EXPERIMENTAL VERIFICATION OF ECOLOGICAL NICHE MODELING IN A HETEROGENEOUS ENVIRONMENT

JESSICA W. WRIGHT,1 KENDI F. DAVIES,2 JENNIFER A. LAU,3 ANDREW C. MCCALL,4 AND JOHN K. MCKAY5

Center for Population Biology, One Shields Avenue, University of California, Davis, California 95618 USA

Abstract. The current range of ecological habitats occupied by a species reflects a combination of the ecological tolerance of the species, dispersal limitation, and competition. Whether the current distribution of a species accurately reflects its niche has important consequences for the role of ecological niche modeling in predicting changes in species ranges as the result of biological invasions and climate change. We employed a detailed data set of species occurrence and spatial variation in biotic and abiotic attributes to model the niche of a native California annual plant, 
Collinsia sparsiflora. We tested the robustness of our model for both the realized and fundamental niche by planting seeds collected from four populations, representing two ecotypes, into plots that fully represented the five-dimensional niche space described by our model. The model successfully predicted which habitats allowed for 
C. sparsiflora persistence, but only for one of the two source ecotypes. Our results show that substantial niche divergence has occurred in our sample of four study populations, illustrating the importance of adequately sampling and describing within-species variation in niche modeling.

Key words: ecological niche modeling; edaphic variation; GAM models; local adaptation; niche differentiation; serpentine soils; transplant experiments.

INTRODUCTION

The niche of a species is the set of environmental variables that determine the geographic distribution of that species, either in the presence of biotic interactions including competition (the realized niche), or without these biotic interactions (the fundamental niche) (Hutchinson 1959, Pulliam 2000, Holt 2003). Ecological niche models use various mathematical techniques to relate the occurrence of species to environmental data (Guisan and Zimmerman 2000). Niche modeling has received increased attention recently because it has important implications for conservation and management efforts (Kearney and Porter 2004, Chefaoui et al. 2005), the spread of invasive species (Peterson and Vieglais 2001, Peterson 2003) and the response of species to global climate change (Ackerly 2003, Peterson et al. 2004).

However, niche models often ignore genetic variation in habitat use and the evolutionary potential for niches to diverge among populations of a species (Peterson and Holt 2003). Spatial variation in the edaphic and biotic variables that comprise a species’ niche can produce selection, ultimately resulting in adaptation to local environmental conditions (provided that sufficient genetic variation exists). Such local adaptation is well documented and can produce divergence among populations of a single species with respect to niche dimensions. Finally, local adaptation, or within species niche divergence, has played an important role in theoretical models for the evolution of species niches (Turesson 1922, Clausen 1962, Holt 2003), speciation (Darwin 1859, Maynard Smith 1966), and the maintenance of genetic variation (Levene 1953, Felsenstein 1981).

There are several approaches to evaluating the ability of a niche model to predict the potential niche of a species (Guisan and Zimmerman 2000). Here we use an experimental technique that we have not found applied in the niche modeling literature—to test the model directly by planting individuals into well-described habitats. If the model represents a meaningful characterization of the realized niche, then it should also predict transplant performance as a function of the measured environmental variables.

Our approach proceeded in three steps. First we collected fine-scale, spatial, environmental data, which we used to construct a model describing the current niche of 
C. sparsiflora. Second, we tested this model by planting seeds into well-characterized environments. Finally, we expand our exploration of the model by
(1) comparing the performance of plants collected from different habitats to determine how intraspecific variation may influence the applicability of ecological niche models and (2) by considering plants growing with and without competition, thus providing insight on the relationship between the fundamental and realized niche.

METHODS

Study system

In the North Coast range of California, habitat types (chaparral, grassland, oak woodland) and soil types (valley sediment, volcanics, serpentine) can change in a matter of meters (Kruckeberg 1984, Stebbins and Hrusa 1995). The McLaughlin University of California Natural Reserve is located in the North Coast range of north-central California. It is principally an open oak woodland, with serpentine chaparral, serpentine meadows, and non-serpentine meadows and woodlands (more information available online)\(^6\).

The native California annual plant, Collinsia sparsiflora (Scrophulariaceae s.l.) occurs on both serpentine and non-serpentine soils at the McLaughlin Reserve. Previous work has shown that there are serpentine and non-serpentine ecotypes occurring in the two soil types at the reserve (Wright et al. 2006).

