Limits to range expansion in the native annual legume *Chamaecrista fasciculata*

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

John Louis Stanton-Geddes

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Ruth Shaw and Peter Tiffin

August 2011

Acknowledgements

I gratefully acknowledge the advice, guidance, patience and encouragement of my advisors, Ruth Shaw and Peter Tiffin. There's no doubt that without them, this work would not be of the quality it is and would most likely not exist.

My undergraduate advisor, Jill Miller at Amherst College, was the person who got me into evolutionary biology research, and she has been a great friend and resource ever since. Everett Dietrick of Rincon-Vitova Insectaries showed me that biology can be a lifetime passion and offered a perspective outside of academia how biology can impact the world.

My friends, colleagues and lab members have all been tremendous resources through this process, and made this a lot more fun.

My research has been generously supported by many sources, including the Rothman Fellowship, Carolyn Crosby Fellowship, Alexander and Lydia Anderson Fellowship, Center for Community Genetics and the University of Minnesota Graduate School. My advisors yet again were fantastic and provided many years of summer support and a year of RA support to complete my research.

Finally, I thank my family. My grandparents are inspirations in pursuing your passion. My parents supported me throughout this process, giving me a place to stay while on the road and even letting me use their yard as a field site! Most importantly, my wife Hilary for everything she does. She put up with me during all these endeavors, and came out to help with field work each year even when she knew it would just be ticks and mosquitoes. Thanks Hil!

Dedication

All happy families are alike; each unhappy family is unique in its own way
-Tolstoy

All species have restricted ranges; each range edge is unique
in its own way.

This thesis is dedicated to ecologists and evolutionary biologists who did the work and wrote the questions that motivate me, and biologists of the future who will further this work. And finally, my daughter Madeleine who arrived just in time for me to finish this thesis, and will soon be pursuing her own passion (maybe biology!).

Abstract

Species range limits are determined by historical (e.g., range expansion), ecological (e.g., biotic interactions) and genetic (e.g, gene flow) processes, but comprehensively understanding the relative role of these processes in limiting any single species' range has been elusive. This research is timely for understanding species' responses to climate change. The goal of this research was to examine the processes that limit the range of the native annual legume *Chamaecrista fasciculata*, by integrating ecological-genetic field studies and population genetic laboratory studies. In Chapter 1, I investigate the extent to which *C. fasciculata* is in demographic range edge equilibrium at its western and northern range edges, and the effect of biotic interactions at these range edges. I find that *C. fasciculata* fitness is reduced to zero when planted beyond the western and northern range limits, indicating it is in equilibrium with its range. Neighbors increase early-season survival, but decrease seedpod production. The goal of Chapter 2 was to examine if the mutualism between *C. fasciculata* and its associated rhizobia was disrupted beyond the range edge, potentially limiting range expansion. The results demonstrate that compatible rhizobia are nearly absent beyond both range edges, which may limit range expansion. In Chapter 3, I ask how the habitat where *C. fasciculata* establishes may change with range shifts. I conclude that habitat type influences *C. fasciculata* fitness, but the outcome depends on both the substrate and competitive environments. Finally, in Chapter 4, I use population genetic methods to gain insight into the history of range expansion, population structure and gene flow. Population genetics indicate that the edge populations have reduced genetic diversity compared to the southernmost interior population, and are highly differentiated from each other. However, there is little evidence for contemporary gene flow between populations at the scale investigated. Overall, this work suggests that ecological-genetic or metapopulation dynamics are likely to be involved in setting the northern and western range limits. Further, it highlights the value of integrated approaches to studying species' range limits.

Table of Contents

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Introduction

The question of why species ranges are restricted is a challenge to evolutionary biologists: *what limits species from expanding their ranges by adaption to conditions beyond the range edge*? Though this question has captivated biologists since Darwin (1859), identifying the proximate ecological and ultimate evolutionary limits underlying range limits in any single species has been elusive (Gaston 2009). Few general patterns to range limits have been identified (e.g., a decline in fitness or abundance towards the range edge), and even these have many exceptions (Gaston 2009, Sexton et al. 2009). This lack of progress is due to both inadequate research and the complexity of the problem. Understanding what limits species' ranges in a comprehensive fashion requires understanding the history of species range expansion, the role of ecological interactions in different geographical regions, and the scale of evolutionary processes such as gene flow and adaptation. Compounding this difficulty, the ecology and evolution of species' distributions were often studied in isolation from each other, and only recently has progress been made to integrate these studies (Gaston 2003). This integration is partially due to the development of experimental and analytical methods (e.g., population genetics, GIS), but most of the progress has come from development of theoretical models of species range limits (reviewed in Sexton et al. 2009). Driving this progress has been the appreciation that understanding the factors controlling the distributions and abundances of species is not only fundamental to our understanding of ecology and evolution, but also crucial for accurate prediction of how species will respond to climate change.

Ecological and evolutionary studies provide independent, but complementary, information on species range limits. Ecological studies of population demographics offer insights on population growth rates in different geographic regions, incorporating many, often unmeasured, environmental factors. Transplant studies beyond the range edge can be used to evaluate the extent to which fitness is reduced beyond the range edge, or infer dispersal limitations if fitness is not reduced. Further, the effect of *a priori* selected environmental factors can be tested on individual fitness. However, associations between

species distributions and environmental factors are manifestations, not causes, of range limits (Antonovics 1976).

Evaluating the ultimate causes of range limits requires evolutionary studies. Genetic causes of range limits fall into two categories. First, populations may lack adequate genetic variation for adaptation to conditions beyond the range edge (Antonovics 1976, Hoffman and Blows 1994, Stanton-Geddes 2011), possibly because of population bottlenecks associated with range expansion (Pujol and Pannell 2008) or genetic constraints (Etterson and Shaw 2001). The role of these factors can be tested in ecological-genetic transplant studies. Second, asymmetric gene flow from interior populations to edge populations may constrain adaptation to local conditions at, and beyond, the range edge (Haldane 1956, Mayr 1963, Antonovics 1976, Kirkpatrick and Barton 1997, Bridle and Vines 2007). This hypothesis can be indirectly tested through population genetics. Moreover, population genetic studies can provide evidence of species range expansion, metapopulation processes, and gene flow among populations, which can be used to evaluate alternative causes of range limits.

A number of insights have come from the integration of ecological and evolutionary studies of species range limits. First, though a decline in fitness from the range interior to edge is assumed in ecological-genetic models of species range limits, a decline is not always observed. For example, of the 14 studies that examine lifetime fitness, Sexton et al. (2009) report that 4 studies do not support this pattern, and for intermediate components of fitness (e.g. survival, fecundity) a full 43% do not decrease towards the range edge. This could be because the duration of these studies has not been adequate to sample the environmental variability at different range locations (Gaston 2009). However, it also indicates that ecological-genetic theories may not adequately explain the stability of range limits in many species.

Second, diversity is often reduced in marginal populations (Eckert et al. 2008), suggesting that edge populations may have experienced population bottlenecks or frequent extinction- recolonization metapopulation dynamics reducing genetic diversity. However, how these measures of neutral molecular diversity relate to ecologically-

relevant quantitative genetic variation is unclear (Reed and Frankham 2001). Finally, the relative roles of abiotic and biotic factors for range limits are not well understood. Most transplant studies control the biotic environment to isolate the effects of climate, potentially overestimating fitness. Further, demonstrating that fitness declines beyond a species' range edge in the absence of competitors does not preclude the hypothesis that adaptation to those conditions is constrained by the presence of competitors (Case and Taper 2000, Case et al. 2005, Price and Kirkpatrick 2009). In fact, across gradual climatic gradients, a negative effect of biotic interactions will result in stable range limits in the presence (Case and Taper 2000) or absence (Price and Kirkpatrick 2009) of gene flow when species would otherwise adapt to fill all the available climate space.

The study of range limits has gathered momentum due to a need to understand how climate change will influence species abundance and distributions. In recent years, the primary method to predict how species will respond to climate change has been the bioclimate envelope approach, based on the relationship between species' current distributions and contemporary climate (Pearson and Dawson 2003). These models are used to project species distributions onto past and future climate scenarios. This method ignores the potential uncoupling of species interactions (Both et al. 2006), evolutionary change (Davis et al. 2005), and the considerable heterogeneity in habitats within geographic regions. An alternative to range shifts for species persistence would be shifts in habitat within a species' current range, likely with adaptation and changes in abundance. In fact, given the plethora of examples of glacial refugia and shifts in species habitat at the edges of their current ranges, it is possible that this will be one of the first observable changes for many species in response to climate change (Jump et al. 2009).

In my dissertation research, I build on the research of species range limits to examine the abiotic and biotic factors that limit the range edge of the native annual legume *Chamaecrista fasciculata*. Further, I evaluate how this species may respond to climate change by range and habitat shifts. I address these goals through multiple field experiments and laboratory work. The research in *Chapter 1 asks whether C. fasciculata is in range edge equilibrium across both its northern (pole-ward) and western range*

edges, using multiple approaches to examine the entire life history of the species and estimate population growth rate in each geographic location. Further, I manipulated the presence of competitors to examine their effects on fitness at each site, and investigated the role of herbivores and disease.

The role of species interactions has featured prominently ever since the question of "what limits species ranges" was first formulated (Darwin 1859). The focus has been primarily on interactions that directly *reduce* an individual's fitness (i.e., competitors, predators). The research *in Chapter 2 asks how the disruption of a mutualism beyond the range edge may limit individual fitness*, and thus the potential for range expansion.

With changes in geographic range location, many aspects of the environment change. Thus, the habitats that species occupy may change with range shifts. The research in *Chapter 3 investigates how the dependence of fitness on habitat type changes with geographic range location* by simulating a colonization event into two habitat types, sand and loam soils, within *C. fasciculata*'s range and at and beyond its northern range limit. Further, I ask how population source influences fitness at these different range locations and habitats, and examine phenotypic selection on ecologically-relevant traits at these sites.

Insights on species range limits can also be gained from information on population history. The research *in Chapter 4 uses population genetics methods to infer patterns of population establishment, population structure, and gene flow between populations of C. fasciculata within its range and at its western and northern range edges*. I interpret these results in the context of the findings from the previous chapters.

In summary, I integrate ecological genetic and population genetic studies to come to a greater understanding of a fundamental question in evolutionary ecology: what limits species' ranges? Given the complexity of the many factors underlying species range limits, I do not unequivocally determine why *C. fasciculata* reaches its range limit where it does, but I exclude some hypotheses, highlight likely explanations, and highlight avenues for further exploration.

Chapter 1

Fitness variation across the range edge of an annual plant: the role of climate and biotic factors

Summary

Ecological-genetic theories of range limits assume that species distributions match their climatic tolerances (i.e. range edge equilibrium), but this assumption is not always supported. Moreover, the relative importance of abiotic and biotic factors for limiting species range edges is not well understood. We test the extent to which fitness of a widespread native annual legume, *Chamaecrista fasciculata*, is in range edge equilibrium with its northern and western range edges in two experiments. First, we examine plant fitness and the effect of neighbors in natural populations at different geographic range locations for three years. Fitness decreases towards the northern, but not western, range edge, and competitors have a consistent negative effect on fitness across years and sites. Second, we use a novel approach to measure both individual fitness and population growth rate in experimental populations planted at sites within the range and at and beyond the northern and western range edges, and tested the effect of competitors with a neighbor removal treatment. Fitness declined to zero beyond both range edges, indicating that *C. fasciculata* will not establish in these regions. The effect of biotic interactions is multifaceted. Neighbors increase early-season survival, but reduce seed production of reproductive individuals at all range locations. Herbivores reduce seedpod production of the Interior experimental population, potentially acting as a density-dependent process limiting population growth in this region.

Introduction

Species survival in response to environmental instability has and will always depend on changes in population abundance, adaptation, habitat shifts and range shifts. However, predictions of how species will respond to contemporary climate change are overwhelmingly based on range shifts only. Projections of current species distributions into future climate scenarios suggest that many species are at risk extinction due to loss of suitable climatic conditions, even with unlimited dispersal (Thomas et al. 2004, He and Hubbell 2011).

These predictions make a number of assumptions (Pearson and Dawson 2003), notably that species' distributions are primarily limited by climatic conditions. Yet it is unclear to what extent species' current distributions match their potential geographic distribution (i.e., range edge equilibrium) because most geographic regions have experienced climatic instability during Quaternary climate change (Williams et al. 2000). Moreover, abiotic factors such as soil type and biotic factors such as competitors and enemies are also important for determining species' current distributions.

Range edge equilibrium can be tested by examining the fitness of individuals within and beyond the species' current range using transplant common garden experiments. Finding that population growth rate is reduced below replacement ($\lambda < 1$) beyond a species' current range edge would indicate that the species is in range edge equilibrium. However, few studies have examined how lifetime fitness varies from range interior to edge and beyond. Of the 14 studies that do so reported in a review by Sexton et al. (2009), ten find the expected decline in lifetime fitness (though they report a much greater proportion of studies that fail to find the expected decrease when examining individual components of fitness only). Further, recent estimates of tree migration rates indicate that many species are not in range edge equilibrium (Svenning and Skov 2007). This evidence, taken together, presents a challenge to the dominance of theoretical models of range limits that assume species are in equilibrium with their ranges (e.g.,

Haldane 1956, Kirkpatrick and Barton 1997, Case and Taper 2000), but clearly more work is needed.

Species may not be able to reach equilibrium with their climatic tolerance limits for a number of reasons. First, dispersal barriers will prevent species from expanding to fill the available climate space (e.g., Gilman 2006, Marsico and Hellmann 2009). The rapid spread of invasive cane toads in Australia is a classic example of how species can rapidly expand their ranges after overcoming a dispersal barrier (Phillips et al. 2006a). Second, low dispersal rates intrinsic to a species' biology may limit expansion. For example, species primarily dispersed by ants, such as *Trillium,* have short dispersal ranges that reduce their expansion potential, even accounting for infrequent long – distance dispersal events (Vellend et al. 2003). Third, lack of available habitat can reduce the potential for a species to reach climatic equilibrium (Prince and Carter 1985, Holt and Keitt 2000). Species endemic to serpentine soils are a pertinent example, as they are restricted by soil conditions, and not climate. Fourth, biotic interactions may prevent individuals of a species from establishing in regions where they could otherwise persist (Price and Kirkpatrick 2009). While the role of biotic interactions in limiting species distributions is commonly cited (Darwin 1859, MacArthur 1972, Hochberg and Ives 1999, Case and Taper 2000, Price and Kirkpatrick 2009), it is rarely investigated in empirical studies.

In this paper, we evaluate the extent to which the native annual legume *Chamaecrista fasciculata* is in range edge equilibrium with its climatic tolerances at its northern and western range edges. To investigate this, we use aster models (Geyer et al. 2007, Shaw et al. 2008) to estimate individual lifetime fitness and population growth rate at common garden sites within and beyond the species' current range edge. Finding that individual fitness or population growth decreases below replacement across the range edge would indicate that *C. fasciculata* will not establish beyond the range edge. Further, we examine how biotic interactions modulate the potential for *C. fasciculata* to expand into climatically suitable regions. We investigate the effect of neighboring plants on plant fitness in both natural populations and experimental plants within and beyond the current

range edge. Additionally, we consider the role of herbivory and disease on plant fitness within and beyond the current range edge. Finding a significant negative effect of competitors, herbivores or disease on fitness beyond the range edge would imply that the distribution of *C. fasciculata* is limited not by climatic tolerance alone, but is limited by the interaction of climate and biotic factors.

Materials and Methods

Chamaecrista fasciculata (Fabaceae), which is native to North America, has an extensive geographic range from southern Mexico in the south to north-central and eastern United States in the north. The approximate northern range limit for this species occurs from southern Minnesota east to Connecticut and the western range limit is from western Minnesota south through Texas into Mexico, based on county-level observations (Kartesz and Biota of North America Program (BONAP) 2011). The southern range limit is unknown. Neither the northern or western range limit occur at distinct geographical barriers, though they do occur along isoclines of mean annual temperature and precipitation, respectively (Hijmans et al. 2005). Populations are distinct and found in old fields, open woodlands, disturbed prairies, and roadsides. At both its northern and western range edges, *C. fasciculata* is often found restricted to sandy sites (Irwin and Barneby 1982, Ownbey and Morley 1993, Chapter 3) *C. fasciculata* is insect-pollinated and self-compatible, but highly out-crossing (Fenster 1991a). Seeds are explosively dispersed from the seedpods in the fall, and typically reach only a few meters from the plant (Fenster 1991a), and there is a limited seed bank with more than 90% of the seeds that will germinate doing so in the first year (Fenster 1991b).

Natural populations: inter-annual fitness and the effect of neighbors at different range locations

In June 2007, we located four populations of *C. fasciculata* from its northern range edge in Minnesota south to Kansas: (i) Grey Cloud Dunes Scientific and Natural

Area in Cottage Grove, Minnesota (MN1), (ii) Conard Environmental Research Area, Grinnell College, Kellogg, Iowa (IA), (iii) Green River Wildlife Area, Harmon, Illinois (IL) and (iv) Konza Prairie Biological Station, Kansas State University, Manhattan, Kansas (KS) (Fig. 1-1, Table S1-1). Within each population, 30 pairs of plants within 50 cm of each other were randomly selected, except at the IL site where only 10 pairs were selected because of small population size in this year. One plant in each pair was assigned at random to a neighbor removal treatment and all above-ground vegetation within 20 cm of the plant was clipped at the beginning of the season (June) when seedlings were tagged. At the end of the growing season in late September and October, we returned to each site to record the number of seedpods produced by each plant. We repeated this experiment for the following two years.

In 2008, we added a population at Tyson Research Station annex (Washington University) in Eureka, Missouri (MO) and a second population at the range edge in Minnesota (MN2). Sample sizes ranged from 20 to 40 pairs of plants depending on the site and year. We collected weather data for each year from the weather station nearest to each site; either a weather station on site (IA and KS) or the nearest airport weather station (data downloaded from wunderground.com). Specifically, we examined total precipitation and average temperature during the growing season of *C. fasciculata* from 1 May to 30 September.

Statistical analyses

To investigate the effect of range location and neighbor removal on plant fitness in natural populations, we used generalized linear models (GLMs) implemented in the MASS library (Venables and Ripley 2002, Crawley 2007) in R (R Development Core Team 2009). We fit GLMs with seedpod production as the dependent variable, and site, year, treatment (NR or NP) and all interactions as predictors. To account for overdispersion in the data, we fit with a quasi-Poisson distribution. Model selection was performed by sequentially dropping terms from the full model, starting with interactions, and testing the significance by likelihood ratio tests. Further, we examined the

relationship between growing season temperature, precipitation, and seedpod production as the correlation between seedpods produced and growing season (May-October) temperature or precipitation.

Experimental populations: fitness and biotic interactions within and beyond the range

We collected seed in September and October 2008 from haphazardly selected plants (maternal families) at least three meters apart in all of the populations studied above (Table S1-1), except for MN2 due to small population size. In April and May 2009, we established five experimental populations: (i) one site within the range (Interior), (ii and iii) one site near each of the northern (N.Edge) and western (W.Edge) range edges, and (iv and v) one site beyond (Beyond.N, Beyond.W) each of these range edges (Fig. 1-1). The north and west edge sites were located orthogonally from the interior site, near the edge of *C. fasciculata*'s range based on county-level observation data (USDA Plants Database: plants.usda.gov), along gradients of mean annual temperature (MAT) and annual precipitation (PPT), respectively (Hijmans et al. 2005, Table S1-2). From the interior site to the site beyond the western range edge, annual precipitation decreases by 48%, while mean annual temperature increases by only 0.4°C. Along the gradient from south to north, the primary change is in mean annual temperature which decreases by 5.2°C. In conjunction with this temperature difference, annual precipitation decreases by 21%, but soil water availability is not expected to decrease as much because the lower temperatures reduce evapotranspiration (Loehle 1998). The beyond edge sites were located along the same orthogonal gradients and placed in suitable habitats (i.e. open fields) well beyond the recorded range limit of *C. fasciculata*. All sites were chosen to have loam soils, as soil type can influence fitness (Chapter 3).

At each site, we planted four seeds from each of 40 maternal families from each population at the interior and edge sites (n = 800 seeds total), and two seeds from each maternal family in each treatment at the beyond edge sites (n = 400 seeds total) due to seed availability. The seeds were planted in 10 blocks containing 20 (interior and edge)

or 10 (beyond edge) plots. Each plot consisted of four randomly selected seeds from a single source population planted in a circular pattern with a diameter of 20 cm. One control plot with no seed additions was included in each block to measure potential recruitment from the seed bank. Each plot was separated from the next plot by 0.5 m. Within this distance, we expected to capture most seed dispersal as > 80% of seeds are reported to land within 0.5 m of the parent plant in natural settings (Fenster 1991a). A nail was placed in the center of the circle so as to facilitate relocation (Fig. S1-1). We began planting at southern sites and moved north so as to track natural timing of germination. Seeds were sterilized with 10% sodium hypochlorite (NaOCl) and scarified with a metal file prior to planting. Dried vegetation from the previous year was removed prior to planting. Otherwise sites were manipulated as little as possible so as to truly represent natural conditions that dispersing seeds might encounter.

Each plot was randomly assigned to one of two treatments, neighbor removal (NR) or un-manipulated neighbors present (NP). The NR treatment was established by spraying glyphosate (Roundup, Monsanto, St. Louis, Missouri) on all vegetation within the 0.5 m circle surrounding the experimental population 24 hours prior to planting, and clipping vegetation immediately before planting. The effectiveness of the NR treatment was confirmed by assessing percent cover of bare ground in each plot when early-season measurements were taken. NR treatments had significantly more bare ground than NP ($F_{1,687} = 803$, $P < 0.0001$) and this varied by site, ranging from 57% more at the interior site to 195% more at the north edge site ($F_{4,683} = 49.5$, $P < 0.0001$).

We returned to each site about four weeks after planting to record early-season survival, which includes differences in germination and survival post-germination. During the middle of the growing season in July, we recorded survival, browsing status (yes/no). We scored foliar herbivory by visually assessing the percent of leaves on each plant that showed signs of herbivore damage in four classes [0%, 1-25%, 26-75%, 76-100%], and similarly recorded foliar disease. At the natural end of the growing season in late September and early October, we recorded survival, browsing status and the number of seedpods produced by each plant.

To calculate population growth rate across an entire generation, from the number of seedlings in 2009 to the number of seedlings in 2010, we returned to each site the following June, after natural germination. We searched exhaustively within 0.6 m of the center of each plot (Fig. S1-1). *C. fasciculata* seedlings had two to four leaves at this time and were easily visible. Germination of seeds that had been planted the previous year was unlikely to contribute to seedling counts in 2010 as we found only one seedling at a toothpick marking locations of seeds planted the previous year. Further, we found no seeds outside the blocks at each site, thus all seedlings recorded are from our experimental populations.

