observation based on real data (Peterson et al. 2007; Phillips 2008). Nevertheless, it is illuminating to realize
that even in relatively simple environments, dispersal limitation can confound methods for comparing the niches of organisms.

Arguably, the best way to mitigate the problems described here is to carefully consider the environments to which each species can disperse. In particular, we need to identify environments that occur in the range of one species, and not the other, as these are a potential source of errors. A simple approach to identify such regions is to use multivariate plots to visualize the background environment surrounding the distribution of each species (Broennimann et al. 2007; Godsoe et al. 2009). If one species can disperse to environments unavailable to the other, the simplest solution may be to exclude this region from subsequent analyses. This approach might make it impossible to detect divergent environmental requirements in some regions and so limit the scope of our inferences, but this seems preferable to relying on metrics that make erroneous inferences in such regions.

The application of niche modeling to study the evolution of environmental requirements is relatively recent. Though a number of papers provide thoughtful discussions (Rissler and Apodaca 2007; Stockman and Bond 2007; Warren et al. 2008; Bond and Stockman 2008), the best way to incorporate this information into systematics is at present unclear. A sensible starting point for this problem may be to consider a 2-part question: First, how could reliable information about an organism’s environmental requirements address a specific research question? Second, what biases might arise from the data available?

Consider, again, the question of whether 2 taxa represent distinct species. It has long been recognized that ecological differences can provide important evidence for species delimitation (Van Valen 1976; Andersson 1990), though there is ongoing work on how to incorporate insights from ecology from the many other sources of data available. Following the proposal of De Queiroz (2007) that distinct species be defined as separately evolving metapopulation lineages, evidence of different ecological requirements only represents one possible criterion for recognizing species. Evidence of strong ecological differentiation may be sufficient to delimit separate species, but in its absence, taxa may still be separated by other mechanisms such as phenology, behavior, or genetic incompatibility (Coyne and Orr 2004).

Our next problem is whether we can use distribution data to reliably infer changes in environmental requirements. The results presented here indicate that existing methods vary widely in their quality. Some methods such as comparisons of model accuracy using GLMs and the background overlap test were reasonably robust, whereas others, particularly the random overlap permutation test of Warren et al. (2008) and MANOVA, were prone to infer significant differences in their absence. Moreover, the validity of our inferences depended on the environment examined, suggesting that distribution data may only represent a useful tool for testing for differences between species in some circumstances. This may mean that distribution data are more useful in some clades than others. For example, some adaptive radiations are associated with heterogeneous environments with many allopatric populations (Hodges and Arnold 1994). In these cases, distribution data would be prone to infer divergent environmental requirements in its absence.

The results presented here investigate whether distribution data can be used to infer that species have identical environmental requirements. However, they have implications for a related issue, whether species’ ecologies are conserved over evolutionary time (Losos 2008;
Warren et al. 2008). A number of recent studies have concluded that the environments organisms occupy can change rapidly (Broennimann et al. 2007; Evans et al. 2008) and suggest that this may be the result of dramatic evolutionary change. In this study, species pairs had perfectly conserved environmental requirements, but the predictions form niche models of each species could be no more similar than expected by chance. If we are to use distribution data to test hypotheses about niche conservatism, it is essential to identify the type of conservatism of interest. Distribution data seem to provide a great deal of information on the similarity of the environments occupied by different species (similarity of an organism’s realized niche). However, when we are interested in changes in environmental requirements or other forms of genetic differentiation, distribution data may only provide reliable inferences in some circumstances.

SUPPLEMENTARY MATERIAL

Supplementary material can be found at http://www.sysbio.oxfordjournals.org/.

FUNDING

Financial support was provided by National Science Foundation grant DEB-0516841. I was supported by a PGS-D fellowship from the Canadian National Science and Engineering Research Council with a supplement from the Foundation for Orchid Research and Conservation. Bioinformatics facilities were supported by the initiative for bioinformatics and evolutionary studies at the University of Idaho, National Institutes of Health/National Center for Research Resources P20RR16448 and P20RR016454. A part of this work was conducted while a Postdoctoral Fellow at the National Institute for Mathematical and Biological Synthesis, an Institute sponsored by the National Science Foundation, the U.S. Department of Homeland Security, and the U.S. Department of Agriculture through National Science Foundation Award #EF-0832858, with additional support from The University of Tennessee, Knoxville.

ACKNOWLEDGMENTS

Jeffery Evans, Melanie Murphy, Jeremy B. Yoder, Dan Warren, Chris Smith, Luke Harmon, Olle Pellmyr, Lydia Gentry, and 3 anonymous reviewers provided helpful comments.

REFERENCES

Godsoe W. 2010. I can’t define the niche but I know it when I see it: a formal link between statistical theory and the ecological niche. Oikos. 119:53–60.
Mayr E. 1968. The role of systematics in biology: the study of all aspects of the diversity of life is one of the most important concerns in biology. Science. 159:595–599.