Environmental data

In 2001, a 600 × 550 m habitat sampling grid was established at the McLaughlin Reserve, with grid points every 50 m. Within the grid, six smaller grids (gridlets) were established, with gridlet points every 10 m (100 × 50 m; Fig. 1). While the grid placement was random, the placement of the gridlets was selected so as to sample areas of dramatic vegetative transitions between putatively serpentine and non-serpentine soils. At each grid point \((n = 528)\), a suite of environmental data were collected, including soil chemical composition and texture (22 variables, see Wright et al. [2006] for a complete description), soil depth (at five random locations within 2 m of the central marker by hammering a 12 mm diameter, 120 cm long, steel rod into the ground until bedrock was encountered or up to 60 cm),

\(^6\) (http://www.nrs.ucdavis.edu/mclaughlin.htm)
slope, aspect, aboveground standing biomass (all vegetation in a 30 cm² area was clipped at the soil surface, dried and weighed), and species occurrence (presence/absence, sampled in the late spring). These points were chosen at random with respect to the distribution of *C. sparsiflora* and by chance included 46 points where the species is present.

**Modeling the niche**

To model potentially nonlinear relationships between candidate explanatory environmental variables and *C. sparsiflora* occurrence, we fitted a generalized additive model (GAM; Guisan et al. 2002), which is a special type of generalized linear model (GLM; McCullagh and Nelder 1989). GAMs are extensions of linear regression models that, for our purposes, allowed us to model the binomial presence–absence data in relation to environmental variables. The strength of GAMs is their ability to deal with highly nonlinear and non-monotonic relationships between species occurrence and abiotic variables (Guisan et al. 2002). All analyses were completed using Genstat, Version 8.1 (VSN International Ltd., Hemel Hempstead, UK).

Backward elimination based on all measured environmental variables was impossible due to the large number of explanatory variables and the amount of collinearity between them. Thus, we built models from the total pool of potential explanatory variables, using an iterative process. First, we used an exploratory phase to detect the variables with the largest effect. Initially, all explanatory variables were fitted one at a time and their significance in predicting the presence or absence of *C. sparsiflora* was noted. This gave a broad indication of which variables were likely to be important in later models. Then, we conducted a conditional test of significance for each variable by dropping it from the fullest possible niche model, which included all significant variables identified in the exploratory phase. *P* values were calculated from deviance ratios, and a variable was considered significant when *P* < 0.05. The final niche model included the significant variables thus identified. We incorporated spatial autocorrelation by including the number of points occupied by *C. sparsiflora* of the eight points that directly surrounded each sample point in the model.

**Testing the model: establishing experimental plots**

To test the predictive ability of our niche model we selected 100 grid points to use for our experimental manipulations. To select these points, we used the Gower dissimilarity measure to calculate the dissimilarity of each point to all other points for the five environmental variables that were determined to describe the realized niche of *C. sparsiflora* (Gower 1971). We then ranked all of the distances and selected the first 100 points (i.e., the points that were most dissimilar from one another in five-dimensional niche space). The presence of naturally occurring *C. sparsiflora* was not included in the selection of test grid points as we were using a manipulative approach, that is, using the model to predict the survival of transplants, not the presence or absence of naturally occurring plants.

At each point, 20 toothpicks holding *C. sparsiflora* seeds were planted into a 25 × 20 cm plot, with minimal disturbance of the background vegetation. To sample the range of available genetic variation in *C. sparsiflora*, our plots each included seeds from four source populations, all within 1 km of the study area, two from serpentine and two from non-serpentine soils (S1, S3, NS1, NS2; see Wright et al. [2006] for a full description of the sites). These populations have been shown to be ecotypically differentiated by soil type (Wright et al. 2006). Seeds were initially collected from the four field populations and reared in the greenhouse for one generation on a standard soil-less growing medium and allowed to set selfed fruit. This greenhouse generation was used to reduce maternal environmental effects between plants collected in serpentine and non-serpentine habitats. Two *C. sparsiflora* seeds were glued onto each toothpick, with five replicate toothpicks from each of the four *C. sparsiflora* populations, two serpentine and two non-serpentine. Two seeds were placed on each toothpick because we wanted to have as many experimental plants as possible. If two seedlings germinated at a given toothpick, we recorded their germination and removed one seedling at random. Each transplanted seed in each of our plots was evaluated for emergence and survival to flowering.