Statistical analyses

To determine the effects of geographic range, source population and competitors on plant fitness, we modeled lifetime fitness for each individual using aster models (Geyer et al. 2007, Shaw et al. 2008) in R (R Development Core Team 2009). Our aster model integrated across four life history stages modeled with appropriate statistical distributions (Fig. 1-2A). We fit aster models with fixed effects for site, population, competition treatment and all interactions. We tested whether each interaction improved the fit of the model to the data using likelihood ratio tests comparing sub-models to the full model. Finding a significant site × treatment interaction would suggest that the effect of competitors varies by geographic region. To investigate at which life history stages each term had a significant effect on fitness above and beyond overall differences in fitness, we performed step-wise forward model comparison beginning with a model that included each main effect at the level of overall fitness (e.g., seedpods). We added a term for each main effect specified at each predecessor life history stage and used likelihood ratio tests to evaluate if the term was significant. Step-wise forward model building was performed to avoid selecting an over-parameterized model. To test for intra-site environmental heterogeneity, we also tested the effect of blocks nested within site. The current aster package only accommodates single-parameter exponential family distributions, so the size parameter for the negative-binomial distribution for seedpods

was determined by fitting a negative binomial distribution (fitdisrt in the library MASS (Venables and Ripley 2002)) to the conditional distribution of the seedpod data. Goodness of fit of the negative binomial distribution for seedpods was confirmed by the Pearson residuals having approximately a mean of zero and a variance of one with few outliers (Shaw et al. 2007b, Section 3.7).

To investigate how browsing intensity varied among sites, we fit a generalized linear model with binomial errors, with browsing status as the response, and site, population and competition treatment as predictors. We expected that browsing might be greater on plants in the competitor removal treatment because the plants were more apparent to herbivores. We examined how foliar herbivory and disease varied among sites, populations and treatments by fitting proportional odds logistic regression models (polr function in the library MASS), which accommodate ordered categorical responses. To evaluate the effect that browsing, foliar herbivory and disease had on fitness at each site, we fit separate aster models for each site including fixed effects for browsing status, foliar herbivory and disease categories, and covariates of population and treatment. We tested the effects of biotic interactions on lifetime fitness by comparing the full model with sub-models dropping the browsing, herbivory and disease terms.

We estimated population growth rate in two ways to set upper and lower bounds on the true value because seedling dispersal into control plots was greater than anticipated (see Results). First, we ignored the plot design, and calculated population growth at the level of each site by summing the number of seeds germinated in 2009 and the number of seedlings found in 2010, and calculated population growth rate as λ = (number seedlings in 2010) / (number seedlings in 2009). To account for the fact that 21% of each block was unsearched because we only searched within a 60 cm diameter of each plot (Fig. S1-1), we estimated the number of unobserved seedlings in each block by assuming that the density of seedlings in the unsearched area was consistent with the searched area. Though this is likely an overestimate because the area we searched was close to the parents, it sets an upper bound on our estimate of population growth rate at each site. Our population growth estimates are similar to measurements of the finite rate

of population increase (Norton et al. 2005) as we planted at low density in natural settings, and thus should give reasonable approximations of the potential for population growth after colonization.

Second, we modeled population growth rate at the level of each plot using aster models (Fig. 1-2B), which allowed us to evaluate how competition treatment influenced plot growth rate. By analyzing plot records from seeds planted in 2009 to seedlings recruited in 2010 (Fig. 1-2B), this aster model takes into account mortality from seed predators and winter conditions as well as differences in survival and reproduction in 2009. Seedlings that were found in empty patches (i.e., those that did not have reproductive plants in the previous fall) were changed to zero in the analysis as it was not possible to definitively associate them with their parent, likely making predictions of seedling production underestimates. We tested fixed effects of block, site, competition treatment and site × treatment in the model as before. Finally, we calculated population growth rate from the aster model as the predicted number of seedlings in 2010 divided by the predicted number of seedlings in 2009 for each treatment combination, and calculated confidence intervals by the delta method (see Supplemental Information, Chapter 3; details available in R script).

Results

Natural populations: inter-annual fitness and the effect of neighbors at different range locations

In natural populations, overall seedpod production was greatest in the southern sites and tended to decrease in northern sites (Fig. 1-3, 1-4). However, there was significant inter-annual variation (significant Year × Site, Table 1-1) such that this pattern varied among years. For example, in 2008 seed production steadily decreased towards the northern range edge (Fig. 1-3). By contrast, in 2009, seed production was greatest at a northern-edge site (MN1) and low at the southern-most site (MO). Neighbor removal significantly increased seedpod production at all sites, with similar effects in the different

sites (Fig. 1-4, Table 1-1). There was a non-significant positive correlation between growing season precipitation and average number of seedpods produced at each site over the three years of the experiment ($r^2 = 0.38$, $P = 0.17$; Table 1-2). This was driven by greater seedpod production in years with higher rainfall at the southern sites (MO, KS and IL; Table 1-2, Fig. 1-3). However, at the northern edge sites (MN1 and MN2), seedpod production was not related to precipitation. There was no relationship between average growing season temperature and seedpod production ($r^2 = 0.09$, $P = 0.77$).

Experimental populations: fitness and biotic interactions within and beyond the range

Lifetime seedpod production of experimental populations of *C. fasciculata* was greatly reduced beyond both the western and northern range edges compared to the edge and interior sites, implying that range expansion is not limited by dispersal alone in either region (Table 1-3, Fig. 1-5). While lifetime fitness was near zero in both regions beyond the range edge, the fitness components underlying this pattern differed. Specifically, beyond the western edge, early-season survival was low, as was the number of seedpods produced given that a plant survived and produced any pods (Fig. 1-5). However, beyond the northern edge, early-season survival was high, but the probability of a plant surviving to the end of the season was low, and very few plants that survived produced any pods. These patterns suggest that factors associated with survival, (for example, water availability) limited fitness early in the west, with resources later limiting reproduction, while beyond the northern range edge a lack of time and resources to complete the life-cycle was the primary limiting factor.

Source population only had a significant effect at early-season survival, with the CRA population having greatest early-season survival at four of five sites, while the KZA population had lowest early-season survival at three of five sites. The largely consistent effect of seed source on early-season survival suggests that maternal environment influences survival through seed provisioning. Lifetime seedpod production did not differ significantly among populations (Table 1-3), indicating that differences in early-season survival diminished across the season. This was likely because of low overall survival

and reproduction, with only 11 - 22 individuals of each population surviving to produce seeds at the interior site and fewer at the remaining sites, and thus limited opportunity for expression of population differentiation. Given this, there was a suggestive pattern with the interior TYS population having greatest or second greatest seedpod production at the interior and edge sites, but low or no seedpod production at the beyond edge sites.

The effect of the neighbor removal treatment varied with life history stage. Consistent with the results in natural populations, NR significantly increased seedpod production, given survival, at all sites except the Beyond.N site where no plants survived in the NR treatment (Fig. 1-5, Table 1-3). However, NR had a large negative effect on early-season survival (significant trt@esurv, Table 1-3) everywhere except the N.Edge site (Fig. 1-5). Surprisingly, the beneficial effect of neighbors on early-season survival was large enough for lifetime fitness to be greater in the neighbor present than the neighbor removal treatment at all sites except the N. Edge (Fig. 1-5).

The percentage of plants that were browsed was greatest at the interior site, and lower at all other sites except the W.Edge site (site, dev = 242, $P_{4,864} < 0.0001$, Fig. 1-6). At all sites, plants in the NR treatment were more likely to be browsed (treatment, dev = 22.5, $P_{1,864} < 0.0001$; Fig. 1-6) and this did not differ among sites (site × treatment: dev = 0.9, $P_{4,864} = 0.92$). There was a significant interaction between sites and treatments affecting foliar herbivory (POLR, site × treatment LR stat = 16.0, $P_{4,703} = 0.003$), with more plants being in the high herbivory categories at and beyond the western edge sites than at the interior or northern sites (Fig. 1-7A). There was also a significant site × treatment for foliar disease (POLR, LR stat = 9.5, $P_{4,703} = 0.05$), but contrasting with herbivory, the effect of disease was greater at the W.Edge site than the other sites (Fig. 1-7B).

The fitness effects of herbivory and disease did not exactly follow their prevalence. Herbivore browsing significantly reduced fitness at the interior site (Table 1-4) where it was most common, but it had no effect at the other sites. The effect of foliar herbivory was positive at the west edge site, but not significant at the other sites (Table 1-

4). Foliar disease significantly decreased fitness at the north edge and beyond western sites, but was positively correlated with fitness at the beyond northern site (Table 1-4).

Population growth rate tracked the overall pattern of fitness in the previous year, with greatest population growth at the interior site, intermediate population growth at the edge sites, and zero population growth at both beyond edge sites (Table 1-5). These estimates should be interpreted with caution as over 25% of the seedlings found in 2010 were located in plots that did not have plants that survived to reproduce in the previous year, indicating that they dispersed from another plot. The observed estimate of λ for each site is not influenced by this as it simply compares numbers of seedlings in 2010 to numbers of seedlings in 2009 at the entire site. Further, we account for any unobserved seedlings in unsearched areas of each site, thus setting an upper limit on λ.

As expected, the aster estimates based on experimental-subpopulations were lower because they did not account for seedlings that may have dispersed outside of plots, but in general, both methods gave qualitatively similar results (Table 1-5). Contrasting with estimates of lifetime fitness in 2009, aster estimates of population growth rate were greater in the NR treatments than the unmanipulated NP sites at the two edge sites, though not at the interior site. This was due to greater recruitment in these plots in 2010. Interestingly, estimates of population growth rate were below one (replacement) even at the Interior site, where a large and stable population has been present for many years.

Discussion

Species are often assumed to be in demographic equilibrium with their range edges. However, recent work has challenged the generality of this assumption (Svenning and Skov 2007, Sexton et al. 2009). We found that population growth rate of a widespread species decrease to zero beyond its western and northern range edges, indicating that it is likely in demographic equilibrium at these range limits. We emphasize that this pattern of a decline in fitness beyond the range edge is supported by

multiple lines of evidence: observations of fitness in natural populations, and estimates of individual fitness and population growth rate in experimental populations. Further, we demonstrate that the effects of interacting species can be multifaceted. While neighboring plants have a negative effect on reproduction as expected, they have a positive effect on fitness at the stage of early-season survival. Thus, if early-season survival is critical for population establishment, as our results suggest for *C. fasciculata*, then neighboring plants may facilitate range expansion.

General patterns of range limits

Our finding that populations are unable to replace themselves beyond the range edge contributes to the majority of empirical studies showing that species are in demographic equilibrium with their climatic tolerances (Sexton et al. 2009). However, given the relatively few studies that transplant organisms beyond their range and evaluate multiple life history stages, it is challenging to come to robust conclusions. Consistent with our findings, some studies do report that fitness declines below replacement beyond the range edge (Jenkins and Hoffman 1999, Geber and Eckhart 2005, Griffith and Watson 2006). However, other studies do not find such a decline in fitness (Prince and Carter 1985, Norton et al. 2005, Samis and Eckert 2009), suggesting that factors such as habitat availability, and not just quality, are important for range limits (Holt and Keitt 2000). Moreover, even in systems with declines in fitness beyond the range edge, fitness is not always lower at the range edge than within the interior (Geber and Eckhart 2005, Griffith and Watson 2005). Such abrupt range edges could be due to multiple interacting that set the range edge, or local adaptation of edge and interior populations.

Local adaptation is a common, though not universal, phenomenon (Hereford 2009). Based on previous work, we know that the KS population is locally adapted (Galloway and Fenster 2000, Etterson 2004, Chapter 4) even though it occurs near the western range edge. This indicates that local adaptation of this population in this region may allow it to maintain fitness equivalent to populations in more central regions of the range. However, in transplants beyond the western range edge, fitness sharply declines in

this region where populations have not successfully established and adapted. The same pattern is seen to a lesser extent at the northern edge, with a slight decrease in fitness in natural populations (Fig. 1-3, 1-4) but a sharp decline beyond the range edge. The slight decline of fitness in the Minnesota population may be because it has undergone bottlenecks following range expansion since the last glaciation in this region (Chapter 4). The role of local adaptation maintaining fitness at the range edge is supported by other experimental work. Populations of the annual plant *Xanthium strumarium* at its northern range edge maintained fitness equal to interior populations by adaptation to reproduce earlier (Griffith and Watson 2005), but transplants beyond the range edge completely failed to reproduce (Griffith and Watson 2006).

Our results indicate that a true test of range edge equilibrium requires transplants beyond the range edge, as inter-annual variability and local adaptation may mask fitness declines in natural populations. We do not suggest that all transplant studies will find declines in fitness beyond the range edge (e.g., Prince and Carter 1985, Norton et al. 2005, Gilman 2006, Marsico and Hellmann 2009, Samis and Eckert 2009) but that alternative hypotheses to dispersal limitation should be considered (e.g. metapopulation processes, Prince and Carter 1985) when locally adapted populations are found at the range edge, but not beyond.

Integrating multiple components of fitness

We used a novel experimental approach where we measured fitness across multiple life history stages on individuals in plots over one season, and then estimated the population growth rate of these plots with data from recruitment the following season. A caveat of this approach is that we likely underestimated population growth. As mentioned before, we missed individuals that dispersed far from the parent plants. Second, we do not account for seed dormancy, which may contribute to population growth in later seasons (Fenster 1991b). However, as our interest is on the fate of individuals that disperse beyond the range edge, where there will be no seed bank, this is not crucial to our conclusions.

There are two major benefits to this approach. First, from examining lifetime seedpod production, the overall decrease in fitness at the edge and beyond sites is clear, even though this decline in fitness is not obvious from any single life history stage (Fig. 1-5). Thus, accounting for multiple life history stages is crucial for understanding range limits, though admittedly a challenge to measure for long-lived perennial species. Second, by estimating population growth rate in experimental populations, we were able to confirm that the decrease in fitness resulted in reduced population growth rates. Our estimates of population growth rate demonstrated the populations beyond the range edge crashed to extinction in a single generation (Table 1-5). Interestingly, even at the interior site the confidence intervals for population growth rate include zero, indicating that even these populations may not persist. As growing conditions were not especially harsh in this year (Table 1-2), this low population growth rate is not likely to be due to climate. One possible explanation is that other factors constrain population growth at this site, potentially browsing by herbivores (Fig. 1-6).

A limitation of this study is that it was done in only one year, which is typical for ecological-genetic studies of annuals. This can be overcome by replicating the experiment over multiple years, which is time-intensive and rarely if ever done, using perennials which sample multiple years (van der Veken et al. 2007), or tracking population growth of annuals across multiple years (Norton et al. 2005).

Biotic interactions

The finding of increased lifetime seedpod production in the presence of neighbors compared to the neighbor removal treatment, mediated through greater early-season survival (Fig. 1-5), was surprising given that the effect of neighbors at range limits is expected to be competitive (Darwin 1859, Case and Taper 2000, Price and Kirkpatrick 2009). Such plant – plant facilitation is typically described in stressful environments, such as alpine habitats (Callaway et al. 2002). Our results suggest that this facilitative effect of neighbors can occur in favorable habitats at stressful life history stages, specifically seedling survival. Two related mechanisms could explain the facilitative

effect of neighbors. First, in arid environments, perennial shrubs act as 'nurse plants' by providing shade for annual seedlings while they grow and develop their own root zones (Went 1942, Brooker et al. 2008). Second, neighboring perennial plants with deep root zones may provide extra soil moisture by hydraulic lift, the passive movement of water through roots from deep to shallow soils (Caldwell et al. 1998). The perennial bunch grasses that often co-occur with *C. fasciculata* could provide both of these benefits. Similar to our results, Marsico and Hellman (2009) report that germination of three long-lived perennial *Lomatium* species transferred beyond their range on Vancouver Island was increased in the presence of neighboring vegetation. The effect of neighbors on population dynamics may not be solely negative in many environments. If early-survival is the key stage for stage for viable population establishment, neighbors may actually facilitate range expansion in some habitats.

Contrasting with the other sites, neighbor removal increased early-season survival and lifetime fitness at the N.Edge site. Two factors related to this site likely explain this observation. First, 2009 was an exceptionally dry year at the north edge, so competition for moisture was especially strong at this site as shown by the overall low survival rates (Fig. 1-5). Second, this site was dominated by weedy herbaceous plants (53% herbaceous cover in the neighbor present treatment), and lowest in grasses (15%). Thus, it is likely that there was greater niche overlap between *C. fasciculata* and neighbors at this site, leading to greater competition for the already limited water resources. Similarly, Carter and Prince (1985) found that another annual ruderal species, *Lactuca serriola,* did not coexist with other competitive ruderals near its range edge. Likewise, the effects of neighbors on individual fitness at and beyond the range edge will depend on neighbor identity (Tilman 1994, Fargione et al. 2003), and the effect of neighbors cannot be ignored at any range edge.

Other biotic interactions that may influence range edge equilibrium include herbivory and disease (Geber and Eckhart 2005, Alexander et al. 2007, Antonovics 2009). We found that browsing from herbivores, primarily deer, was greatest at the Interior site (Fig. 1-6), and that this was the only site where browsing significantly

influenced fitness (Table 1-4). This may be because there were more deer at this site, there were fewer attractive co-occurring plants, or because *C. fasciculata* plants were taller and more apparent to herbivores at this site than the other sites. The negative effect of browsing on fitness is likely partially responsible for the low population growth rate at this site. We also found that foliar herbivory was greater at the two western sites than the interior or northern sites (Fig. 1-7), consistent with the hypothesis of greater insect damage on drought-stressed plants (Mattson and Haack 1987). However, foliar herbivory was actually associated with increased fitness at the W.Edge site, likely because larger plants that were more likely to survive and produce more seed were also more likely to attract herbivores. (Table 1-4).

By contrast, foliar disease was greater at the Beyond.N site than the other sites (Fig. 1-6), however there was no significant effect of disease on fitness at this site. In a companion study at these same sites, we found that mutualistic rhizobia are nearly absent beyond both range edges (Stanton-Geddes and Anderson 2011), and that this reduces fitness. Thus, while the relevant factors are different, our results add to other studies that have found negative effects of biotic interactions beyond "harsh" range edges (Kavanagh and Kellman 1986, Geber and Eckhart 2005). This puts into question the partitioning of abiotic factors for limiting the harsh range edge, and biotic factors for limiting the benign range edge (MacArthur 1972).

Conclusions

Our results shows that the absolute fitness of individuals of *C. fasciculata* transplanted beyond the range edge is near zero, and populations cannot establish in these regions. This reduction in fitness may be caused by the direct effect of climate, or more likely, the compound negative impact of climate and biotic interactions. Though we find that competitors may facilitate population establishment, contrary to expectations, their negative effect at later life history stages may be sufficient to prevent populations from establishing, or other biotic interactions may limit fitness beyond the range edge (Stanton-Geddes and Anderson 2011). What prevents *C. fasciculata* from adapting to

survive and reproduce in conditions beyond these edges is unclear. Populations may lack the genetic variation necessary for range expansion (Griffith and Watson 2006). Relatedly, a paucity of genetic variation for particular trait combinations may be due to antagonistic genetic correlations between traits (Etterson and Shaw 2001). Alternatively, populations may simply not have had enough time to accumulate the mutations necessary for range expansion due to inadequate time or extinction – recolonization dynamics at the range edge (Holt and Keitt 2000).

Funding and Assistance

We thank Larissa Mottl at Conard Environmental Research Area, Jean Knops at University of Nebraska, Shawn Schottler at St. Croix Watershed Research Area, and Bryan Wood at the Audubon Center of the Northwoods for the use of field sites. Konza Prairie, Tyson Research Station, Conard Environmental Research Area and the Minnesota and Illinois Departments of Natural Resources all generously allowed seed collection. Charlie Geyer provided statistical assistance. This work was funded by the Alexander and Lydia Anderson Fellowship and a Center of Community Genetics Fellowship to J.S.G.

Tables and Figures

Table 1-1. Summary of model comparisons to tests for the effects of site, year, competition treatment and all interactions on the number of seedpods produced per plant in six natural populations across three years.

Term	Resid df	Test df	Deviance	F-statistic	P-value
Year:Site:Trt	619	8	45.9	0.32	0.96
Site:Trt	627	5	49.7	0.57	0.73
Year:Trt	632	2	3.1	0.09	0.91
Year:Site	634	8	2296.7	16.5	< 0.0001
Trt	642	1	620.2	27.3	< 0.0001
Year	642	2	1199.8	26.4	< 0.0001
Site	642	5	4231.1	37.2	< 0.0001

Table 1-2. Summary of climate data during the growing season (1 May – 30 September) for each natural population in neighbor removal experiment, arranged from south to north, with locations in Fig. 1-1. Data comes from either weather stations at field stations (IA, KS) or the nearest airport weather station available (data downloaded from wunderground.com). The two MN were closest to the same weather station.

	2007		**2008**		**2009**	
	Temp (˚C)	**Precip (mm)**	**Temp (˚C)**	**Precip (mm)**	**Temp (˚C)**	**Precip (mm)**
MO	na		22.1	598	21.8	500
KS	17.3	561	21.9	823	21.3	574
IL	20.9	317	19.6	442	18.7	345
IA	21.1	587	19.4	628	18.9	581
MN1/MN2	20.6	353	18.9	292	18.9	236

Table 1-3. Summary of results from aster model comparisons testing the effects of transplant site (site), source population (pop), competition treatment (trt) and their interactions on individual fitness. The effect of block and the interactions were tested against the full model. The effect of the predictors at each life history stage was tested by adding them to the "base" model including the effect of site, population and treatment specified at the fitness level of seedpods. Thus these models are testing the significance of each predictor at each life history stage above and beyond its effect on lifetime fitness.

Term	**Resid df**	**Test df**	**Deviance**	**P-value**
Full	82	-	-	-
Block	37	45	157.6	<0.0001
Trt × Pop	78	4	6.2	0.18
Site × Pop	62	16	31.1	0.013
Site × Trt	74	4	43.5	<0.0001
Main effects	13	-	-	-
Trt	12	1	0.31	0.58
Pop	8	4	3.81	0.43
Site	4	4	61.70	<0.0001
Life history effects above and beyond main effects				
Base	13	-	-	-
Site@esurv	17	4	350.0	<0.0001
Site@fsurv	21	4	325.3	<0.0001
Site@anypods	25	4	52.6	<0.0001
Pop@esurv	29	4	17.8	0.001
Pop@fsurv	33	4	5.5	0.24
Pop@anypods	33	4	7.5	0.11
Trt@esurv	30	1	102.0	<0.0001
Trt@fsurv	31	1	0.81	0.37
Trt@anypods	32	1	9.4	0.002

Table 1-4. Summary of aster model tests for the effect of browsing, disease or herbivory at each site.

Site	Term	Effect	P-value
Interior	Browse	-	< 0.0001
Interior	Disease	+	< 0.0001
Interior	Herbivory	0	0.27
W.Edge	Browse	0	0.99
W.Edge	Disease	0	0.47
W.Edge	Herbivory	+	< 0.0001
N.Edge	Browse	0	0.59
N.Edge	Disease	-	< 0.0001
N.Edge	Herbivory	0	0.18
Beyond.W	Browse	n.t.	-
Beyond.W	Disease	-	< 0.0001
Beyond.W	Herbivory	0	0.44
Beyond.N	Browse	0	0.99
Beyond.N	Disease	0	0.26
Beyond.N	Herbivory	0	0.93

n.t Not tested because no plants were browsed at this site.

Table 1-5. Confidence intervals (95%) for population growth rates at each site. The first column is the 95% CI for population growth rate calculated from the total number of seedlings in 2009 and 2010, accounting for unobserved seedlings in 2010. The second and third columns give the 95% CI for aster estimates of population growth rate. With the aster estimates, the effect of neighbors present (NP) can be compared to neighbors removed (NR), but because unobserved seedlings are not included, these are likely underestimates.

	Observed λ	**Aster λ**	
Site		**NP**	**NR**
Beyond.W	-0.01 – 0.03	-0.03 – 0.07	0
W.Edge	-0.17 – 1.19	0.11 – 0.44	0.23 - 0.70
Interior	0.47 – 1.42	0.52 – 1.29	0.27 - 0.78
N.Edge	-0.03 – 1.40	0	0.10 – 0.41
Beyond.N	0.00 – 0.00	0	0

Figure 1-1. Map of transplant sites (triangles) and seed source populations (circles). The dotted line is the approximate range edge of *C. fasciculata* based on Biota of North American maps (www.bonap.org) and personal observation. Details of sites are given in Tables S1-1 and S1-2.

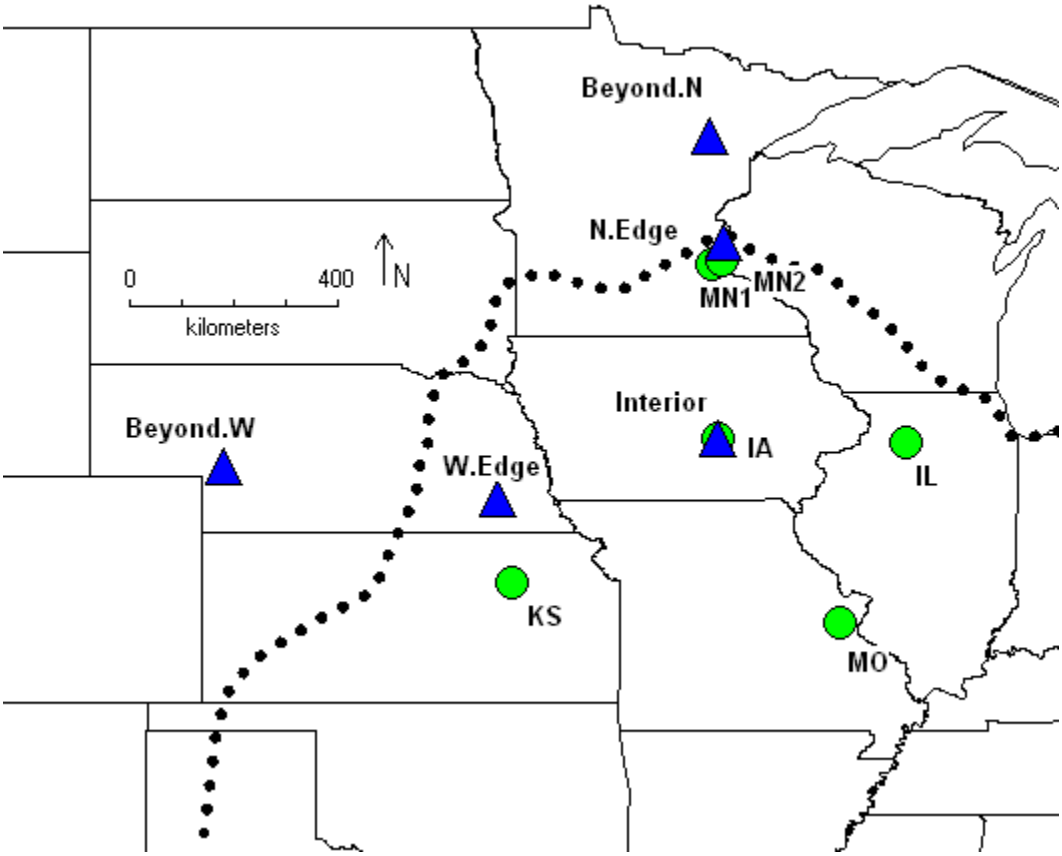


Figure 1-2. Aster life history model for (A) fitness in the first growing season and (B) population growth rate through seedling recruitment. The statistical distribution used for each stage is listed beneath the stage.

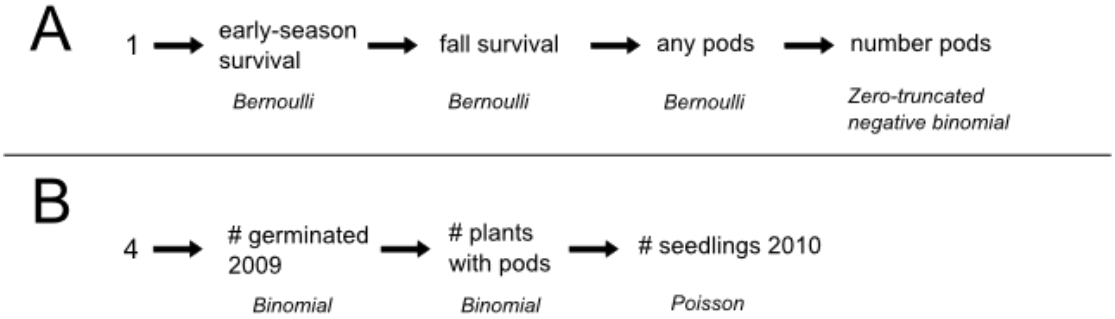


Figure 1-3. Number of seedpods produced per plant across three years in six natural populations at different geographic range locations, arranged from south to north (locations shown in Figure 1-1).

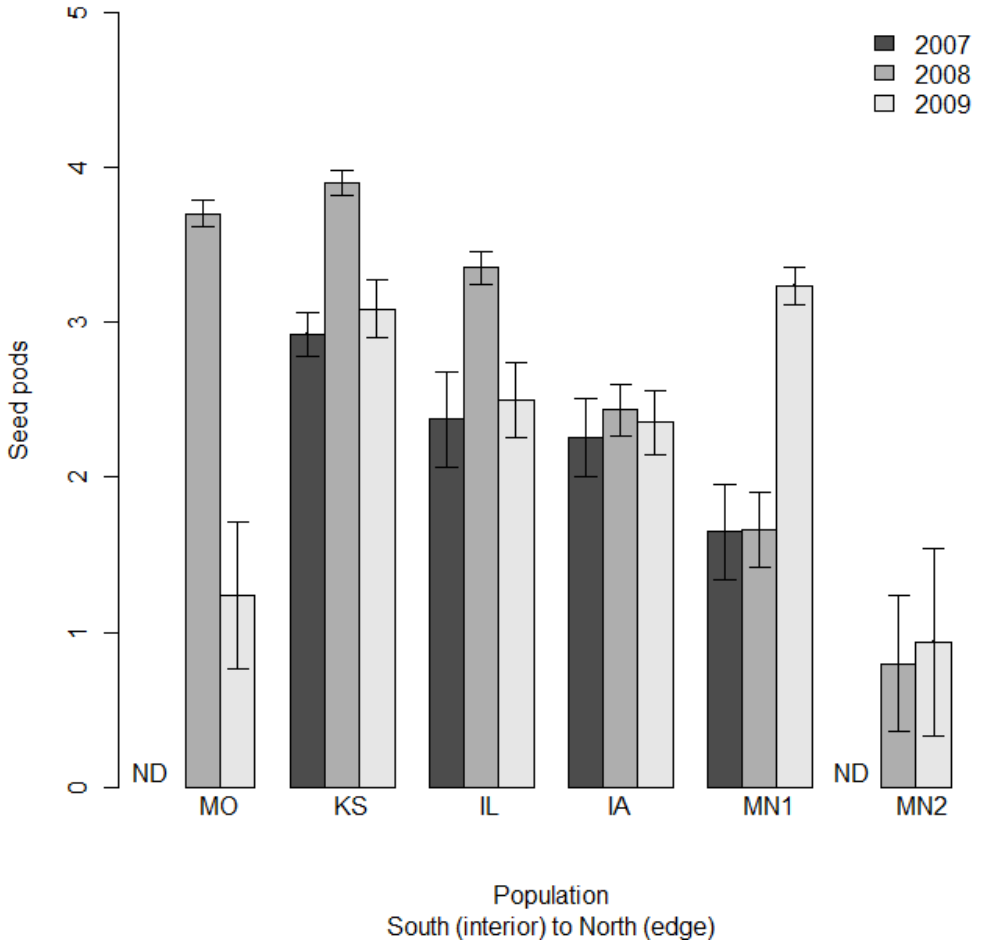


Figure 1-4. Effect of neighbor removal on seedpod production averaged across three years at the same sites as in Fig. 1-3.

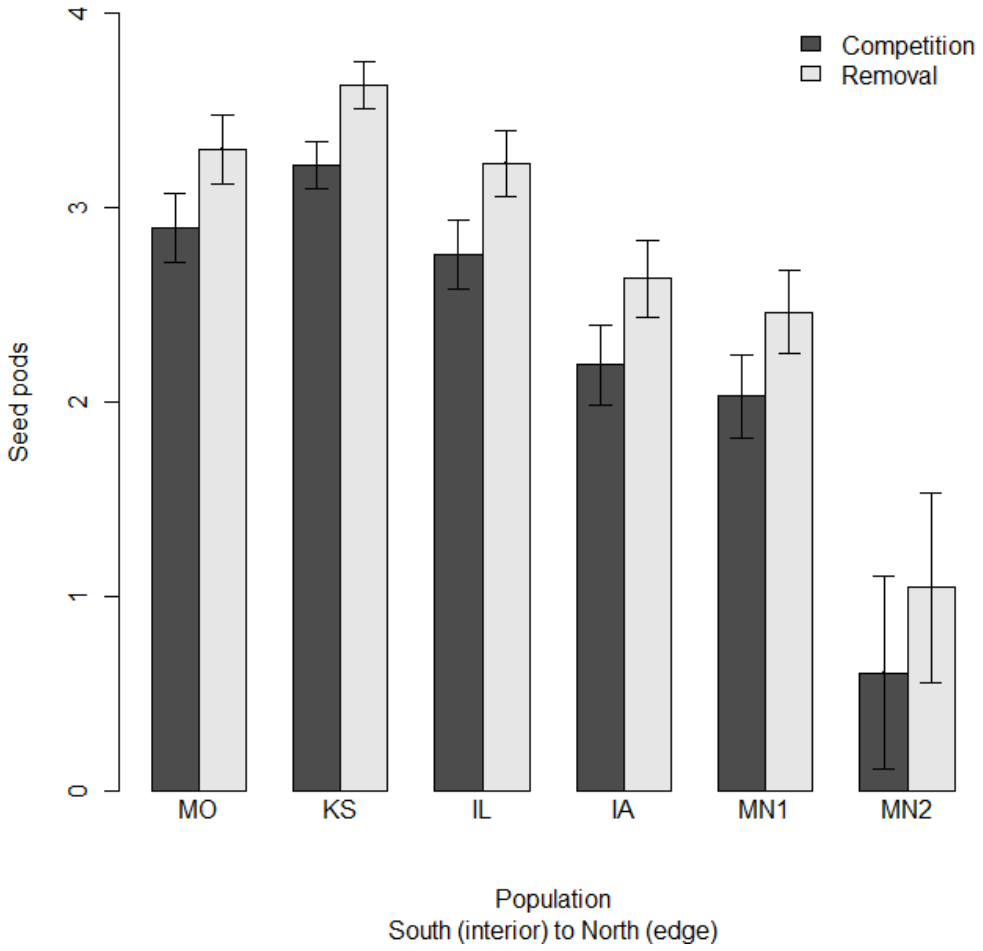


Figure 1-5. Aster model predictions of overall lifetime seedpod production and its underlying components (seedpods produced given survival, whether a plant produced any pods or not, fall survival and early-season survival) for *C. fasciculata* transplanted into sites within the Interior (middle), at the western range edge (W.Edge), at the northern range edge (N.Edge) and beyond these range edges (Beyond.W, Beyond.N).

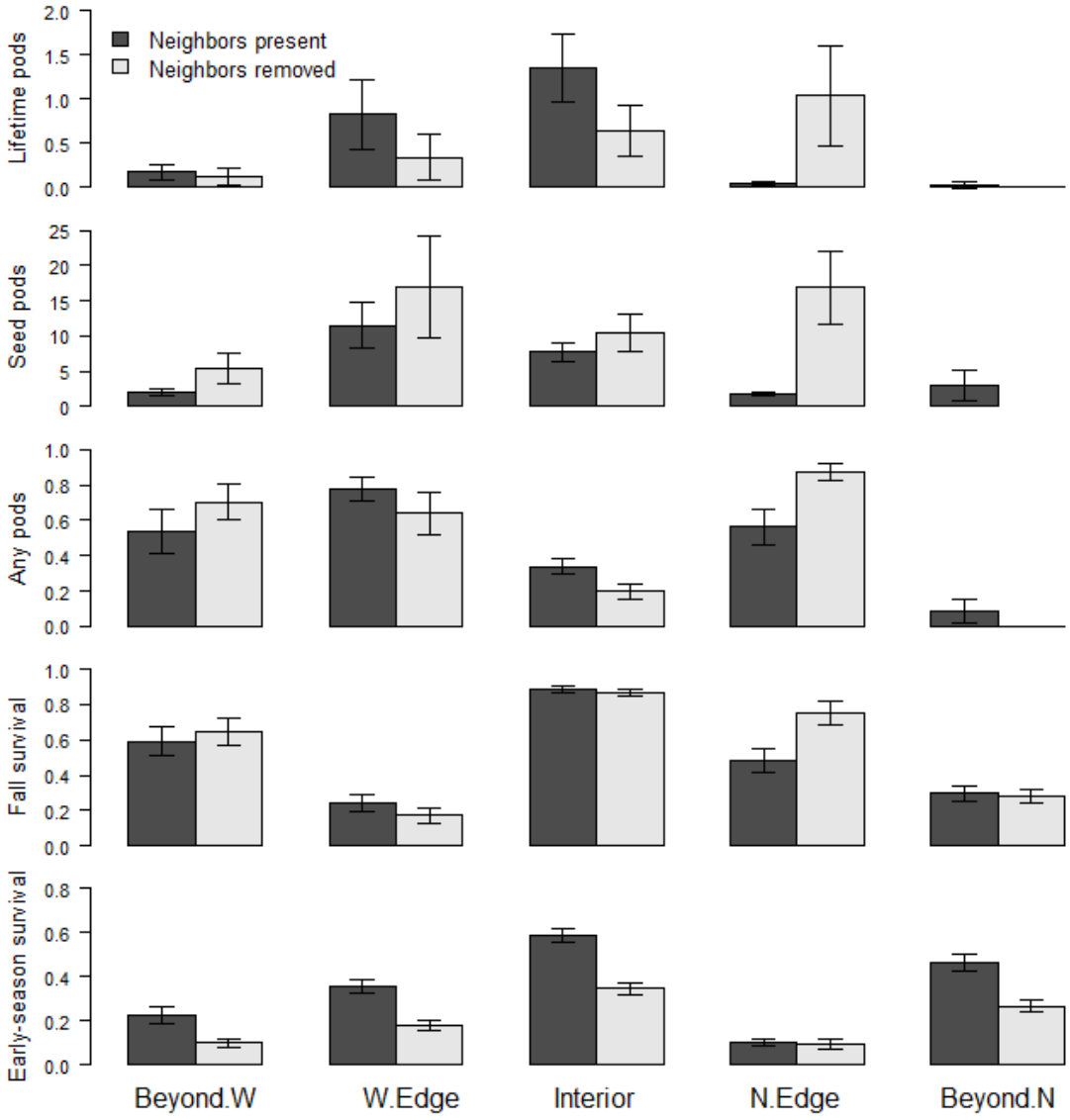


Figure 1-6. Fraction of plants browsed by herbivores at each transplant site in each treatment.

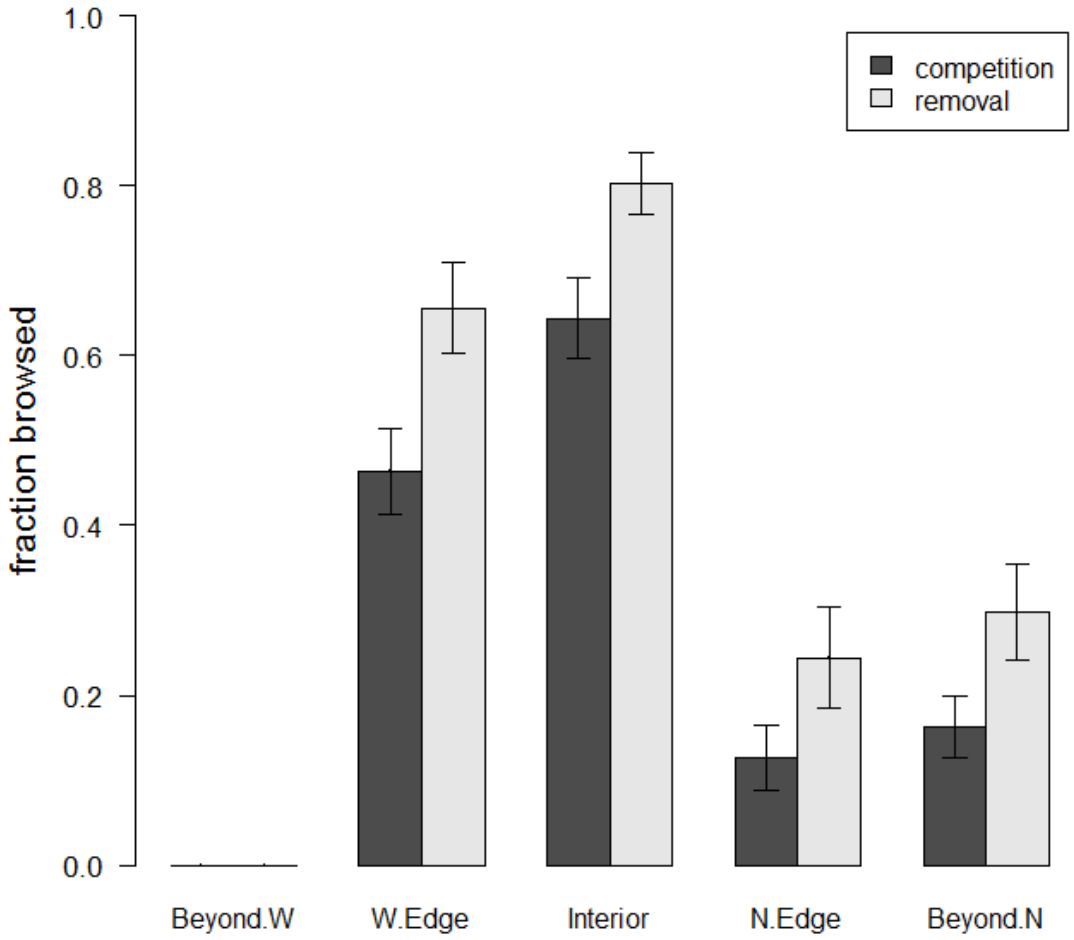
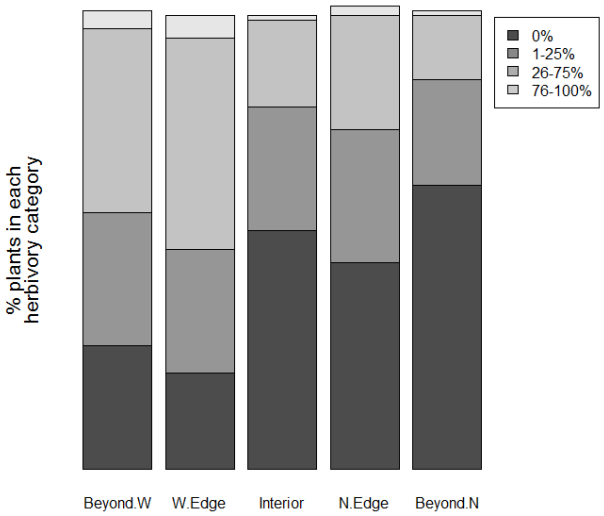
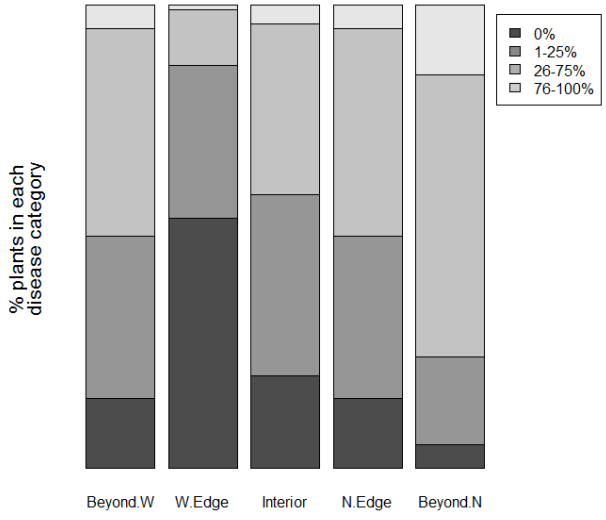


Figure 1-7. Percent of plants at each transplant site in each (A) foliar herbivory category and (B) disease category predicted from proportional odds logistic regression model.