**Testing the model: data analysis**

Based on the five measured environmental attributes found to be important in our model, we predicted the establishment success of *C. sparsiflora* at each grid point used in the transplant experiment. We then used logistic regression to compare the predicted establishment success to the observed number of transplants that survived to flowering. This was done for each of the four *C. sparsiflora* populations together and separately. In our initial survey data we recorded only flowering plants. With this sampling, any sites where *C. sparsiflora* plants emerged but did not survive to flowering would not have had their presence recorded. Therefore we were comparing plants at a similar time in their growth cycle. By using proportion of plants that survived to flowering we were able to directly compare the observed data with the predicted ability to exist at each of the 100 points.

**Testing the model: the fundamental niche**

Our model was designed to describe the realized niche of *C. sparsiflora*, as it was based on presence/absence data recorded in unmanipulated sites. To test for effects of competition and other interactions with aboveground neighboring plants on limiting the realized niche, at 50 of our 100 study points we established a second plot (150 plots in total). These 50 competition removal plots were cleared of vegetation before the seeds were planted and
weeded regularly to be kept clear of plants throughout the study period, thereby removing competition and other aboveground biotic interactions associated with neighboring plants. If our model accurately describes the fundamental niche of the species, then it should predict the success of plants in cleared plots as well as uncleared plots. If the realized and fundamental niche of the species differ, then the model may only accurately describe the realized niche and model fit would be better in the uncleared plots than in the cleared plots. The data analysis proceeded as described previously for uncleared plots.

**RESULTS**

**Modeling the niche**

The model revealed five environmental factors that had significant predictive value for *C. sparsiflora* occurrence. The model predicted that *C. sparsiflora* was most likely to occur at sites within the grid with slopes between 15° and 25°, northerly aspects, intermediate soil phosphorus concentrations, little soil organic matter, and low soil calcium-to-magnesium ratios (indicating serpentine soils; adjusted $R^2 = 64.4\%$, Appendix A).

While our model describes where *C. sparsiflora* is currently found, it may have a limited ability to predict the patchy distribution. It is possible that *C. sparsiflora* occurs only at 46 out of 528 grid points because (1) it is dispersal limited; (2) it is currently occupying its realized niche, and is unable to perform well in unoccupied sites; or (3) our randomly chosen sampling points do not fully describe the distribution of *C. sparsiflora*. Further inspection of the initial environmental data revealed that the random sampling nature of the grid resulted in sampling only individuals growing on serpentine soils. Therefore our model is based only on plants found on serpentine soils, and may not represent the full niche of the species. We investigate this point further in the following section.

**Testing the model: the realized niche**

When we examined model fit for each population separately, our model quite accurately predicted where the serpentine ecotype of *C. sparsiflora* would survive to flower, but was unable to predict survival of the non-serpentine ecotype, suggesting that the niches of the two ecotypes differ (Table 1, Fig. 2). To establish a basis for comparison for the degree of niche divergence between soil ecotypes, we compared our model predictions for *C. sparsiflora* to all other species that occurred at a similar frequency as *C. sparsiflora* on the grid. Of the 11 species evaluated, the model for the *C. sparsiflora* serpentine ecotype more accurately described the distribution of four unrelated species than the distribution of the non-

<table>
<thead>
<tr>
<th>Source population</th>
<th>With competition</th>
<th>Without competition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>Combined data</td>
<td>&lt;0.001</td>
<td>0.057</td>
</tr>
<tr>
<td>Serpentine source populations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-1</td>
<td>&lt;0.001</td>
<td>0.124</td>
</tr>
<tr>
<td>S-3</td>
<td>&lt;0.001</td>
<td>0.170</td>
</tr>
<tr>
<td>Non-serpentine source populations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS-1</td>
<td>0.679</td>
<td>†</td>
</tr>
<tr>
<td>NS-2</td>
<td>0.205</td>
<td>0.006</td>
</tr>
</tbody>
</table>

Notes: A significant result shows that the model accurately predicted where *C. sparsiflora* transplants would perform well. These results show that the model works best for transplants from serpentine source populations, compared to those originating from non-serpentine source populations, both in the presence (100 test points) and absence (50 test points) of competition.

† The residual mean deviance exceeded the mean deviance of the response variable.

The positive y-intercepts indicate the potential role of dispersal limitation in *Collinsia sparsiflora*.

![Fig. 2](https://example.com/f2.png)

**Fig. 2.** There is a significant positive association between the observed and predicted establishment success for the combined data, as well as for the transplants that originated from serpentine sites. However, there was no significant ability to predict recruitment success for plants originating from the two non-serpentine sites. The small dotted line indicates a 1:1 relationship between the observed and expected observations. The positive y-intercepts indicate the potential role of dispersal limitation in *Collinsia sparsiflora*.
serpentine ecotype of *C. sparsiflora*, suggesting that the level of niche differentiation between the ecotypes is biologically significant (Table 2). Such differences in model fit between populations of a single species indicate that it is imperative to consider within-species variation when constructing niche models.