(A)



(B)



Supplemental Information

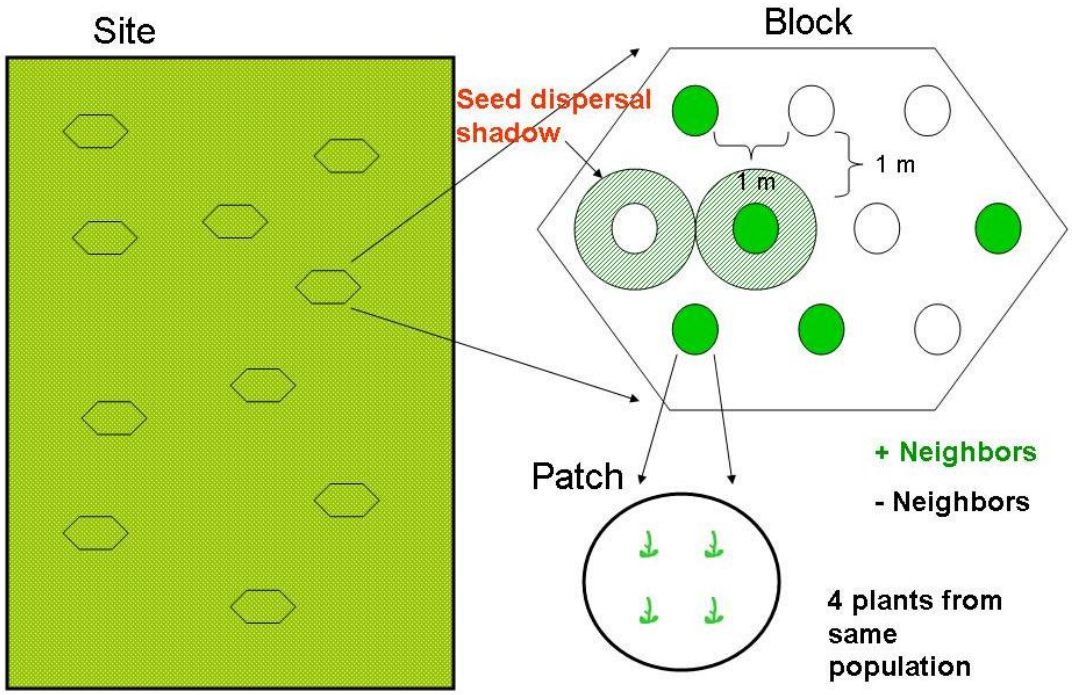
Table S1-1. *C. fasciculata* natural population information, arranged from north to south origin. Mean annual temperature (MAT) and annual precipitation (PPT) were collected from the WorldClim data set implemented in DIVA-GIS (Hijmans *et al*. 2004).

Population	Location	Lat/Long	Region	MAT) (°C)	PPT (mm)
MN2	Afton State Park Afton, MN	44°51’35N 92°46’21W	N. Edge	7.0	778
MN1	Grey Cloud Dunes Scientific and Natural Area, Cottage Grove, MN	44°47’19N 92°57’29W	N. Edge	7.5	754
IA	Conard Environmental Research Area (Grinnell College), Kellogg, IA	41°41’14N 92°52’16W	Interior	8.9	882
IL	Green River Wildlife Area Harmon, IL	41°38’64N 89°31’20W	Interior	9.2	909
KS	Konza Prairie Biological Station, (Kansas State Univ.), Manhattan, KS	39°07’07N 96°32’15W	West edge	11.9	876
MO	Tyson Research Station annex (Washington Univ.), Eureka, MO	38°25’30N 90°40’58W	Interior	12.6	993

Table S1-2. Transplant site locations and climate characteristics. Mean annual temperature (MAT) and annual precipitation (PPT) were collected from the WorldClim data set implemented in DIVA-GIS.

Site	Location	Lat/Long	MAT (°C)	PPT (mm)
INTERIOR	Conard Environmental Research Area (Grinnell College), Kellogg, IA	41°41’03N 92°51’42W	8.9	882
W.EDGE	Reller Natural History Area (Univ. of Nebraska), Sprague, NE	40°36’48N 96°45’41W	10.7	754
N.EDGE	St. Croix Watershed Research Station, Marine-on-St. Croix, MN	45°10’04N 92°45’53W	6.8	774
BEYOND.W	Cedar Point Biological Station (Univ. of Nebraska), Ogallala, NE	41°11’33N 101°38’57W	9.3	458
BEYOND.N	Audubon Center of the Northwoods Sandstone, MN	46°06’41N 92°59’34W	3.7	693

Figure S1-1. Diagram of experimental design at each transplant site. Ten blocks were haphazardly located at each site, depending on space available. Within each block, either 20 (interior and edge sites) or 10 (beyond edge sites) plots were regularly spaced 120 cm from each other in 3 to 5 rows. Rows were staggered so that plots all had a 60 cm diameter surrounding area. Each plot was planted with 4 seeds from the same source population, with seeds placed 10cm from the center of the plot in each cardinal direction, and marked with a toothpick. The corners of each block were marked with posts to facilitate finding the plots, and the center of each plot was marked with a nail. Each plot was randomly assigned to one of two treatments, neighbor removal or unmanipulated neighbors present.



Chapter 2

Does a facultative mutualism limit species range expansion?

A version of this paper was published:

Stanton-Geddes, J., & C.G. Anderson. 2011. Does a facultative mutualism limit species range expansion? Oecologia. DOI 10.1007/s00442-011-1958-4

Summary

The paucity and poor quality of mutualists beyond a species' range edge may limit range expansion. With the legume *Chamaecrista fasciculata*, we asked to what extent the availability and quality of rhizobia beyond the range edge limits host range expansion. We tested the effect of rhizobia availability on plant growth by transplanting seed from three locations into five sites spanning *C. fasciculata*'s range (interior, at the northern and western range edges, and beyond the range edges), and inoculating half the seeds with rhizobia. We recorded growth of all surviving plants, and for the uninoculated plants, whether they had formed nodules or not. We isolated rhizobia from nodules collected on the uninoculated plants, and cross-inoculated seed from four populations (both range edge and interior populations) in the greenhouse to determine whether the quality of rhizobia differed among regions. We found that seeds transplanted beyond the range edge were less likely to be nodulated when they were not experimentally inoculated, and there was benefit to inoculation at all sites. In the greenhouse, the three inocula that formed nodules on plants, from the range interior, northern edge and beyond the northern edge, did not detectably differ in their effect on plant growth. These results suggest that low densities of suitable rhizobia beyond the range edge may limit range expansion of legume species.

Introduction

Many factors including climate (Angert and Schemske 2005, Geber and Eckhart 2005, Rehfeldt et al. 2008, Purves 2009), competitors (Case and Taper 2000, Price and Kirkpatrick 2009), natural enemies (Holt and Barfield 2009) and pathogens (Antonovics 2009) have been implicated in limiting species' geographical distributions. However, the role of a mutualism has received little attention as a potential factor limiting species' distributions. Dependence on a symbiont can constrain range expansion of a host species if the symbiont is rare or absent beyond the range edge (Parker 2001), reducing fitness below that necessary to maintain positive population growth rate. Furthermore, even if the symbiont is present, low abundance or poor quality of the symbiont for the host (Burdon et al. 1999, Heath and Tiffin 2007) may constrain range expansion. As the effect of mutualists on plant performance has been shown to be equal in magnitude to the effect of enemies (Morris et al. 2007), the role of mutualisms in limiting range expansion needs to be further considered. For example, in the case of plant invasion, it has been shown that the absence or low densities of mutualistic rhizobia (Parker et al. 2006) and mycorrhizae (Nuñez et al. 2009, Pringle et al. 2009) may limit invasion success. However, the role that mutualisms may have in limiting native ranges is unclear.

The symbiosis between legumes (Fabaceae) and rhizobia (i.e. nitrogen-fixing soil-dwelling bacteria) is well-suited for examining the role of a facultative mutualism in limiting plant range expansion for both empirical and theoretical reasons. First, legumes experience significant reductions of fitness if compatible rhizobia are not present, demonstrating the dependence of legume population growth on both the availability and identity of symbionts (Bushnell and Sarles 1937, Burdon et al. 1999, Heath 2010). Second, legumes are under-represented in island flora relative to nearby mainland areas (Parker 2001) suggesting that their establishment is constrained by factors other than climate. Third, *Medicago* species that exhibit less specificity for mutualists have larger geographic distributions (Béna et al. 2005), suggesting that the evolution of reduced symbiont specificity allows range expansion. Further, there are theoretical reasons to

expect the availability and quality of rhizobia to differ beyond the range edge. Models show that both partners may reach an equilibrium at a low population size (Parker 2001), which may happen in marginal conditions at the edge of a species range (see examples and counter-examples in Sagarin and Gaines 2002), and thus are more likely to face local extinction. If present beyond the edge of the range of a particular host, rhizobia may be of lower quality (i.e. provide less benefit compared to other strains) to that host plant because selection is not acting to maintain the symbiosis between them. Any rhizobia present either persist saprophytically in the soil, with selection favoring survival in the soil potentially at a trade-off to symbiotic nitrogen fixation (Ratcliff et al. 2008), or the rhizobia persist by forming symbioses with other legume species.

In this study, we examine how the legume – rhizobia mutualism affects the potential for range expansion of the native annual legume *Chamaecrista fasciculata*. Specifically, in a transplant experiment, we ask how rhizobia availability influences fitness at different geographic range locations by inoculating some seeds with rhizobia and leaving other seeds uninoculated. We emphasize that because rhizobia are ubiquitous, we are testing for the presence of strains that have the ability to infect and benefit *C. fasciculata,* and not simply the presence of rhizobia. Finding that uninoculated plants are less likely to nodulate than inoculated plants, and that the inoculated plants have greater fitness than uninoculated plants beyond the range edge, would suggest that the rarity of compatible rhizobia availability does limit legume range expansion. Using rhizobia strains isolated from the field experiment, we conducted a second experiment in the greenhouse asking to what extent rhizobia quality differs among geographic range locations. Finding that rhizobia from the range interior provide a greater benefit to plant growth than rhizobia from the range edge or beyond would suggest that rhizobia quality also may limit plant range expansion.

Materials and methods

Study system

Chamaecrista fasciculata (Fabaceae), an insect-pollinated native annual legume, is one of the most northerly species in the genus *Chamaecrista*, with a distribution from Mexico or Central America in the south to north-central and eastern United States in the north (Irwin and Barneby 1982). Symbiosis with rhizobia is common in the genus *Chamaecrista*, which is one of few genera in the subfamily Caesalpinioideae with species known to nodulate (Doyle and Luckow 2003). In North America, *C. fasciculata* is nodulated by rhizobia in the genus *Bradyrhizobium* (Tlusty et al. 2004, Parker and Kennedy 2006), specifically the lineage *B. elkanii* (Parker and Kennedy 2006). *C. fasciculata* is also known to nodulate with rhizobia isolated from co-occuring species (e.g. *Dalea purpurea*) in the lab, indicating it has the potential to utilize rhizobia from alternate hosts in the soil (Tlusty et al. 2004), though the relatedness of these strains and the prevalence of sharing of rhizobia strains in the soil is unknown. Throughout the text, we use rhizobia to refer to the functional type, as opposed to the generic identity, of the bacteria.

Field experiment assessing rhizobia availability

Plant growth at different geographic range locations may be limited by the availability of compatible rhizobia in the soil. To test this, we examined the growth of *C. fasciculata* plants inoculated with rhizobia (+ rhiz) or left uninoculated (- rhiz) at five sites: the range interior (Interior: Conard Environmental Research Area (Grinnell College), Kellogg, Iowa), western edge (W.Edge: Reller Natural History Area, Spague, Nebraska), beyond western edge (Beyond.W: Cedar Point Biological Station, Ogallala, Nebraska), northern edge (N.Edge: St. Croix Watershed Research Station, Marine-on-St. Croix, Minnesota) and beyond northern edge (Beyond.N: Audubon Center of the Northwoods, Sandstone, Minnesota) (Fig. 2-1). We selected these sites to span two distinct range edges (northern and western) of *C. fasciculata* that may be limited by different combinations of environmental factors (Chapter 1). At the interior and edge

sites, *C. fasciculata* was not growing at the site where the seeds were planted, but was found within1 km. We planted 20 seeds collected in 2008 from each of three source populations (MN, IA, KS; Fig. 2-1) randomly assigned to locations one meter apart at each site between late April and early May 2009. Ten seeds from each source population were randomly assigned to either inoculation with 1 mL inoculum or 1 mL sterile H_2O at the time of planting. The rhizobia inoculum was a mixture of strains UMR6404 and UMR6437, which are known to be beneficial to *C. fasciculata*, from the University of Minnesota Rhizobium collection provided by the late Dr. Peter Graham. Though the identity of these strains has not been confirmed, slow growth suggests that they are *Bradyrhzobium* spp., likely *B. elkanii* given the only previous study of rhizobia taxonomy from *C. fasciculata* (Parker and Kennedy 2006*)*. We used multiple plant populations and a mixture of rhizobia strains to reduce the chance of complete incompatibility between plant genotype and rhizobia inocula (Heath 2010). To inoculate the plants, the rhizobia strains were grown separately in tryptone-yeast (TY) media (Somasegaran and Hoben 1994) until turbid (approx. 10^6 cells/ml, about 5 days), mixed and stored at 4°C until the inoculum was applied to the seeds (not longer than 4 days).

Early-season survival was recorded about four weeks after planting to determine whether the treatments influenced germination and survival. We recorded plant height, which is correlated with fitness ($r^2 = 0.49$ between height and seedpod production, Stanton-Geddes, unpublished data), when plants began to flower in July so that we could also record nodulation before the nodules began to senesce, which happens prior to seed maturation (JSG, personal observation). Sample size at this stage was low because of drought and herbivory (Beyond.N: 11 [- rhiz], 9 [+ rhiz]; Beyond.W: 4 [- rhiz], 5 [+ rhiz], N.Edge: 6 [- rhiz], 7 [+ rhiz]; W.Edge: 10 [- rhiz], 3 [+ rhiz]; Interior: 18 [– rhiz], 15 [+ rhiz]). To record nodulation, we excavated the roots of all surviving uninoculated plants and collected nodules. We recorded only observation of at least one nodule, and not nodule number, because of the difficulty of extracting all roots. Nodules were stored at 4°C with a dessicant (Drierite, W.A. Hammond Company).

While geographic range location may influence the effect of rhizobia on plant growth, abiotic soil nitrogen can also influence plant growth. At high levels of soil nitrogen, *C. fasciculata* does not form symbiotic nodules with rhizobia (Naisbitt and Sprent 1993), and thus, all plants should grow equally well regardless of inoculation treatment. To examine how soil nitrogen influenced nodulation and plant growth, we determined total nitrogen in soil samples collected from each site using a Costech ECS 4010 elemental analyzer (Costech Analytical Technologies, Valencia, CA).

Greenhouse experiment to assess rhizobia quality

To examine whether rhizobia from the edge or beyond a host plant's range differed in quality from rhizobia from within the range, we cross-inoculated four plant populations from different geographic regions with rhizobia strains sampled from within and beyond the range in a greenhouse experiment. *C. fasciculata* seed was collected in September 2008 from four sites; two within the species range (IA and MO), one at the northern edge (MN) , and one at the western edge (KS) (Fig. 2-1). We obtained rhizobia from the nodules collected on uninoculated plants at each site in the field experiment, plus an additional interior site (InteriorMO) so as to have two rhizobia strains for each region (interior, edge, beyond). Nodules were rehydrated in sterile H_2O, surface sterilized with sodium hypochlorite (NaOCl) and 95% ethanol, crushed in 1.5 mL tubes in liquid TY media and the supernatant was spread on TY plates. Where possible (InteriorMO, InteriorIA, N.Edge), we selected three nodules from three different plants, but at other sites (Beyond.N, W.Edge, Beyond.W), most plants did not form many nodules so we used three nodules from one plant. We grew the bacteria at 30ºC for 5-7 days, and then randomly selected one colony from each nodule (three per site), except for the Beyond.W site where no rhizobia grew on the plates. The strains were re-streaked to check for contamination. We grew each strain individually in 50 mL TY media for 7 days at 30ºC, and then we standardized the concentration of each strain at approximately 10^6 cells per mL with OD_{600} readings, likely far greater than the concentration of rhizobia in natural settings. We mixed the three strains isolated from each site in equal proportions to

generate a single inoculum representative of that site and applied the inoculum to the plants within 24 hours. A mixture was used to avoid genotype-specific effects, as we were interested in the effects of the rhizobia population at each geographic location, and not any specific rhizobia strain by plant genotype interactions.

In January 2010, we surface sterilized 60 *C. fasciculata* seeds from each population in 10% sodium hypochlorite. We scarified the seeds, germinated them on filter paper, and planted each emerging seed in a conetainer filled with a 50:50 soil mixture of Turface and SunGro SB-500 to facilitate nodule counting at the end of the experiment. The soil mix was steam-sterilized before planting to reduce potential rhizobia contamination. We used a split-plot experimental design with each of three blocks divided into ten whole plots (conetainer racks) that were randomly assigned a rhizobia treatment (one of five rhizobia inoculum or a control) to reduce the chance of cross-contamination between treatments. Each treatment was represented twice in two blocks and once in the third block because of space constraints on the benches (i.e. blocks). Each plot was divided into eight sub-plots, with two individuals from each plant population randomly assigned. Because of differences in emergence rate, seeds were not all planted on the same day.

Approximately two weeks after planting, we inoculated all *C. fasciculata* plants on the same day with 1.5 mL of the prepared inoculum or sterile H_2O for the control. Plants were watered daily with a fine-nozzle spray hose to reduce splashing between racks. We recorded initial plant height on the day of inoculation, and when plants began flowering (after about eight weeks), we measured height again, harvested all plants and counted all the nodules on each plant. Two inocula (InteriorIA and W.Edge) did not form any nodules on most plants, and it was unclear if this was due to biological reasons or experimental error. Thus, we analyzed the data without these two rhizobia inocula, though the results were similar if they were included (Fig. S2-1, Table S2-1).

Data analysis

To test the influence of site, population, treatment and the site x treatment and pop x treatment interactions on early-season survival, we used a generalized linear model with a binomial link function using glm in R (R Development Core Team 2009). For the plants that survived, we tested the effects of these terms on height by ANOVA (aov in R). For both responses, we fit a full model and then compared nested models using the likelihood ratio test, beginning with the interactions, and removing terms that did not significantly ($P < 0.05$) improve fit of the model to the data. We calculated estimated mean values and standard errors from the final model for plotting.

To assess whether the percentage of uninoculated plants that formed nodules differed beyond the range than within the range, we grouped data from the two sites beyond the range, and the three sites within the range or at the edge. We compared frequencies of nodulated plants between these two groups (beyond vs interior) using Fisher's exact test because of small sample sizes. To determine if soil nitrogen influenced nodulation, we tested if there was a significant correlation between site nitrogen and the percent increase in growth between uninoculated and inoculated plants. Finding a significant negative correlation would suggest that at sites with high soil nitrogen, plants benefit less from the rhizobia, potentially not entering in symbiosis with them.

To assess differences in rhizobia quality on plant growth in the greenhouse, we examined the effects of rhizobia treatment, population, rhizobia x population and the covariates of block and date planted on the difference in plant growth from the date of inoculation to the date of harvest. One plant died and was removed from the analysis. We performed a split-plot ANOVA with the effect of rhizobia treatment and rhizobia x plant population tested against the variance among plots, here the rhizobia x block interaction while the main effects of population and date planted were tested with the residual variance as the appropriate error term (Milliken and Johnson 2009). We repeated these analyses for nodule number, with a ¼ power transformation of nodule number which came closest to the assumption of normality for the residuals.

Results

Rhizobia availability in field

Inoculation with rhizobia had no effect on early-season survival, but significantly increased mid-season plant height (Table 2-1) with inoculated plants being taller at all sites (Fig. 2-2). The site by treatment interaction was not significant for plant height (Table 2-1), as the absolute increase in height was similar across sites. Site had a significant effect on early-season survival (Table 2-1) with high survival at the interior (63%) and beyond northern edge site (61%), intermediate survival at the western edge (42%) and low survival at the northern edge (15%) and beyond western edge site (15%). Plant height in July was also significantly influenced by site (Table 2-1), with plants being tallest in the interior site and shortest in the beyond range sites (Fig. 2-2). Plant population source had a significant effect on early-season survival but not plant height (Table 2-1). The interaction between population and treatment was not significant for either early-season survival or height (Table 2-1).

The percentage of uninoculated plants that had nodules differed among sites beyond the range and sites within the range (Fisher's exact test, $P = 0.002$; result consistent using GLM with binomial link, $dev_{1,44} = 10.1$, $P = 0.0001$). Specifically, 13% of uninoculated plants beyond the range were nodulated, compared to 64% within the range, though this result should be regarded as tentative because of the small sample size due to low survival. Soil nitrogen varied from 0.13% to 0.25% and there was no significant correlation between soil nitrogen and the effect of inoculation ($r = -0.67$, $P = 0.2$).

Greenhouse experiment to assess rhizobia quality

Plant height was significantly influenced by rhizobia treatment (Table 2-2). Rhizobia mixtures from the inocula that formed nodules (one interior site, the northern range edge site and the beyond northern range edge site) all significantly increased plant growth relative to the control treatment, though they did not differ from each other (Fig. 2-3). Plants inoculated with the two rhizobia mixtures that did not form nodules did not

grow significantly taller than the control plants (Fig. S2-1). Plant population also influenced plant height, but there was no significant interaction between plant population and rhizobia treatment (Table 2-2).

The number of nodules per plant differed significantly with rhizobia treatment but not by plant population or the interaction between the two (Table 2-2). Control plants had zero nodules in 28 of 40 cases and never had more than two nodules, allowing us to compare ecologically-relevant differences in infection and benefit, though limiting our ability to detect potential low levels of infection. Of the three inocula that formed nodules, the interior inoculum produced the most nodules (~98 nodules/plant), the inoculum from beyond the northern range edge next (~57 nodules/plant) and the northern edge inoculum produced the fewest nodules (~37 nodules/plant) (Fig. S2-2).

Discussion

A mutualism may limit species range expansion if either the availability or quality of mutualistic partners is decreased beyond the range edge. In the *C. fasciculata* – rhizobia facultative mutualism, we found uninoculated seeds were less likely to form nodules when transplanted beyond the range edge than within the range or at the edge, and that there was a benefit to plant growth from inoculation, indicating that rhizobia availability does limit range expansion. We emphasize that we observed this pattern beyond two distinct range edges, reinforcing the generality of this conclusion. However, we did not find evidence that rhizobia from beyond the range edge were of lower quality in their effect on growth of the legume.

Theoretical work has shown that low densities of compatible rhizobia, and not just the absence thereof, can limit the potential for plants to invade (Parker 2001). Our results support this prediction, as we found a reduction of nodulation beyond the range edge of *C. fasciculata*. While rhizobia availability was limited, some plants did nodulate and we were able to isolate rhizobia from the nodules collected at the northern site.

The potential to form a symbiosis, and the benefit that each partner gains in symbiosis, depends on both plant and rhizobia genotype (Spoerke et al. 1996, Burdon et al. 1999, Heath 2010). We expected to find differences in rhizobia quality among strains isolated from within *C. fasciculata*'s range compared to strains from beyond the range because selection between symbionts is no longer maintaining the mutualism. Specifically, given the trade-off between allocation of resources to carbon storage and symbiotic nitrogen fixation that has been shown for rhizobia (Ratcliff et al. 2008), we expected rhizobia from beyond the range edge to be of lower quality to the host. However, we did not detect any differences in the effect on plant growth of rhizobia inocula isolated from sites within *C. fasciculata*'s range, at the northern range edge and beyond the northern range edge (Fig. 2-3). One potential explanation is that by using a mixture of strains isolated from each site, we averaged across genotypes that may differ in quality. Alternatively, these could be strains of rhizobia that are generalists and are maintained in the soil at these sites by alternate hosts, at no cost to symbiosis with *C. fasciculata*. Finally, these strains may be broadly distributed (e.g. dispersal by wind) and are the same as those within *C. fasciculata*'s range, though at lower densities where *C. fasciculata* is not present and thus has not had the opportunity to augment populations of compatible rhizobia via the symbiosis. Regardless of any differences among individual strains, this result suggests that at least at one site beyond *C. fasciculata*'s range, the average effect of rhizobia (even if not abundant) on host fitness is not strikingly different from that of rhizobia within the species' range.

In conclusion, we find evidence that the availability of rhizobia in the soil beyond the range edge of the legume host limits plant growth, and thus fitness, but that rhizobia from beyond the range edge are not different in quality to the plant from rhizobia within the range. We emphasize that rhizobia alone do not prevent range expansion, as both survival and plant growth are reduced beyond the range edge even with added rhizobia (Fig. 2-2; Chapter 1), but that low densities of rhizobia beyond the range edge may interact with other environmental factors to constrain range expansion (e.g. Case et al. 2005). This result suggests that the rate at which legume species can shift their

distributions to track climate change (Davis and Shaw 2001, Ackerly 2003) may be limited not only by the rate at which they can disperse, but the rate at which they encounter compatible rhizobia in the soil. More generally, it implies that the role of mutualisms should be further considered in studies of the factors that limit species' ranges.

Funding and Assistance

We thank Larissa Mottl at Conard Environmental Research Area, Jean Knops at University of Nebraska, Shawn Schottler at St. Croix Watershed Research Area, Bryan Wood at the Audubon Center of the Northwoods for the use of field sites, and Konza Prairie, Tyson Research Station, Conard Environmental Research Area and the Minnesota Department of Natural Resources for seed collection, and two anonymous reviewers for helpful comments. This work was funded by a University of Minnesota Undergraduate Research Opportunity Fellowship to C.A., the Center of Community Genetics Fellowship to J.SG. and National Science Foundation DEB 0515466 to Peter Tiffin. All work performed in this study complies with laws of the United States of America.

Data accessibility: DRYAD entry: doi:10.5061/dryad.8566

Tables and Figures

Table 2-1 Summary of results from likelihood ratio tests of the effects of site, plant population, rhizobia treatment (+/- rhizobia) and interactions on early-season survival (GLM with binomial link) and mid-season height (ANOVA) in the field. Residual df differ for height and early-season survival because fewer plants survived to mid-season.

	Early-season survival			**Height**			
Source	**df**	**Dev**	**P > χ^2**	**df**	**SS**	**F value**	**P > F**
Site	4, 293	59.2	< 0.0001	4, 82	16987	44.69	< 0.0001
Plant pop	2, 293	6.8	0.03	2, 80	285.2	1.52	0.23
Rhizobia	1, 292	0.86	0.35	1, 80	3744.7	39.91	< 0.0001
Site x Rhiz	4, 286	1.44	0.84	4, 74	589.0	1.58	0.19
Pop x Rhiz	2, 290	0.16	0.92	2, 78	20.91	0.11	0.90

Table 2-2 Summary of results of split-plot ANOVA testing for effects of rhizobia treatment and plant population on change in plant height and nodule number. The effects of block and rhizobia treatment were tested against the variance among plots (rhizobia x block interaction), while the remaining terms were tested against the overall variance. The error terms for each level of the split-plot design are designated in italics below their respective terms**.**

		Change in plant height				**Nodule number**		
Source	**df**	**SS**	**F value**	**P > F**	**df**	**SS**	**F value**	**P > F**
Block	2	138	4.36	0.07	2	15.9	12.60	0.007
Rhizobia	3	3083	65.1	< 0.0001	3	169.0	89.27	0.0001
Rhiz x Block	*6*	*95*	-	-	*6*	*3.8*	-	-
Date planted	11	903	2.64	0.005	-	-	-	-
Plant x Rhiz	9	256	0.91	0.52	9	0.26	0.24	0.99
Plant pop	3	464	4.97	0.003	3	0.71	1.99	0.12
Residuals	*124*	*4552*	-	-	*135*	*16.1*	-	-

Figure 2-1. Map of source locations for *C. fasciculata* seed (yellow circles) and transplant sites (blue triangles) where the field experiment was performed and rhizobia were trapped. Rhizobia were also collected from plants at the MO site for the greenhouse experiment. The black line is the approximate range edge of *C. fasciculata* based on USDA Plants Database (plants.usda.gov) county level information. Plant populations are: MN, Grey Cloud Dunes Scientific and Natural Area (44°47'19N 92°57'29W), Minnesota Department of Natural Resources, Cottage Grove, Minnesota; IA, Conard Environmental Research Area of Grinnell College (41°41'14N, 92°52'16W), Kellogg, Iowa; MO, Tyson Research Station annex of Washington University (38°25'31N 90°40'58W), St. Louis, Missouri; KS, Konza Prairie Biological Station at Kansas State University (39°07'07N, 96°32'15W), Manhattan, Kansas. Transplant sites are: Beyond.N: Audubon Center of the Northwoods (46°06'41N 92°59'34W), Sandstone, Minnesota; N.Edge: St. Croix Watershed Research Station (45°10'42N 92°45'53W, Marine – on – St. Croix, Minnesota; Interior (same as IA above); W.Edge, Reller Natural History Area of University of Nebraska – Lincoln (40°35'48N 96°45'41W), Sprague, Nebraska; Beyond.W, Cedar Point Biological Station of UN – Lincoln (41°11'34N 101°38'57W), Ogallala, Nebraska. The map was made using DIVA-GIS (www.diva-gis.org).

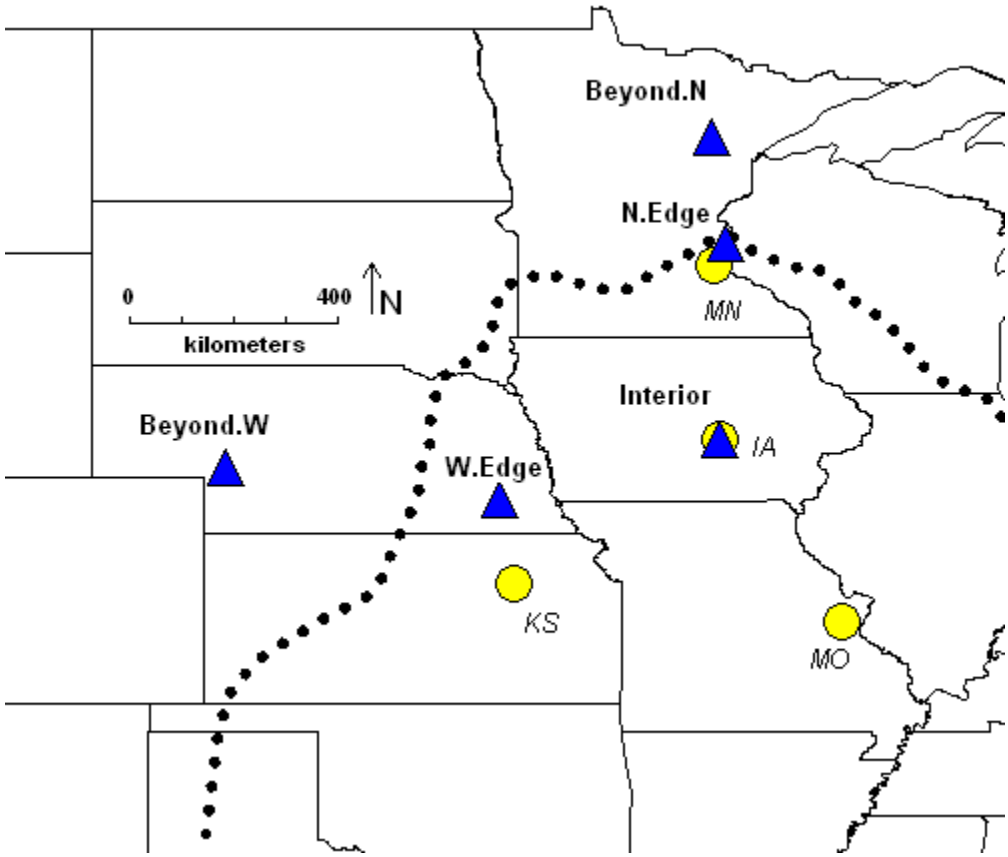

Beyond.N
N.Edge
0
400
N
kilometers
MN
Interior
Beyond.W
IA
W.Edge
KS
MO

Figure 2-2. Estimated mean (± SE, N = 88) of mid-season height for uninoculated (dark bars) and rhizobia inoculated (light bars) plants grown in the field at five sites: the interior of *Chamaecrista fasciculata*'s range, at the western (W) and northern (N) range edges, and beyond the western and northern range edges. Specific locations are given in Supplemental Figure S3-1.

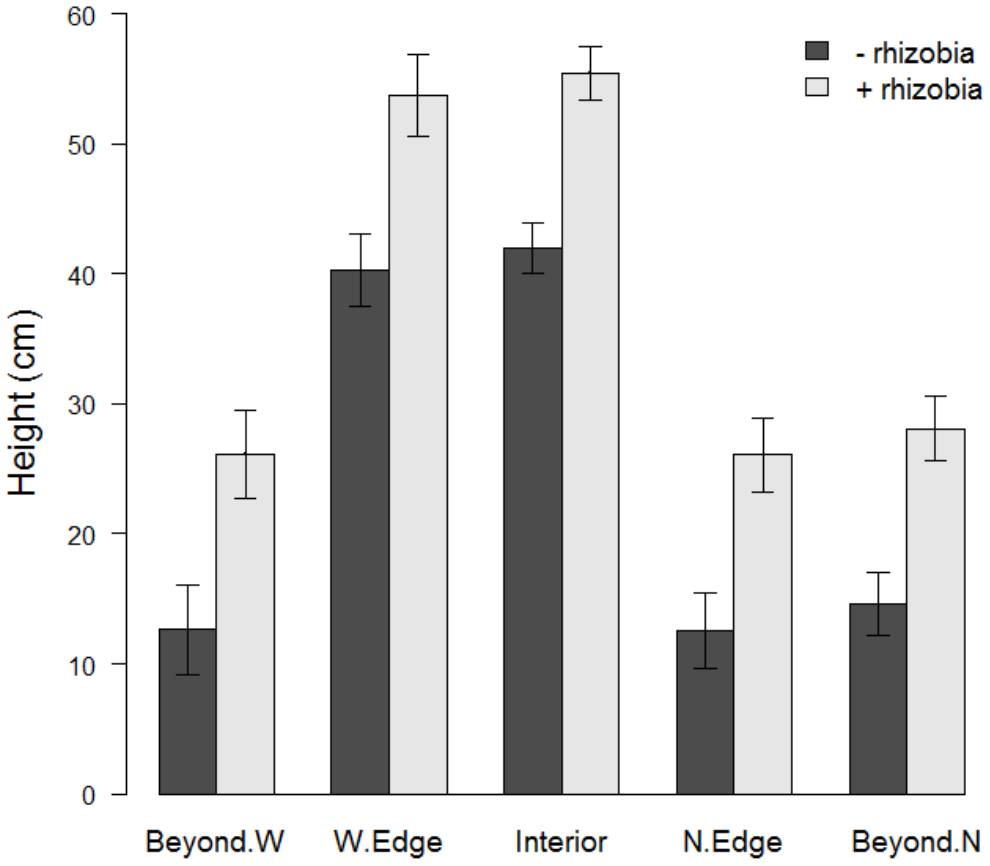
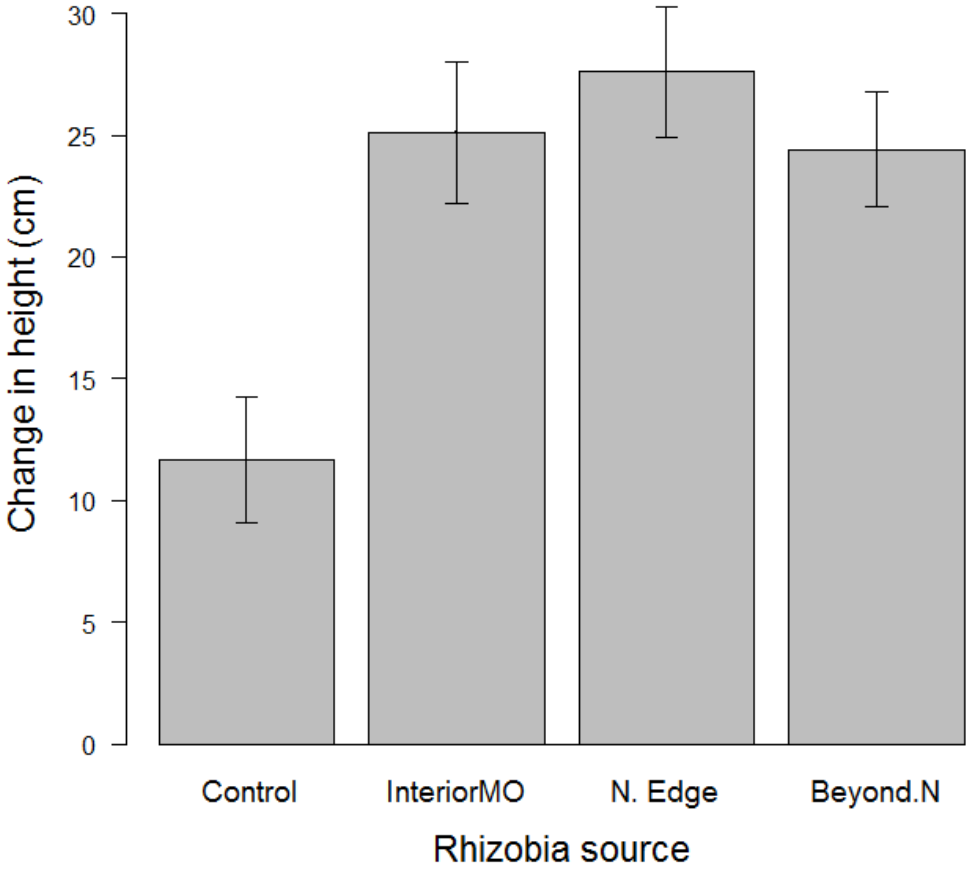


Figure 2-3. Estimated mean (± SE, N = 239) of change in height since inoculation for plants grown in the greenhouse with or without one of three rhizobia inocula (interior, northern range edge or beyond northern range edge)



Supplemental Information

Table S2-1. This table is the same as Table 2-2, but includes all 5 rhizobia inocula and the control in the "Rhizobia" term. Two of these inocula (rINTERIOR, rRNHA) were not included in the analysis presented in the paper because they did not produce any more nodules than control.

		Change in plant height				**Nodule number**		
Source	**df**	**SS**	**F value**	**P > F**	**df**	**SS**	**F value**	**P > F**
Block	2	360.3	3.5	0.07	2	15.2	8.4	0.007
Rhizobia	5	5890.5	22.7	< 0.0001	5	316.8	70.3	< 0.0001
Rhiz x Block	*10*	*518.9*	-	-	*9*	*9.0*	-	-
Date planted	11	989.4	2.3	0.01	-	-	-	-
Plant x Rhiz	15	479.0	0.8	0.68	15	1.63	0.6	0.89
Plant pop	3	492.6	4.1	0.007	3	0.68	1.19	0.32
Residuals	*192*	*8427.6*	-	-	*203*	*38.85*	-	-

Figure S2-1 Estimated mean values (± SE) of change in height for plants grown in the greenhouse with one of five rhizobia inocula or control, including the two inocula (Interior2 and W.Edge) not included in Figure 2 in the text as they did not produce significantly more nodules than control. Rhizobia inoculum source location is given in Figure S2.

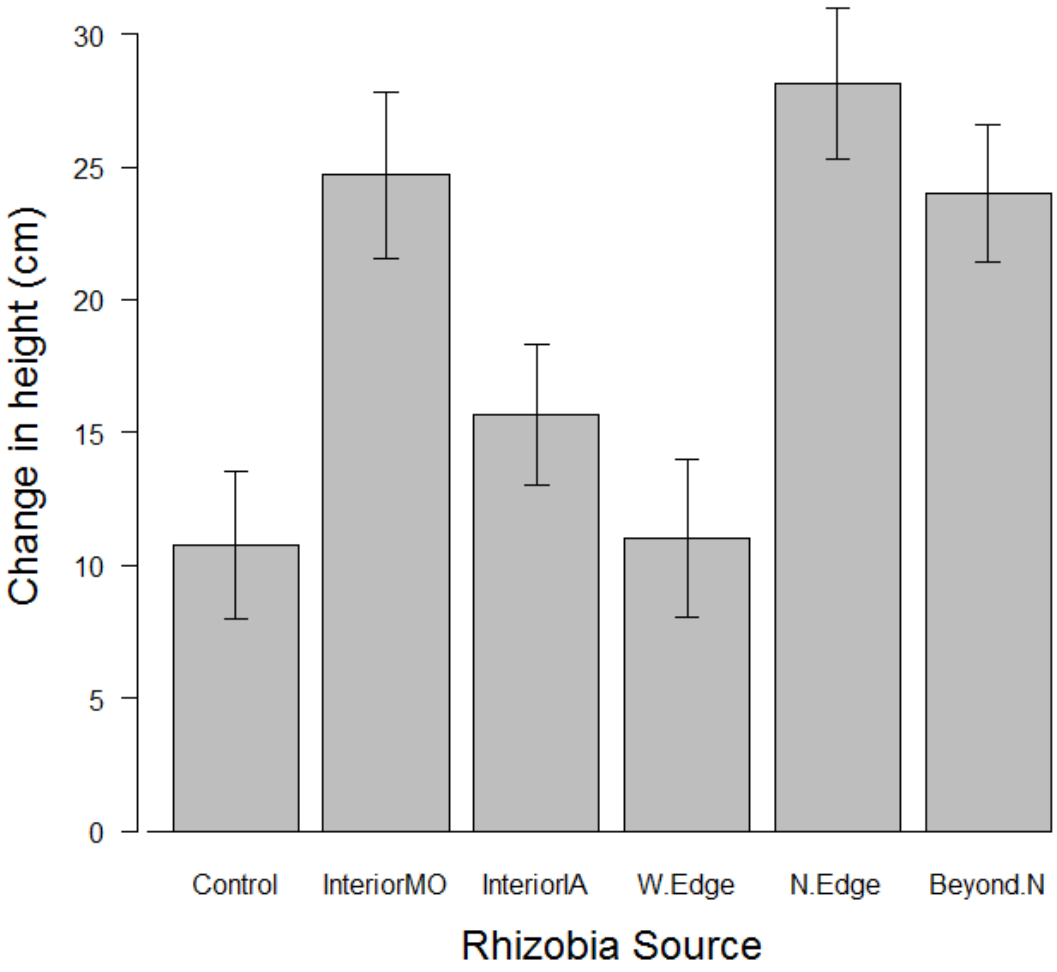
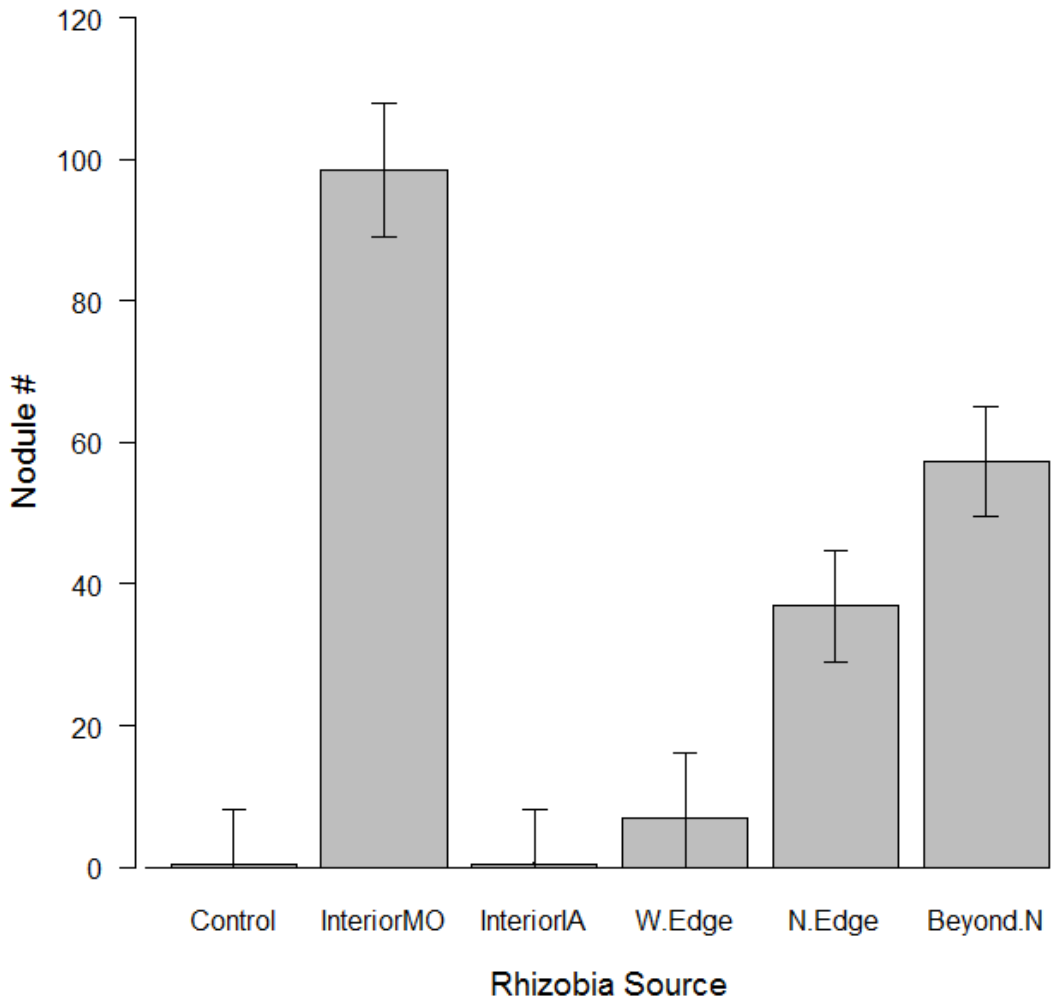


Figure S2-2. Estimated mean values (± SE) of nodule number on plants grown in the greenhouse with five rhizobia inocula and a control. Rhizobia inoculum source location is given in Figure S2.



Chapter 3

The influence of soil type on lifetime fitness with changes in geographic range location

Summary

Range shifts are a primary expected response of species to climate change. However, we know little about how species will change the habitats they occupy in the course of range shifts. We planted seed from five populations of the native plant *Chamaecrista fasciculata* within its range, at its northern range edge and beyond this range edge to examine whether this species will shift habitats (e.g. soil type) if its range shifts northward. Further, we investigated how phenotypic selection differs among regions and habitat types. We found that with reduced competition, lifetime fitness was always greater or equivalent at one habitat type, loam soils, though early-season survival was greater on sandy soils. At the range edge, natural populations are typically found on sandy soils, which are also less competitive environments. In light of this, our results suggest that the habitats species occupy may change because of the combined effect of differences in establishment and biotic effects (e.g. competitors) among soil types. Selection varied with range location, with traits associated with greater resource acquisition favored beyond the range edge, and resource conservation favored within the range.