**Testing the model: the fundamental niche**

To investigate whether the fundamental and realized niche of *C. sparsiflora* differ, we also compared how our model of the realized niche predicted survival to flowering for transplants growing in 50 plots that had been cleared of vegetation to release plants from the effects of above ground competition and other biotic interactions. Using the same statistical technique described above, we determined that our model predicts survival to flowering of transplants even in cleared plots (Table 1; Appendix B: Fig. B1B). However, the model provided a better fit in the presence of competition. This result makes sense, given that the initial model was based on the observed distribution of *C. sparsiflora*, which is determined both by abiotic environmental niche axes as well as biological interactions (which in this case are correlated; J. W. Wright, K. D. Davies, J. A. Lau, A. C. McCall, and J. K. McKay, unpublished data), and therefore described the realized rather than the fundamental niche. However, the reduced fit of the model in the absence of competition provides evidence that biotic forces are also important in limiting the distribution of *C. sparsiflora* and that the fundamental and realized niches of this species differ.

A further result from this experiment is that the model of the *C. sparsiflora* realized niche under-predicted the presence of *C. sparsiflora* at sites with a low probability of occurrence for all ecotypes (Fig. 2), in both cleared and uncleared plots. This result suggests either that the distribution of *C. sparsiflora* is partially dispersal limited, (by planting seeds, we were able to consistently boost the probability of establishment and flowering at marginally appropriate sites), or, that source-sink dynamics are important in this system and that we essentially created “sink” populations that would have been unable to persist. Our experimental design does not allow us to tease apart these effects as plots were cleared of all *C. sparsiflora* seeds at the end of the growing season to prevent genetic contamination of already established populations.

**Discussion**

While some theory and prior empirical work suggest that niches should be conserved through evolutionary time (Holt and Gaines 1992, Peterson et al. 1999, Prinzing et al. 2001), the niche shift that we have observed between the two ecotypes of *C. sparsiflora* suggests that niches may evolve relatively quickly. In a niche modeling paper exploring within-species variation in niche use, Peterson and Holt (2003) found that for four species of Mexican birds, each having identified subspecies, three did not show evidence for niche differentiation between the subspecies, while the fourth did. However, Graham et al. (2004) found evidence for niche divergence between closely related taxa of Dendrobatid frogs in Ecuador, particularly along environmental axes of temperature and elevation. Our study differed from these two studies, and most GIS-based niche models, in that our sample grid points were as few as 10 m apart. This difference in scale may have implications for the interpretation of our results. It is possible that niche models based on larger scale data are more likely to show niche conservatism than those based on a smaller scale. Given the lack of studies based on a smaller scale and the observation that the scale of environmental heterogeneity and adaptation varies among species (McKay et al. 2005) more investigation is needed on the influence of scale of habitat sampling on the accuracy of niche modeling. In any case, it is clear that it is important to be able to model the distribution of species at all spatial scales, particularly in light of global climate change, invasive species and other human-altered habitat changes.

From an evolutionary perspective, we argue that our populations are potentially interbreeding and therefore these populations are in sympathy. Gene flow via seed dispersal and particularly by pollinators seems likely within this 1-km area, particularly given the fact that flowers are visited by a range of bees, including bumble and honey bees (J. Wright, personal observation). Our results, coupled with unpublished data indicating that inter-population crosses show evidence of post-zygotic barriers to gene flow between ecotypes (L. C. Moyle, J. W. Wright, and M. T. Levine, unpublished manuscript),

**Table 2.** Model predictions for 11 species that occurred at a similar frequency as *Collinsia sparsiflora* on the grid.