Introduction

Plant species are predicted to respond to climate change by range shifts and adaptation (Huntley 1991, Davis and Shaw 2001, Ackerly 2003). However, little consideration has been given to how the habitats species occupy may change beyond their current range edges. For a number of reasons, the habitat occupied by a species in its contemporary range may not be representative of the sites where dispersing individuals and propagules (hereafter, colonists) will successfully establish. First, changing climatic factors, such as mean annual temperature and precipitation, may become unlinked from other factors that directly influence individual fitness (Jackson et al. 2009), altering the fundamental niche of a species. Second, other abiotic factors that influence individual fitness such as photoperiod (Griffith and Watson 2006), insolation (Warren 2010) and soil type (Macnair 1987, Sambatti and Rice 2006) will not change on the same temporal scale as climate and could influence the habitats suitable for colonists. Finally, biotic factors such as competitors, disease and mutualists vary spatially (Thompson 2005), are likely to be affected differentially by climate change (Both et al. 2006), and thus may influence which sites colonists can occupy beyond the current range.

Soil characteristics are spatially variable within climate regions, and because they vary in nutrient levels, water diffusion, metal concentrations (Macnair, 1987), soil biota (Bissett et al. 2010) and support different biotic communities (Fine et al. 2006, Henne et al. 2007) they can have strong effects on plant growth and fitness (Antonovics and Bradshaw 1970, Sambatti and Rice 2006). Further, the environment a plant experiences is

dependent on interactions between soil type and climate. For example, at *Clarkia xantiana*'s eastern range edge in southern California, water availability is much lower than expected based on precipitation because of a change in soil type (Eckhart et al. 2010). As species shift their ranges in response to climate change, it has been suggested that they will track (passively, via propagule dispersal, in the case of sessile species) the conditions to which they have the narrowest environmental tolerance and adapt to novel conditions in other niche dimensions (Ackerly 2003). This may occur by establishing in habitats with different properties in novel regions beyond the range edge. Thus, the soil type on which individuals establish and reproduce may differ within and beyond the current range.

Where colonists establish may also depend on their origin. Populations are often locally adapted (Leimu and Fischer 2008, Hereford 2009) to factors including climate (Etterson 2004, Angert and Schemske 2005, Geber and Eckhart 2005, Bischoff et al. 2006), competitors (Ehlers and Thompson 2004, Bischoff et al. 2006), natural enemies (Crémieux et al. 2008) and soil (Macel et al. 2007). If populations at the range edge have already adapted to local conditions, individuals from these populations may be most likely to establish new populations beyond the current distribution. However, peripheral populations are often small (Gaston 2003) and thus adaptation to local conditions may be constrained by drift or gene flow from interior populations (Antonovics 1976, but see Holt and Gomulkiewicz 1997, Lopez et al. 2009). Interior populations are often larger and more stable (Gaston 2003) and thus likely retain greater genetic diversity. In this

case, interior populations may retain greater fitness and potential for further adaptation than edge populations when dispersing beyond the range edge.

The evolutionary potential of colonists to adapt to conditions beyond the range edge is also likely to be influenced by the ecological context of the habitat where colonists arrive (Holt 2003). Species' range limits have been explained by selection for different optimum trait values at different geographic locations, combined with gene flow among populations (Kirkpatrick and Barton 1997). However, empirical evidence for this is limited (Jenkins and Hoffman 1999, Angert et al. 2008). Moreover, these studies do not examine how different abiotic habitats influence fitness and optimum trait values with changes in range location. By evaluating selection on trait distributions of populations transplanted into common conditions in nature, we can obtain ecologically and evolutionarily meaningful estimates of selection. Specifically, do range edge populations have trait values that are favored by selection at and beyond the range edge relative to interior populations?

In this study, we use the native annual legume *Chamaecrista fasciculata* to investigate; (1) the extent to which plant fitness is influenced by habitat at different geographic range locations, (2) whether range edge populations maintain higher fitness than interior populations when transplanted beyond the range edge, and (3) how selection on ecologically-relevant traits differs among habitats and range locations. We accomplished these objectives by transplanting individuals from five populations into two habitat types, loam and sandy soils, within the species' range, at its current range edge, and beyond the range edge. Although *C. fasciculata* is found on soils ranging from sandy

to clay within its range (Foote and Jackobs 1966; JSG, personal observation), it is primarily found on sandy soils at its range edge (Irwin and Barneby 1982, Ownbey and Morley 1993), suggesting that with range expansion, *C. fasciculata* may only be able to establish and persist on a subset of the habitats on which it occurs within its range. However, the reason for this habitat shift is unknown. We note that while attempted to reduce the effects of competitors, the competitive environment also differed among habitats and regions. Thus, we interpret our results in the context of the habitat, including both abiotic and biotic properties, which colonists may establish in and not the soil type itself.

Materials and methods

Chamaecrista fasciculata (Fabaceae) is an annual legume native to North America that is widely distributed from central Minnesota to Massachusetts and south into Mexico (Irwin and Barneby 1982). It is found in prairie remnants, old fields, open woodlands, and disturbed areas on a wide variety of soil types from sandy to waterlogged clay soils (Foote and Jackobs 1966, Irwin and Barneby 1982). Within the range in Illinois, *C. fasciculata* occurs on all these soil types but most often on silty clay loam soils (Foote and Jackobs 1966). However, at both the northern (Ownbey and Morley 1993; JSG, personal observation) and the western range edge (Irwin and Barneby 1982) it appears restricted to open habitats with sandy soils, though loam soils adjacent to all these sites are not frequently occupied by *C. fasciculata*.

In September and October 2007, we collected seed from five populations at different locations, from the northern range edge in Minnesota south to Kansas and Missouri (Fig. 3-1, Table S3-1). In May and June 2008, we established common gardens at three geographic regions chosen to represent different range locations: interior (central Iowa), edge (south-eastern Minnesota within ~50km of the furthest north known naturally occurring populations) and beyond the range (central Minnesota, approximately 120 km beyond the furthest north population recorded in the area) (Fig. 3-1). The locations were chosen by the availability of nearby sites with the desired soil properties in each region where we could establish common gardens. Principal components analysis of climatic variables demonstrated climatic similarity of sites within regions (Fig. S3-1). Within each region, we chose two sites differing in soil types, sand and loam, characteristic of the habitats *C. fasciculata* occupies at the range edge and interior, respectively (Table S3-2). Transplant sites are referred to by region and soil type (e.g. "interior – sand") throughout the text.

At each of the transplant sites, we planted 100 seeds from each of three source populations (CRA, GCD, KZA) and because of limited seed availability, 52 and 24 seeds from the AFT and CUI populations, except at the beyond-loam site where no CUI seeds were planted and edge-loam where no seeds from either AFT or CUI were planted. Seeds were planted in late May and early June, starting at the southern sites and moving north. Prior to planting, seeds were sterilized with 10% sodium hypochlorite (NaOCl) and scarified with a metal file. Each site was sprayed with glyphosate (Roundup, Monsanto, St. Louis, MO) at least 24 hours prior to planting, and above-ground vegetation was

removed by mowing and raking to lessen differences in competition among sites. All sites except the beyond-loam site, due to restrictions, were fenced to exclude deer. Seeds were planted in a randomized block design with four blocks per site. In each block, seeds were planted 1 cm deep, 20 cm apart in staggered rows 17.5cm apart, such that each plant had three neighbors 20 cm away. Additional seeds were planted at the end of each row as edge plants. Pairs of rows were 50 cm apart to allow walking and mowing between plants. Rows were mowed to reduce competitors; three times at edge and beyond edge sites, and once at interior sites due to logistical constraints. Because of the limited mowing at the sites within the range, aboveground vegetation was taller and denser at the loam site than the sand site. There were minimal differences in vegetation among sites in the edge and beyond edge regions due to less competitive vegetation and more frequent mowing (JSG, personal observation).

We returned to each site 2-4 weeks after planting to record early-season survival, which includes differences in germination as well as survival of early-season conditions. Approximately 8 weeks after planting, we recorded survival, flowering status and plant height and collected the top-most fully expanded leaf to measure specific leaf area (SLA, leaf area (m^2) / dry mass (g)), which was calculated with five leaflets per leaf, removing the petiole (Cornelissen et al. 2003). In late September and early October when plants had begun to senesce, we recorded survival, branch number, the number of seedpods produced by each plant, and collected either 10 or 10% (whichever was larger) of the seedpods on each surviving plant except at the interior-loam site where all plants were eaten or trampled by deer late in the season after the fencing was damaged. In cases

where plants had fully senesced, it was possible to count pods that had already dehisced because the stiff pedicels with pod fragments remain attached to the plant. Pods were stored in coin envelopes at room temperature, and the average number of seeds per collected pod was recorded. Inviable or aborted seeds, judged by size and color, were not counted.

Statistical Analyses

To determine the effects of geographic region and habitat type on plant fitness, and how the response of populations to each transplant environment differed according to source location, we modeled individual lifetime fitness at each site, taking into account multiple life history stages. We then estimated the fitness landscape at each site to investigate differences in selection among regions and habitats.

We modeled individual lifetime fitness using aster models (Geyer et al. 2007, Shaw et al. 2008) implemented in R (R Development Core Team 2009), a likelihood approach that allows modeling of multiple components of life history (Fig. 3-2) in a single analysis, with an individual's response at each stage conditioned upon its response at the previous stage. Aster models are an improvement over previous attempts to model lifetime fitness because an appropriate distribution is specified for each life history stage, and the dependence of later life history stages on previous stages is explicitly modeled (Geyer et al. 2007, Shaw et al. 2008). The details of fitting aster models are given by Shaw *et al*. (2007b, 2008). Because seeds were counted in a subsample consisting of a random number of pods, a stage for pods sampled was included between the stages for

number of pods produced and the number of seeds counted in the sampled pods (Fig. 3-2, node 5; see Supporting Information Technical Report for details). As the current aster package automatically accommodates only single-parameter exponential family distributions, the size parameters for the negative binomial distributions were chosen by fitting a negative binomial distribution (fitdistr function in library MASS (Venables and Ripley 2002) in R) to the conditional distribution of the data. Goodness of fit for the conditional distributions of seedpods and seeds counted was assessed using Pearson residuals (Shaw et al. 2007b, Section 3.7) and found to have mean approximately zero and variance one with few outliers, demonstrating that these distributions appropriately model their respective stages.

Dependence of fitness on region and habitat

To examine how region and habitat influenced individual fitness and its components, we fit an aster model with fixed effects for site, blocks nested within site, region, soil type and the interaction among region and soil type. All model terms were specified at each life history stage. We used likelihood ratio tests to compare the fit of the full model to reduced models that sequentially dropped terms, beginning with the interaction, and retaining terms that improved the model fit when testing later terms. The block term was retained in all models. Maximum likelihood estimates of the response for each life history stage were obtained for a typical individual at each site from the final model that included all model terms. Because the estimate of seed count was obtained from a sub-sample of seedpods, it was necessary to transform the estimated average

number of seeds counted per plant to the average total number of seeds produced per plant (see Supplemental Information Technical Report for details). For the interior-loam site, in which herbivory by deer prevented us from obtaining data on seedpods, we predicted the average number of seedpods produced by each plant using a regression model of seedpods on mid-season height for all other populations at all sites ($r^2 = 0.49$).

Dependence of fitness on population

Maternal effects, due to the causal influence of the maternal phenotype on the offspring phenotype (Wolf and Wade 2009) and maternal environment (i.e. seed provisioning) often have the greatest effects at earlier life stages (Roach and Wulff 1987). To account for potential effects of source environment mediated by maternal phenotype or environment (hereafter, maternal environmental effects), the effect of population was included in the analysis in two ways. First, population was included in the model at the stage of early-season survival only. Second, population was included with effects at the later fitness stages of pods and seeds, in addition to early-season survival. The significance of population at different life history stages was then determined by using likelihood ratio tests to compare nested models as above. A significant population effect on later fitness stages when it was already included at the early-survival stage implied differences among population with respect to fitness, beyond early-expressed maternal environmental effects on survival. The effect of population on early-season survival beyond the effect of population on later life history stages was also tested by dropping the effect of population on early-season from the larger nested model. Because two

populations were not planted at all sites, the model including the interaction between population and region was over-specified and therefore failed to converge, though sub-models without this term converged. We tested this interaction by restricting the data to the populations planted at all sites (CRA, GCD, KZA). For the model terms tested in both sets of nested models, the results were consistent, so only the latter results including the population × region interaction are presented. Maximum likelihood estimates for each stage and total reproductive output were made for a typical individual from each population in each region – habitat combination as before.

Region- and habitat- specific selection

To examine how selection differed among regions and habitats, we focused on one growth trait, branch number, and one ecophysiological trait, SLA, because they, or traits correlated with them, are both expected to be subject to selection in this species (Etterson 2004). We expected larger plants, with more branches, to be favored at all sites. SLA is a trait on the spectrum between rapid growth and resource conservation, with higher values associated with faster growth and lower values with tolerance to resource limitation (Ackerly et al. 2002, Reich et al. 2003). Thus, we expected positive selection on SLA at the range edge and beyond, but stabilizing or negative selection within the range.

We fit aster models using lifetime fitness as the dependent variable, with linear and quadratic terms for branch number and SLA as covariates. Because of the complexity of the full aster model due to sub-sampling of pods, we used a reduced life history

through seedpods only for the selection analyses, which gave results similar to those with the full life history when populations were considered together (unpublished results). We focused on differences in selection among sites, rather than differences among populations, analyzing selection across all populations.

We performed four groups of hierarchical analyses to examine how selection differed between soil types and among regions. First, we asked if there was an overall dependence of the non-linear components of selection on site, by comparing a model where an interaction was specified between site and all trait terms (linear, quadratic and the cross-product) against a model without the interactions between the non-linear terms and site. We then tested for an overall dependence on site of the linear components of selection by comparing a model with the interaction between the linear terms and site to a model without this interaction. Second, to determine if soil type altered phenotypic selection within a region, we repeated these tests separately for the range edge and beyond range regions (it was not possible to do this test for the interior region because fitness data were available for only one site). Third, to examine how region influenced phenotypic selection, we repeated this model comparison for all sites within a soil type. Finally, at each site, we determined if there was curvature to the fitness landscape and linear selection on each trait.

We used the method of Lande and Arnold (1983) to determine the strength and direction of selection. For this analysis, we performed ordinary least squares regression using as the dependent variable the expected relative fitness for each individual (seedpods produced divided by average seedpods produced) obtained from the aster analysis, and

with phenotypic traits transformed to mean zero and standard deviation one as predictors (details in Geyer and Shaw 2010). The coefficients from this aster model are the maximum likelihood estimate of the linear selection gradient (β) and can be used to interpret the direction and magnitude of selection on each trait taking the other trait into account; use of aster predictions addresses the problem that fitness does not fit a parametric distribution (Shaw and Geyer 2010) and thus yields statistically appropriate confidence intervals via the delta method (Geyer and Shaw 2010). We examined the shape of the fitness landscape by calculating its mathematical stationary point, and plotting its contours of the estimated fitness function (Shaw et al. 2007b, Section 3.4). Stabilizing selection is inferred if the curvature is negative with a maximum within the phenotypic trait distribution, and disruptive selection if the curvature is positive with a minimum within the phenotypic trait distribution (Mitchell-Olds and Shaw 1987). Finally, we determined if populations differed from each other with respect to branch number and SLA at each site using ANOVA. We then compared the patterns of selection at each site to the population trait values, though these comparisons must be interpreted with caution because of the differing duration of time that populations have had to adapt to the regions they are growing in, and the potentially confounding effects of phenotypic plasticity.

Results

Region and soil type effects on fitness

Lifetime seed production was significantly higher in the interior (an average of 710 seeds produced for each seed planted) than in the edge (94 seeds/seed planted) or beyond edge (~ 10 seeds/seed planted) regions (Table 3-1, Fig. 3-3). Soil type also influenced lifetime seed production. At the range edge and beyond edge, seed production was greater at the loam site than at the sand site. In the interior region, the sites did not significantly differ (Fig. 3-3), with the caveat that our estimate of fitness at the interior-loam site came from mid-season height due to late-season herbivory.

Although more seeds were produced at loam sites in the edge and beyond edge regions, early-season survival was greater at the sand sites in both locations (Fig. 3-3). By contrast, in the interior region, early-season survival was slightly greater at the loam site than the sand site. Similarly, reproductive status (whether a plant produced any pods given survival) was greater at the sand sites in the range edge and beyond regions, and roughly equivalent at the interior sites. However, seedpod production, given that a plant reproduced, was much greater at the loam site than the sand site in both the edge and beyond edge regions (Fig. 3-3).

Population effects on fitness

Early-season survival differed significantly among source populations (Table 3-2), with the rank order of populations generally consistent across sites (Fig. 3-4). Northern edge population AFT had highest survival at three of four sites where it was planted, while the southern KZA population was second highest or highest at all sites (not shown). By contrast, lifetime seed production (i.e. overall fitness) showed some patterns

consistent with regional adaptation (Fig. 3-4). The northern populations AFT and GCD had greatest lifetime seed production beyond the range, and the southernmost population KZA having the greatest seed production within the range. However, in other cases the geographically closest populations to the transplant site had lower seed production than more distant populations, such as the northern AFT at the edge-sand site and the interior CRA population at the interior-sand site (Fig. 3-4). Although the northern populations had estimates of seed production greater than one (i.e. replacement) at both sites in the beyond edge region, the 95% confidence intervals of seed production included values below one (not shown), indicating that even these populations, though best-adapted to this region, may not maintain themselves here based on data from this year of study.

Patterns of selection

Among all sites, there was a significant interaction among site and both linear (deviance = 93.6, $P < 0.0001$) and non-linear components of selection on branch number and SLA (deviance = 61.0, $P < 0.0001$), implying that both the curvature and strength or direction of selection differed among sites. Beyond the range edge, soil type had a significant effect on non-linear (deviance = 10.6, $P = 0.01$) and linear (deviance = 19.7, $P < 0.0001$) components of selection. In the edge region, soil type did not significantly influence non-linear (deviance = 2.7, $P = 0.44$) selection, though it did influence linear selection (deviance = 47.4, $P < 0.0001$). Across the two loam sites, region had a significant effect on linear selection (deviance = 19.3, $P < 0.0001$), but not on non-linear selection (deviance = 5.4, $P = 0.15$). At the three sand sites, region had a significant

effect on both the non-linear (deviance = 24.0, $P = 0.0005$) and linear (deviance = 42.6, $P < 0.0001$) components of selection.

At all sites, there was significant curvature to the fitness landscape (Table S3-3) and linear selection on both branch number and SLA was highly significant (Table 3-3), with greater branch number favored at all sites. Selection on SLA was stabilizing (negative curvature and fitness optimum within range of phenotypic data) or slightly negative at all sites except the beyond-sand site where an increase in SLA was favored (Table 3-3, Fig. 3-5). For both branch number and SLA, the magnitude of selection varied as much between soil types within regions as it did among regions (Table 3-3). Across the sand sites, selection on branch number and SLA was stronger at the interior site than at the edge or beyond (Table 3-3). However, at the loam sites, selection on branch number and SLA was stronger at the beyond site than at the edge site.

Branch number differed significantly by site ($F_{4,605} = 210$, $P < 0.0001$), decreasing from the interior to the beyond region (Fig. S3-2). Though the populations differed from each other in branch number ($F_{4,590} = 8.1$, $P < 0.0001$), the relative order of the populations did not significantly change (pop × site: $F_{61,529} = 1.2$, $P = 0.20$) and there was no clear geographic pattern of differentiation (Fig. S3-2). SLA also differed significantly by site ($F_{4,592} = 79.5$, $P < 0.0001$), and on average was greatest at the interior-sand site and lowest at the edge-sand site (Fig. S3-3. The two southernmost populations (KZA, CUI) had slightly lower SLA than the three northernmost populations (CRA, GCD, AFT) ($F_{4,577} = 30.3$, $P < 0.0001$) at the sand sites in the edge and beyond regions, but not at the interior-sand site (Fig. S3-3) .

Discussion

While range shifts are a commonly studied response to climate change (Parmesan and Yohe 2003), whether species will establish in different habitats is rarely considered. In this study, we asked how habitat may influence where a native annual legume, *C. fasciculata*, will initially establish with a shift in its range beyond its northern limit and evaluated selection in those habitats. We found that lifetime seed production of *C. fasciculata* was always equivalent or greater in loam soil habitats, indicating that this is the preferred habitat for this species in the absence of competition. Further, we found that selection differed between regions and the strength of selection was influenced by the habitat within the region.

The finding of greater fitness on loam soils at the range edge conflicts with the observation of *C. fasciculata* being restricted to sandy habitats at its range edge in Minnesota. One explanation for this inconsistency is the biotic community associated with each soil type. In general, heterospecific vegetation was denser and more abundant at the loam sites (JSG, personal observation). Previous work has shown that neighbors have a negative effect on seedpod production (Chapter 1). By removing competitors at these sites, we likely masked the negative effect of competitors on seed production. Due to greater competitive intensity at the loam sites, this would have resulted in over-estimating fitness at the loam sites to a greater extent than at the sand sites (Fig. 3-3). Neighbors also have a facilitative effect on early-season survival (Chapter 1), and in this study we observed an increase in early-season survival at sandy sites in the edge and

beyond edge regions (Fig. 3-3). Thus, we tentatively suggest that at the range edge and beyond regions, sandy habitats may have the optimum level of heterospecific vegetation such that *C. fasciculata* seedlings gain maximum benefits of early season facilitation, and experience the least competitive effects on seedpod production.

Contrasting effects of region and habitat on different life history stages

Identifying the factors that determine species distributions requires inference of individuals' lifetime fitness. Our results indicate that inferences based on single life history stages can be misleading. For example, early-season survival was greater at sand than at loam sites at the range edge and beyond, though equivalent on the two soil types within the range (Fig. 3-3). Moreover, there was a significant effect of population source on early-season survival, and this effect was consistent across sites, suggesting that maternal effects have a significant effect on establishment, and therefore lifetime fitness. However, in reduced competition, the lower survival of individuals at loam sites at the range edge and beyond was countered by greater reproductive output at the loam sites, perhaps because of greater nutrient or water availability in loam soils. Thus, lifetime fitness is greater on loam soils at the range edge, in spite of lower survival there. Future work should elucidate if the effect of competitors differs between soil types. A greater negative effect of competitors on seed production at the loam sites would augment differences in early-season survival, with fitness effects at both survival and seed production stages explaining the shift to sandy habitats at the edge of *C. fasciculata*'s range, as suggested above.

Regional adaptation

Local adaptation is commonly, though not universally, demonstrated in populations (Leimu and Fischer 2008, Hereford 2009). We found some evidence for regional adaptation at the extreme transplant sites (beyond and interior), while there was no clear pattern at the edge sites (Fig. 3-4). This is suggestive given the small sample sizes and non-adaptive differences in early-season survival among populations attributable to maternal environmental effects (see above). The lack of evidence for local adaptation in the edge region is likely because all the populations came from within ~1000km, a distance in which previous studies have not found evidence for local adaptation (Galloway and Fenster 2000, Etterson 2004). However, the environments of the beyond edge and interior transplant sites were sufficiently distant from the furthest transplanted populations in each region to influence the expression of local adaptation.

Though the northern AFT population had greater than average fitness beyond the range edge, it had low fitness at the edge-sand site, which was geographically closest to the population source. This lack of consistent local adaptation could be due to drift, as the source population was small, with fewer than 1000 individuals in the year seed was collected, a critical size below which there are no examples of local adaptation in plants (Leimu and Fischer 2008). For the interior CRA population, the lack of local adaptation may be due to contrasting soil conditions between its source, a loam site, and the interior transplant site where fitness was evaluated, a sand site (~50km north of its source). At its home site on loam soil, the CRA population was tallest, implying that it would have had

greatest seed production at this site (barring the major episode of herbivory). However, at the sand site, the CRA population produced about half as many seeds as the southern KZA population (Fig. 3-4). This finding suggests that as colonists disperse, they may have larger than expected changes in fitness over short geographic distances if they land in different habitats than those where they originated.

Shifts in the optimum phenotype

At the sandy sites, there appears to be a shift of the optimum value of SLA among regions, with selection favoring slightly intermediate values of SLA within the range, but higher values beyond the range at the sand site (Table 3-3, Fig. 3-5). As increased SLA is associated with a suite of traits that favor rapid growth (Reich et al. 2003), this pattern of selection is likely due to the cumulative effect of selection on these traits favoring faster growth beyond the range edge. Such shifts in optima, when combined with gene flow among populations, are commonly implicated in explaining why species' have limited ranges (Kirkpatrick and Barton 1997). For example, Angert *et al.* (2008), using experimental hybrids of two monkeyflower species with divergent altitudinal distributions, showed that adaptation to either high or low elevations came at a fitness cost to growth in the contrasting environment.

The direction of selection on SLA also corresponds with patterns of phenotypic plasticity among populations. We found that at the beyond-sand site, where greater SLA was favored (Fig. 3-5), the northernmost populations (AFT, GCD, CRA) had greater mean SLA, showing greater plasticity at that site than the two southern populations, in

comparison with the expression of SLA at the edge-sand site (Fig. S4-3). This suggests that plasticity of the northern populations contributes to their fitness, as previous studies of *C. fasciculata* have also found (Etterson 2004).

Implications for range expansion

Though we cannot definitively conclude that any of the populations have seed production sufficient to maintain themselves beyond the range edge, some individuals did survive and reproduce demonstrating the potential for persistence in these locations and, given suitable genetic variation, the potential for further adaptation to support range expansion. Thus, why *C. fasciculata* has not expanded its range and adapted to persist in this region is unclear. There are many potential explanations for this lack of range expansion. First, we have likely overestimated seed production because we do not account for dormant-season mortality and post-dispersal seed predation. Thus, absolute fitness may be much lower, such that colonizing populations go extinct before they can further adapt. The issue of not completely measuring complete lifetime fitness is widespread in ecological genetic studies, often for practical reasons, as most studies go from planted seed to seed production (Geber and Eckhart 2005), seedling to seed production (Etterson 2004), or use some correlated measure of lifetime fitness (Angert and Schemske 2005, Baack et al. 2006). Typically, this will result in overestimates of fitness and species' potential distributions.

Second, the current range limit may not be stable, and *C. fasciculata* is slowly expanding its range toward the north, but the rate of colonization is limited because seed

dispersal distances are small (Fenster 1991a), so few colonists may arrive at sites beyond the range, resulting in little opportunity to adapt. Moreover, the rate of range expansion is likely to be constrained by interannual variability in weather, with benign summers allowing colonists to establish, and harsh years eliminating nascent populations. The conditions at the range edge and beyond in this year were similar in temperature but considerably drier than average years (Table 3-4), suggesting that in more favorable years colonists will have greater fitness, and thus the opportunity to establish viable populations and adapt.

Third, adaptation to conditions beyond the range edge may be prevented by shifts in optimum trait values in different regions combined with swamping gene flow from central populations (Antonovics 1976). While gene flow occurs over a few meters in this insect-pollinated species (Fenster 1991a), we did observe such a shift in the trait optimum in SLA, and thus infrequent long-distance gene flow from interior populations could limit the rate of adaptation at the range edge.

Finally, adaptation to conditions across the range edge may be prevented by competition or other biotic interactions. As we discuss above, if the effect of competitors is strong enough, individuals may survive, but fail to establish viable populations at any sites beyond the range edge. An evolutionary stable range limit may occur because other organisms have already filled *C. fasciculata*'s ecological niche beyond its current range edge, and thus it does not have the opportunity to adapt to climatic conditions in this region (Price and Kirkpatrick 2009). Alternatively, other biotic interactions with mutualists, herbivores or disease may limit range expansion. For instance, in another

study, we have found that the availability of mutualistic rhizobia is limited beyond the range edge, reducing plant fitness (Chapter 2)(Stanton-Geddes and Anderson 2011).

Conclusion

In conclusion, the results of this study show that soil type will influence where colonists will establish as species shift their ranges in response to climate change, and that other aspects of the environment that are influenced by soil type, such as competitors, may also play important roles in determining where populations will persist. As species are unlikely to shift their ranges synchronously, this suggests that range expansion may be limited not only by the rate at which colonists disperse, but also the rate at which competitors in the novel region retract their ranges or decrease in competitive ability.

Funding and Assistance

We thank the managers at Conard Environmental Research Area of Grinnell College, Cuivre River State Park, Konza Prairie, Grey Cloud Dunes SNA and Afton State park for permission to collect seed, Larissa Mottl at Conard Environmental Research Area, Mike Stegman at the Marshall County Conservation Board, Shawn Schottler at St. Croix Watershed Research Station, Al Lieffort at Douglas County Parks, and the Stanton-Geddes family for the use of field sites, Charlie Geyer for statistical assistance, and Peter Reich for comments. This work was supported by the Rothman and Dayton Research Fellowships at the Bell Museum, the Carolyn Crosby Fellowship from the Graduate

School at the University of Minnesota, the Minnesota Center for Community Genetics, the Ray C. Anderson Fellowship and the John Hall Memorial Research award to PT and RGS. NSF DEB-0544970 supported RGS and DEB-0515466 supported PT.

Tables and Figures

Table 3-1. Summary of aster model comparisons for effects of region and soil on *C. fasciculata* lifetime seed production. Analysis of deviance (-2 log likelihood) and χ^2 *P*-values for each model test, block and the interaction term are tested against the full model, while the region and soil terms were tested against the (region:soil) model without the interaction term.

Model	**Model d.f.**	**Model deviance**	**Test d.f.**	**Test deviance**	***P*-value**
Full model	45	50430	-	-	-
Block	30	50450	15	21.4	0.13
Region × soil	40	50881	5	458	<0.0001
Region	30	54376	10	3542	<0.0001
Soil	35	51299	5	422	<0.0001

Table 3-2. Summary of aster model comparisons to test for effects of population, region and soil and all interactions on *C. fasciculata* fitness for populations planted at all sites (CRA, GCD, KZA). The effect of population was tested at multiple levels of the life history graph (@seeds, @ pods and early-season survival (@esurv). Analysis of deviance (-2 log likelihood) and χ^2 *P*-values for each model test; block and the interaction terms are tested against the full model, while the population, region and soil terms were tested against the model without interactions.

Model	**Model d.f.**	**Model deviance**	**Test d.f.**	**Test deviance**	***P*-value**
Full model	93	42649	-	-	-
Block	78	42666	15	16.8	0.33
Pop × region	69	42738	24	88.8	<0.0001
Pop × soil	81	42677	12	28.7	0.005
Region × soil	88	42738	5	360.8	<0.0001
Without interactions	46	43177	-	-	-
Pop@esurv	40	43243	2	61.1	<0.0001
Pop@pod	42	43182	2	1.1	0.59
Pop@seed	44	43181	2	4.0	0.14
Region	32	46292	10	3109	<0.0001
Soil	37	43672	5	490.1	<0.0001

Table 3-3. Standardized selection gradients of branch number and specific leaf area (SLA) with 95% confidence intervals from ordinary least squares (OLS) regression using relative fitness estimated from aster models as the dependent variable and standardized phenotypic traits as predictors.

Region	**Soil**	**Trait**	**Selection gradient (β)**
Beyond	Loam	branch #	3.62 ± 0.65
		SLA	-0.12 ± 0.34
Beyond	Sand	branch #	1.07 ± 0.22
		SLA	0.65 ± 0.11
Edge	Loam	branch #	1.95 ± 0.44
		SLA	-0.06 ± 0.36
Edge	Sand	branch #	1.07 ± 0.19
		SLA	0.20 ± 0.14
Interior	Loam	branch #	n.d.
		SLA	n.d.
Interior	Sand	branch #	2.36 ± 0.49
		SLA	-1.06 ± 0.47

Table 3-4. Weather data for the growing season (May-September) in the year of the experiment, 2008, compared to the 1971-2000 climate normals for these months. For 2008, daily data for 1 May – 30 September 2008 were downloaded from www.wunderground.com from the closest weather station to the field sites in each region, and monthly mean temperature and total precipitation were calculated. The climate normals are available from NOAA* and mean temperature and precipitation were calculated for the same date range. The weather stations were the same for the interior and edge regions, but were 24km apart in the beyond edge region.

	Mean Temperature (°C)			**Precipitation (mm)**		
	2008	**average**	**% change**	**2008**	**average**	**% change**
Beyond	18.2	17.5	+4%	294	376	-22%
Edge	19.8	19.3	+2%	314	525	-40%
Interior	19.0	19.7	-4%	627	576	+9%

* http://lwf.ncdc.noaa.gov/oa/climate/normals/usnormals.html

Figure 3-1. Map of seed source populations (circles, Table S3-1) and transplant common garden locations (triangles, Table S3-2). The dotted line is the approximate range edge for in this region, based on USDA Plants Database county level information. The source populations are: AFT, Afton State Park; GCD, Grey Cloud Dunes Scientific and Natural Area; CRA, Conard Environmental Research Area; KZA, Konza Prairie Biological Station and CUI, Cuivre River State Park. The transplant common gardens are: Beyond-sand (LIDA), field near Lake Ida, MN; Beyond-loam (KRCP), Kensington Runestone County Park, Kensington, MN; Edge-sand (CCES), Cedar Creek Ecosystem Science Reserve; Edge-loam (SCWR), St. Croix Watershed Research Station; Interior-sand (IRCA), Iowa River Conservation Area; Interior-loam (CERA), Conard Environmental Research Area. The map was created using DIVA-GIS (Hijmans et al. 2004).

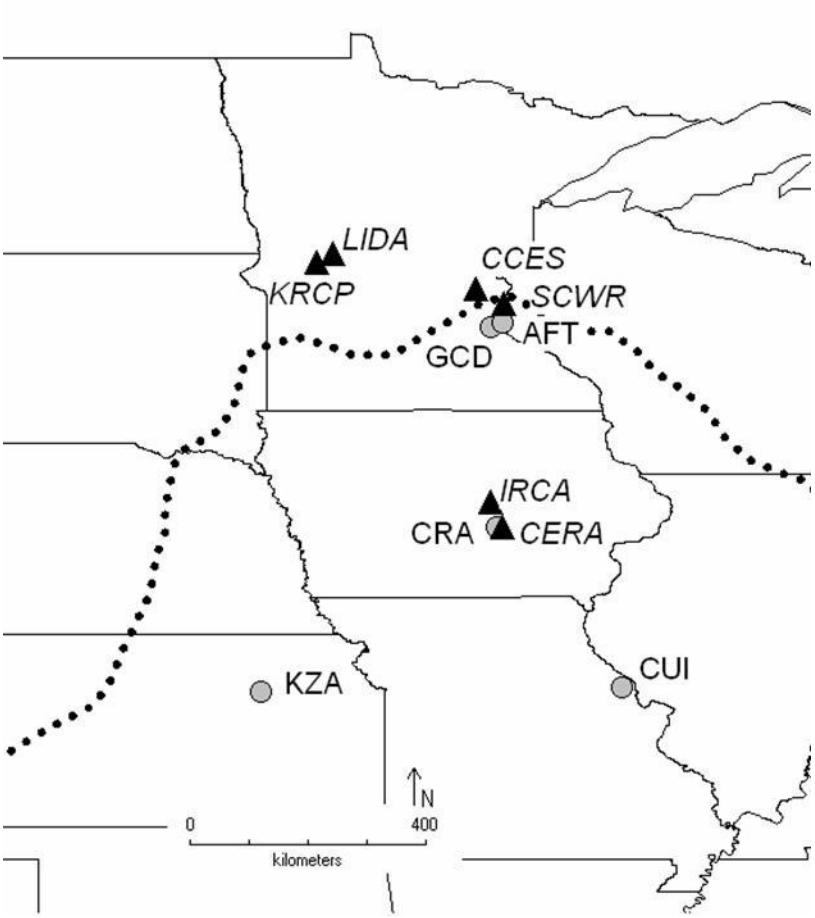

LIDA
KRCP
CCES
SCWR
AFT
GCD
IRCA
CRA
CERA
KZA
CUI
N
0
400
kilometers

Figure 3-2. Graphical model for aster analysis with the distribution for each life history stage given below. Reproductive status is whether a plant reproduced or not, and pods sampled is the random sample of the total pods that were collected to count seeds per pod. For the selection analysis, a separate graphical model that went only through seedpods was used.

Figure 3-3. Maximum likelihood estimates and standard errors from aster models for early-season survival, reproductive status given survival, seedpods produced given reproducing, and lifetime seed production, integrating across the previous stages, at each site (white = sand, black = loam) in each region. At the interior loam site (dashed bars), all plants were destroyed before end of season measurements were made so the values for survival and reproductive status come from the observed data, and the seedpod and lifetime seed production data comes from the regression of each variable on mid-season height at the other sites. Using this regression equation, the estimated lifetime seed production in the interior region at the sand site (1107 seeds) is greater than at the loam site (846 seeds). The inset in the top panel shows the differences between sites in lifetime seed production at the "beyond" region.

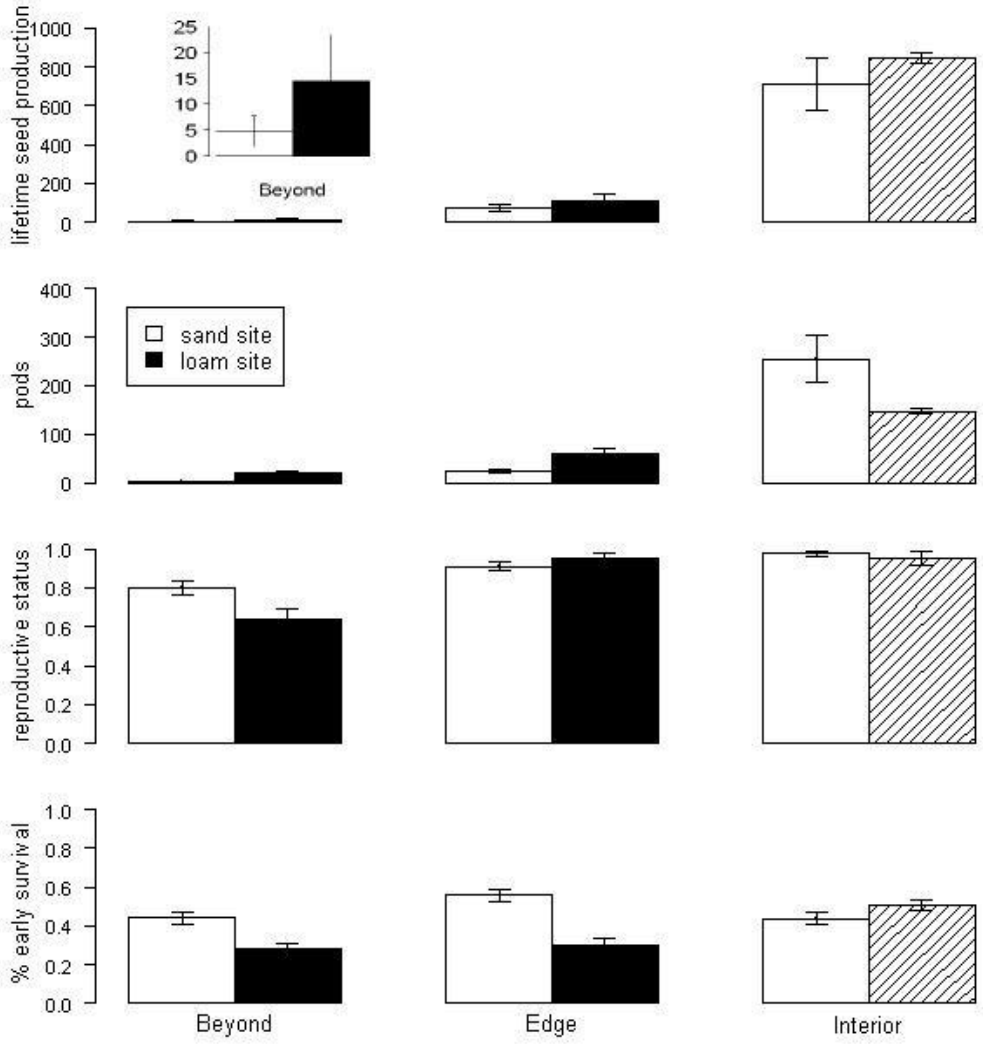

lifetime seed production
1000
800
600
400
200
0
25
20
15
10
5
0
Beyond
pods
400
300
200
100
0
sand site
loam site
reproductive status
1.0
0.8
0.6
0.4
0.2
0.0
% early survival
1.0
0.8
0.6
0.4
0.2
0.0
Beyond
Edge
Interior

Figure 3-4. Maximum likelihood estimates of lifetime seed production with standard errors from aster models for each population (source locations in Figure 3-1) at each site. Left column is sand sites, right column is loam sites; top row is beyond edge region, middle row edge region, and bottom row is interior region. Populations are organized from north to south, and a box designates the two range edge populations (AFT, GCD) across the regions. At the interior loam site, mid-season height is presented because of herbivory at this site prior to end-of-season data collection. The plotted values for the AFT and CUI populations (dashed boxes) are means of the observed data because we were unable to include them in the aster model as they were not planted at all sites (see text for details).

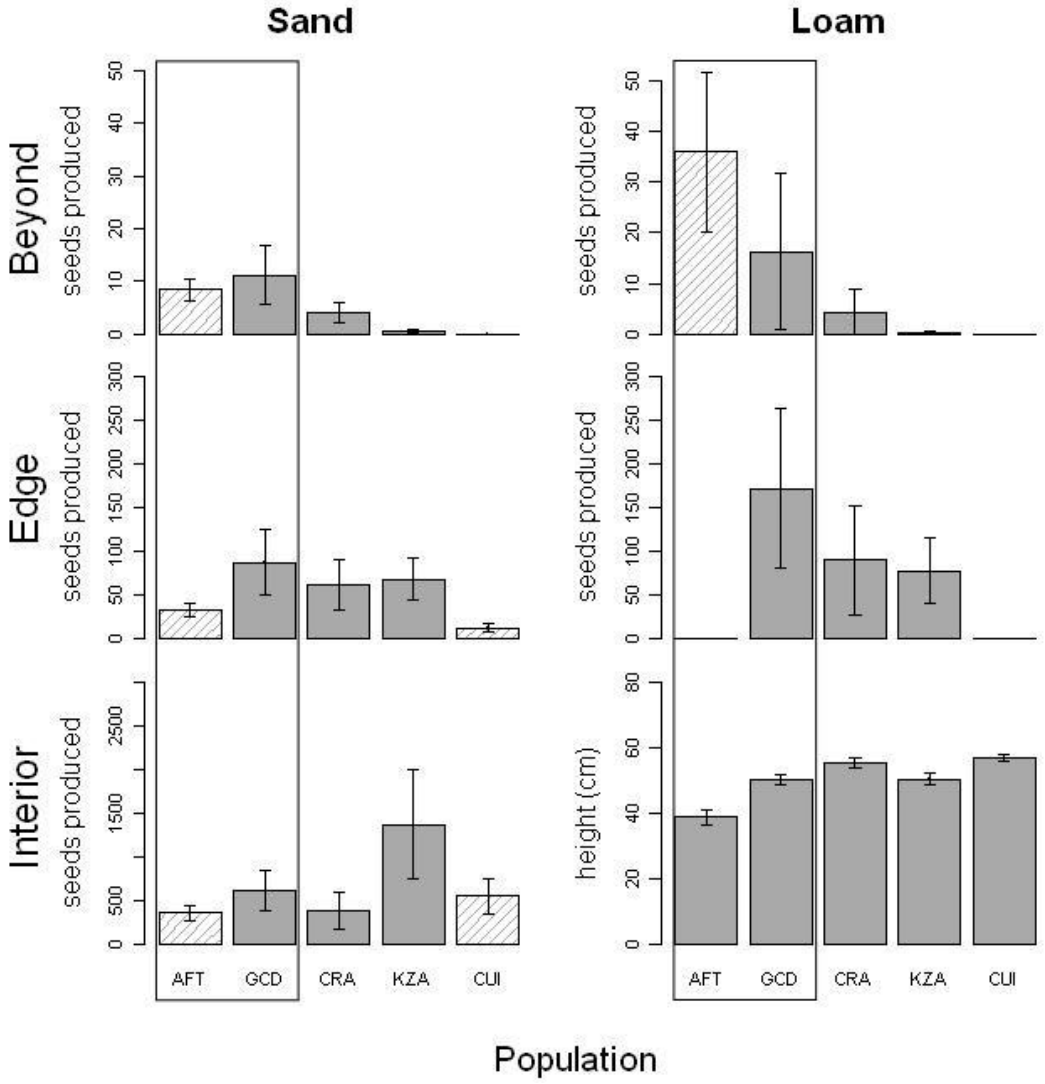
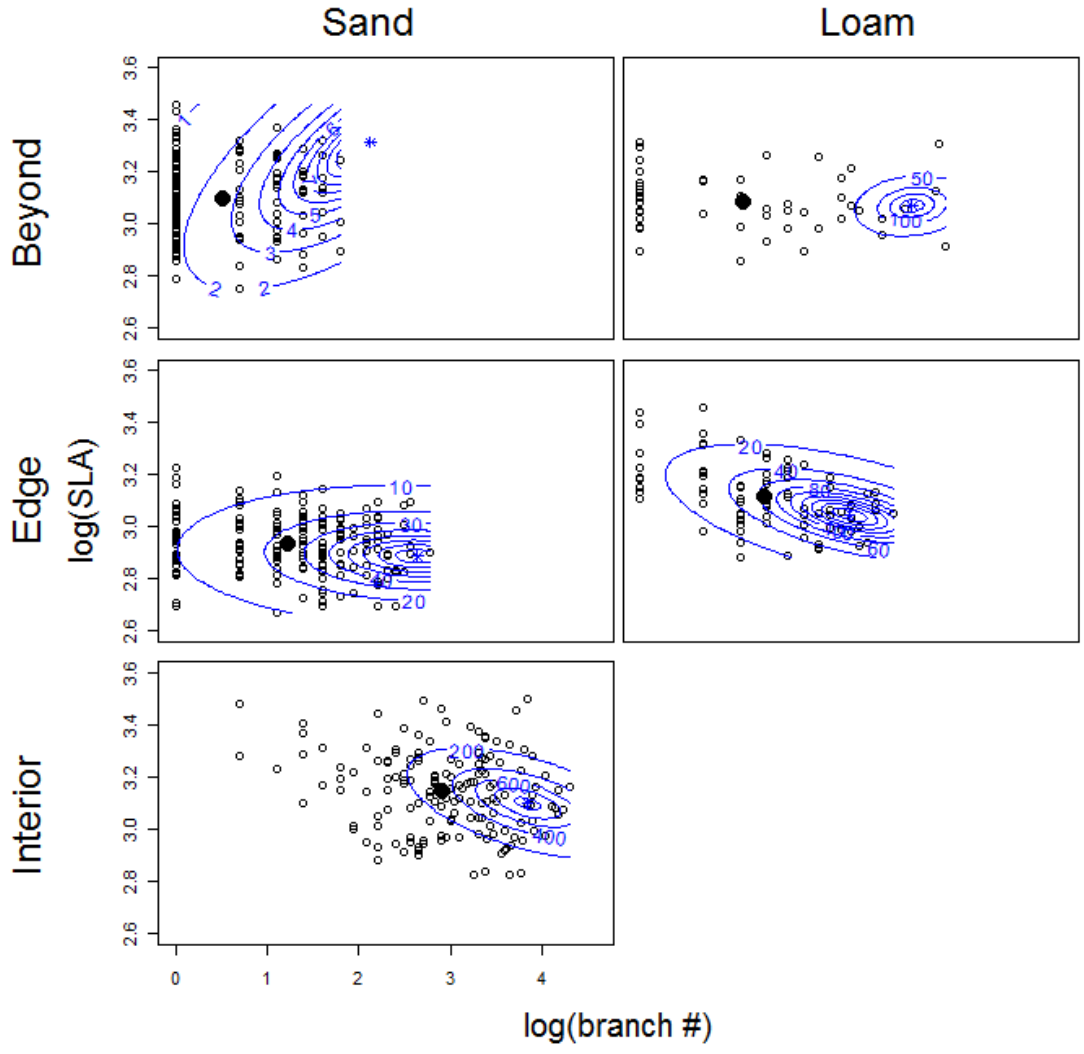