<table>
<thead>
<tr>
<th>Species</th>
<th>Family</th>
<th>Slope</th>
<th>P</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Avena fatua</em></td>
<td>Poaceae</td>
<td>–</td>
<td>0.002</td>
<td>0.065</td>
</tr>
<tr>
<td><em>Bromus hordeaceus</em></td>
<td>Poaceae</td>
<td>–</td>
<td>&lt;0.001</td>
<td>0.086</td>
</tr>
<tr>
<td><em>Lolium multiformum</em></td>
<td>Poaceae</td>
<td>–</td>
<td>0.005</td>
<td>0.048</td>
</tr>
<tr>
<td><em>Nassella pulchra</em></td>
<td>Poaceae</td>
<td>–</td>
<td>&lt;0.001</td>
<td>0.124</td>
</tr>
<tr>
<td><em>Taeniatherum caput-medusae</em></td>
<td>Poaceae</td>
<td>–</td>
<td>&lt;0.001</td>
<td>0.190</td>
</tr>
<tr>
<td><em>Vulpia microstachys</em></td>
<td>Poaceae</td>
<td>+</td>
<td>0.004</td>
<td>0.025</td>
</tr>
<tr>
<td><em>Dichelostemma capitatum</em></td>
<td>Liliaeae</td>
<td>+</td>
<td>&lt;0.001</td>
<td>0.069</td>
</tr>
<tr>
<td><em>Holostemma caperata</em></td>
<td>Plantaginaceae</td>
<td>+</td>
<td>0.321</td>
<td>0.121</td>
</tr>
<tr>
<td><em>Achyrachneus mollis</em></td>
<td>Asteraceae</td>
<td>–</td>
<td>&lt;0.001</td>
<td>0.300</td>
</tr>
<tr>
<td><em>Agoseris heterophylla</em></td>
<td>Asteraceae</td>
<td>+</td>
<td>0.001</td>
<td>0.025</td>
</tr>
</tbody>
</table>

Notes: The model more accurately described the distribution of four unrelated species (in boldface type) than the distribution of the non-serpentine ecotype of *C. sparsiflora*. A positive slope indicates a positive association between the predicted and observed distributions, while a negative slope suggests that model predictions tended to be opposite to observed distributions.

† The residual mean deviance exceeded the mean deviance of the response variable.
suggest the potential role of niche divergence (i.e., local adaptation) in speciation among serpentine and non-serpentine populations (Levin 2005).

The experimental nature of our data allows us to investigate ecological mechanisms that may influence C. sparsiflora distribution. Seed dispersal appears to be limited in this species. C. sparsiflora seeds have no obvious dispersal mechanism other than gravity (J. Wright, personal observation). Fig. 2 illustrates this dispersal limitation as there is a modest observed success of transplants in sites predicted to have no C. sparsiflora. There are two possible explanations of this result: dispersal limitation and competitive exclusion. However, we observed similar results for both cleared and uncleared plots, suggesting that it is indeed dispersal limitation, as there should be no competitive exclusion in a competition-free environment. Source–sink dynamics may also be relevant, as the sites predicted to have no C. sparsiflora may be “sinks.” Longer-term observations of the ability of each site to establish a population would allow us to examine the relative role of seed dispersal and source–sink dynamics.

Our experimental data indicate that ecological niche models can accurately predict the success of C. sparsiflora populations over very small spatial scales. Such experimental manipulations are necessary for investigating the mechanisms underlying observed species distributions. For example, only by experimentally transplanting seeds into well-characterized environments were we able to investigate the roles of intra-specific variation, dispersal limitation, and competition in limiting the distribution of a common annual plant. Such experiments are imperative for expanding the relevance of ecological niche models and to determine what factors influence species distributions. A mechanistic understanding of factors underlying species niches and distributions may aid in more accurately predicting species responses to environmental change.

ACKNOWLEDGMENTS

We thank K. Rice for his help throughout this project. H. M. Easlon, R. Gertz, C. Jorgensen, T. Park, K. Rice, and J. Sambatti provided assistance in field. K. Rice, M. Stanton, P. Gepts, R. Grosberg, S. Harrison, and S. Collinge provided their helpful insight during the design of this experiment. D. Ackerly, R. Grosberg, S. Harrison, M. Holyoak, L. Moyle, B. Melbourne, K. Rice, M. Stanton, and an anonymous reviewer provided helpful comments on this manuscript. We also thank the McLaughlin Reserve managers P. Aigner and C. Koehler as well as the University of California Natural Reserve System for their assistance throughout this project. This research was funded through generous grants by the Andrew Mellon and the David and Lucile Packard Foundations.

LITERATURE CITED


Levene, H. 1953. Genetic equilibrium when more than one ecological niche is available. American Naturalist 87:331–333.


APPENDIX A

A figure showing the results of the GAM model used to predict *Collinsia sparsiflora* survival (*Ecological Archives* E087-148-A1).

APPENDIX B

A figure showing the association between the observed and predicted establishment success for the combined data, as well as for the transplants that originated from serpentine and non-serpentine sites (*Ecological Archives* E087-148-A2).