Sand
Loam
Beyond
Edge
Interior
seeds produced
height (cm)
AFT
GCD
CRA
KZA
CUI
Population

Figure 3-5. Unconditional fitness landscape of log(SLA) and log(branch #) for each site. The contours give estimated seedpod production (i.e. fitness) from the quadratic aster models at each site, restricted to the trait area where plants were observed. Circles are the observed bivariate trait values for each plant at a site, the large filled circle is the mean phenotype, and the asterisk is the maximum of the fitness landscape. Large areas without contours are 'flat' regions of the fitness landscape. The interior loam site is missing because no fitness data was collected at this site due to deer herbivory.



Supplemental Information

Table S3-1. *C. fasciculata* seed source population information, arranged from north to south origin. Mean annual temperature (MAT) and annual precipitation (PPT) were collected from the WorldClim data set implemented in DIVA-GIS (Hijmans *et al*. 2004).

Population	Location	Lat/Long	Region	MAT) (°C)	PPT (mm)
AFT	Afton State Park Afton, MN	44°51’35N 92°46’21W	N. Edge	7.0	778
GCD	Grey Cloud Dunes Scientific and Natural Area, Cottage Grove, MN	44°47’19N 92°57’29W	N. Edge	7.5	754
CRA	Conard Environmental Research Area (Grinnell College), Kellogg, IA	41°41’14N 92°52’16W	Interior	8.9	882
KZA	Konza Prairie Biological Station, (Kansas State Univ.), Manhattan, KS	39°07’07N 96°32’15W	West edge	11.9	876
CUI	Cuivre River State Park Troy, MO	39°10’30N 90°55’17W	Interior	12.6	993

Table S3-2. Transplant site locations, and climate and soil characteristics. Mean annual temperature (MAT) and annual precipitation (PPT) were collected from the WorldClim data set implemented in DIVA-GIS. Soil types were verified using the hydrometer method (Dane *et al*., 2002) to determine the fraction of soil that was sand, silt and clay, except at CCES where soil data were already available (Grigal *et al*., 1974).

Site	Location	Lat/Long	Region	Soil	MAT (°C)	PPT (mm)	% sand	% clay
KRCP	Runestone County Park, Kensington, MN	45°48'43N 95°39'55W	Beyond	Loam	5.2	635	38	33
LIDA	Lake Ida, Douglas County, MN	45°57'05N 95°25'31W	Beyond	Sand	5.2	596	71	11
SCWR	St. Croix Watershed Research Station, Marine-on-St. Croix, MN	45°10'04N 92°45'53W	Edge	Loam	6.8	774	68	29
CCES	Cedar Creek Ecosystem Science Reserve, Bethel, MN	45°24'10N 93°11'28W	Edge	Sand	6.3	751	94	1
CERA	Conard Environmental Research Area, Grinnell College, Kellogg, IA	41°41'03N 92°51'42W	Interior	Loam	8.9	882	33	14
IRCA	Iowa River Conservation Area, Marshalltown, IA	42°04'31N 92°57'02W	Interior	Sand	8.4	854	83	6

Table S3-3. Summary of aster model tests for non-linear selection at each site. All models have three test degrees of freedom, including both quadratic terms and the cross product, tested against the larger model with 11 degrees of freedom.

Region	**Soil**	**Model deviance**	**Test deviance**	***P*-value**
Beyond	Loam	468.16	105.17	< 0.0001
Beyond	Sand	490.97	77.51	< 0.0001
Edge	Loam	509.45	171.62	< 0.0001
Edge	Sand	954.29	274.26	< 0.0001
Interior	Loam	no data	-	-
Interior	Sand	1227.6	189.9	< 0.0001

Figure S3-1. Principal components analysis of annual mean temperature, mean temperature of warmest quarter, annual precipitation and precipitation of warmest quarter for transplant sites, implemented using the prcomp function in R (R Development Core Team, 2009). PC1 explains 95% of the total variation, and PC2 an additional 4%. PC1 is associated with increasing mean annual temperature and precipitation, while there are no clear loadings on PC2. Sites in each region cluster together along PC1. Interior sites (IRCA, CERA), edge sites (CCES, SCWR), beyond edge sites (KRCP,LIDA).

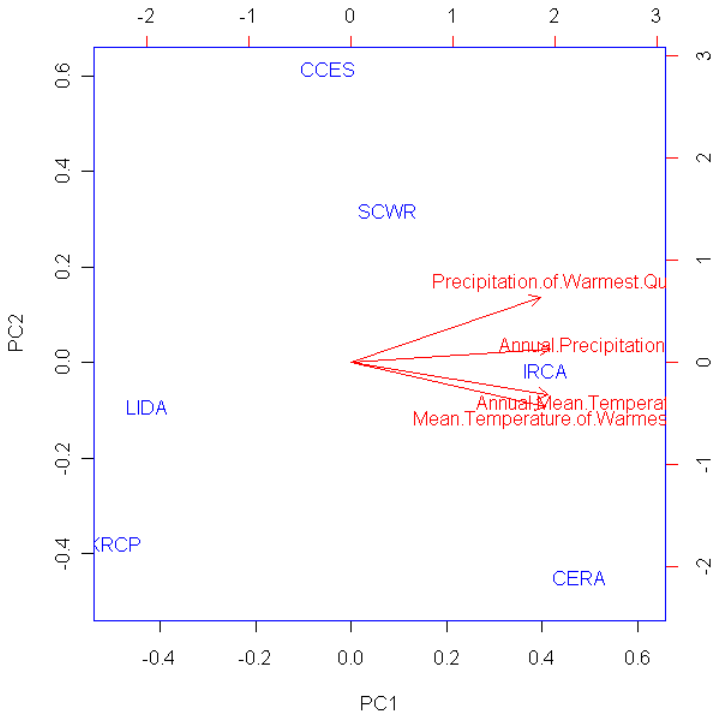


Figure S3-2. Least squares means and standard errors for log branch number for each population at each site.

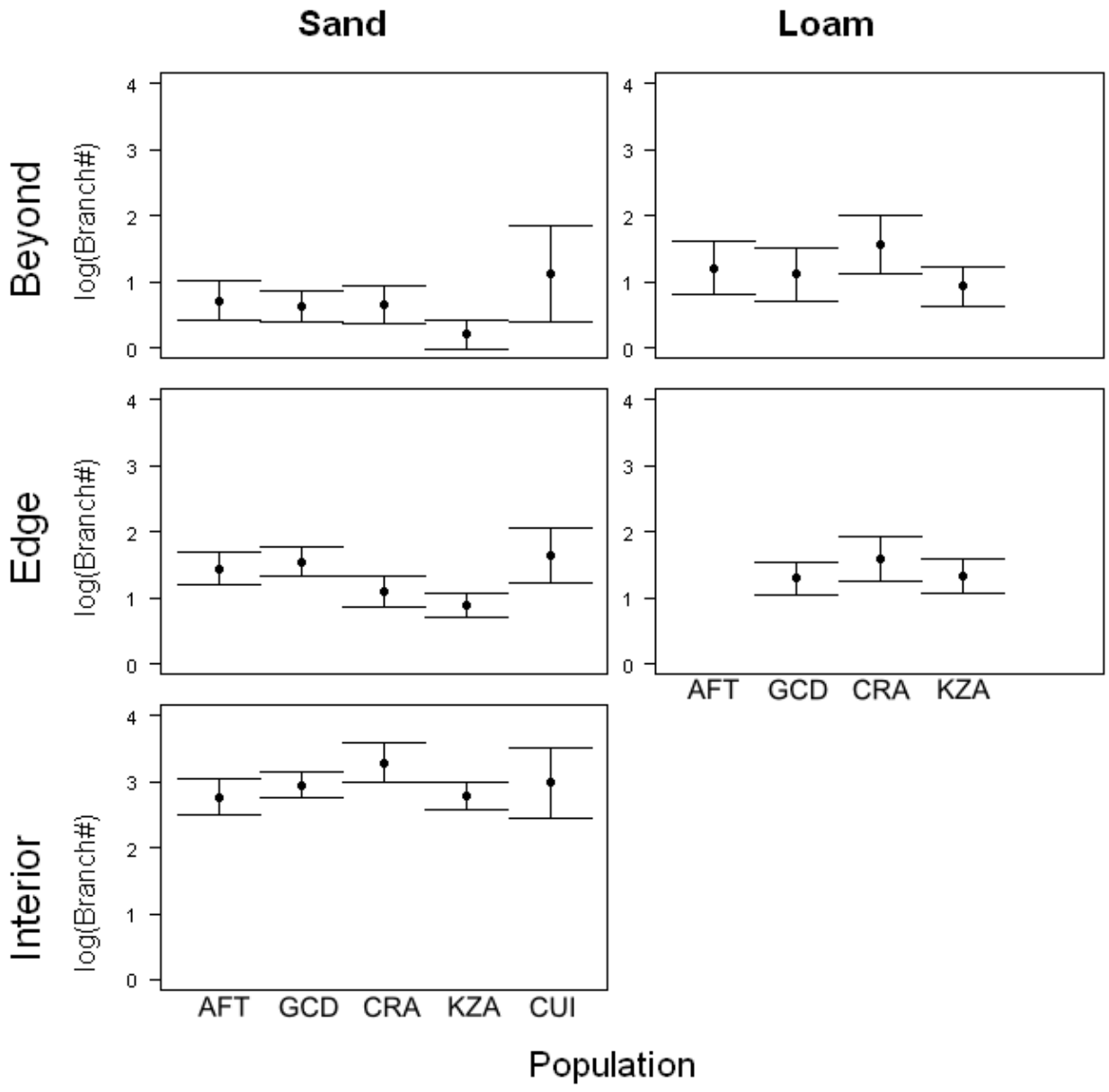
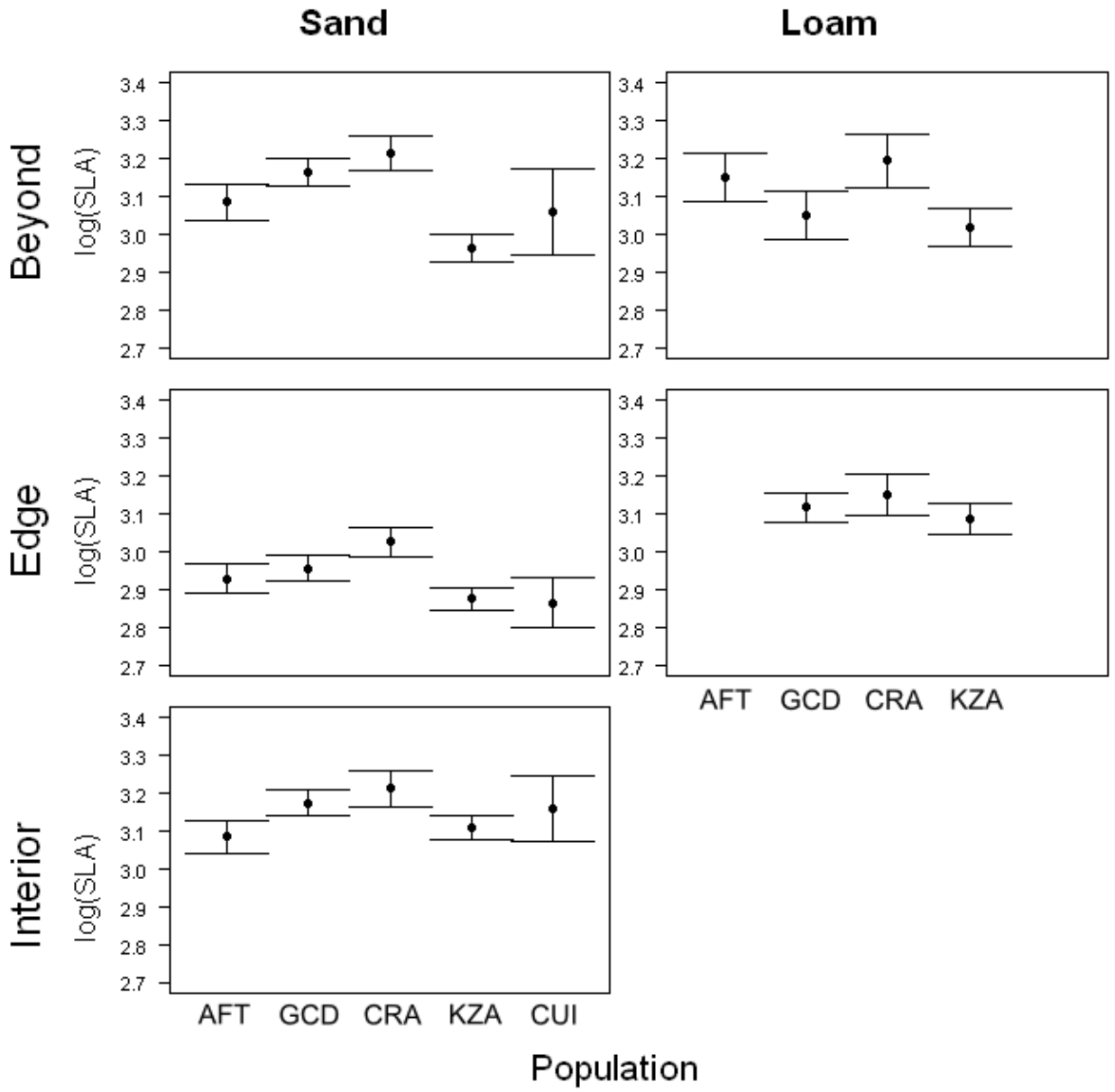


Figure S3-3. Least squares means and standard errors of log(SLA (m^2/g) for each population at each site.



Technical report for the sub-sampling of a fitness stage and transformation to estimate absolute fitness in aster models

Aster models allow joint analysis of multiple life history stages conditioned upon the previous stage (Geyer et al. 2007, Shaw et al. 2008). Ideally, all data (i.e. survival, offspring produced) is recorded at each stage, and can then be appropriately modeled in the aster analysis to give overall fitness. However, as pointed out by Shaw *et al*. (2008, p E43) , it is not always possible or practical to record all data for all life history stages, such as seeds produced, and that sub-sampling is a common practice. Shaw *et al.* (2008, p E43) show how it is possible to specify a node for the stage that is sub-sampled, and then the result for overall fitness would be proportional to the value for total fitness estimated from the aster model. It is necessary that the sub-sampling stage is distributed according to an exponential family distribution such that it can be included in the aster graph, and thus cannot be a constant value. That is, sampling of, for example, seedpods should be a percentage (i.e. 10%) of the total seedpods and not a fixed value (i.e. 10).

In the "Aster Technical Report No. 661", Shaw *et al.* (2007a, Section 8) give a simulation example showing that if data is collected appropriately, sub-sampling will give appropriate estimates for total fitness with aster models. A key point is that the sub-sampling must be done with *a priori* knowledge of the requirements of aster modeling for the stage to be modeled with an exponential family distribution. The method recommended for the sub-sampling node is to choose a binomial (*n*,*p*) distribution, where *n* is specified by the sample size of the previous node that is being sub-sampled, and *p* is the success probability, which is determined in advance (i.e. 10% of all seedpods), but simply estimated by maximum likelihood as are the other conditional parameters. The model will estimate a different *p* in each class where *p* is known to be different (i.e. for each site, see below). The number of seedpods collected does not need to be exactly *p* for each individual, as the coefficient will simply be estimated as are the other coefficients, and thus only the expectation of seedpods sampled in each class should be equal to *p*. For our analysis, we choose to subsample a different proportion of seedpods

at each site because plants varied and size and seedpod production varied by an order of magnitude. Thus, we sub-sampled approximately 10% of seedpods on each plant at the range interior site, approximately 30% of seedpods at the range edge sites, and approximately 75% of seedpods at sites beyond the range.

When setting up the aster graphical model for this analysis, we included a "sample" node in the graph (Figure 2, reproduced below) that is the fraction of seedpods sampled from the total number counted ("seedpods" node). This node is fit with a Bernoulli distribution, which is in effect a Binomial(*pods*, *p*) distribution, where *pods* is determined by the previous node of the graph (number of seedpods) and *p* is the known fraction of pods sampled (as above). It is possible to check that the "sample" node is correctly estimated by getting the conditional mean value parameters for each stage (Shaw et al. 2007a, Section 8.2.1) and verifying that the estimate for "sample" reflects the known value of *p*.

Though the aster modeling is simple, difficulties arise when making estimates of total fitness for each individual, as the sub-sampling stage results in the unconditional estimates of fitness being proportional, but not equal, to true total fitness. To get true lifetime fitness, it is necessary to transform the final node by the proportion sub-sampled. For our data, to get individual plant lifetime fitness, we needed to transform the estimated average number of seeds counted per plant, to the average total number of seeds produced per plant. In this case, this is:

$$w = \frac{u_6 * u_3}{u_4}$$

where *w* is lifetime fitness and *U* is the vector of the estimated mean value parameters from the aster model, thus u_6 is the maximum likelihood estimate for seeds counted (node 6 of the graph), u_3 is the maximum likelihood estimate for pods (node 3) and u_4 is the maximum likelihood estimate for pods sampled (node 4).

The calculation of standard errors and confidence intervals for this transformation requires using the delta method to get the asymptotic variance – covariance matrix for *w*.

To do this, we calculated the Jacobian matrix for the transformation of u_i to w:

$$J = \left(\frac{u_6}{u_4}, \frac{-u_6 u_3}{{u_4}^2}, \frac{u_3}{u_4} \right)$$

and then we calculated the asymptotic variance – covariance matrix (M_u)for U using the component gradient,which is $\partial U / \partial w$, from the output of the predict.aster function (Geyer and Shaw 2010) and the component fisher, which is the expected Fisher information, from output of the aster function (in this example, called "model") follows:

$$M_u = \frac{\partial U}{\partial w} \times (fisher)^{-1} \times (\frac{\partial U}{\partial w})^T$$

In R, this is written as:

M_u <- gradient %*% solve(model$fisher) %*% t(gradient)

Then, by the delta method, the asymptotic variance – covariance matrix for w is:

M_w <- $J * M_u * J^T$

and the diagonal elements of this matrix give the standard errors for w. Code for this was modified from (Geyer and Shaw 2010, p 6-7).

The R script and data needed to repeat this analysis are available upon request.

Chapter 4

Population genetics of range edge and interior populations in a widely-distributed annual plant

Summary

Species range limits are determined by a combination of historical, ecological and genetic factors. While most of our understanding of range limits comes from ecological genetic studies, molecular population genetic methods can provide unique insight on the role of range expansion and gene flow in determining range limits. We sequenced nine nuclear loci (5410 base pair) in 68 individuals from four populations, located within the range and at both western and northern range edges, of the native annual legume *Chamaecrista fasciculata* in the Midwest of North America. We calculated standard estimators of nucleotide diversity and demographic history, inferred population structure and used coalescent methods to assess the directionality of gene flow. Molecular diversity was greatest in the most interior population, and significantly reduced at both the western and northern range edges. The western edge population was distinct from the remaining populations, which showed evidence of northward range expansion. Coalescent models indicate that gene flow was high among all populations, and predominately northward. In the context of data from transplant studies, we tentatively suggest that the western range edge is limited by extinction-recolonization dynamics, and the northern range edge is limited by ecological-genetic processes.

Introduction

The edges of species' distributions occur where they are limited by dispersal or where their growth rate is no longer adequate to maintain viable populations. A combination of historical (e.g. range expansion), ecological (e.g. climate, biotic interactions) and genetic (e.g. genetic constraints and gene flow) factors determine the locations of these edges. Establishing the relative importance of history, ecology and evolution for determining the distributions and abundances of species will further our understanding of past range shifts as well as species' potential to shift their ranges and adapt to contemporary climate change (Davis and Shaw 2001, Parmesan and Yohe 2003)

Our understanding of species range limits comes largely from ecological studies. These include observational studies of individual fitness, or fitness surrogates, (Carter and Prince 1985, Kavanagh and Kellman 1986, Caughley et al. 1988) and demographic rates (Norton et al. 2005, Angert 2006, Eckhart et al. 2011) at different geographic range locations. Further, ecological genetic experiments are used to evaluate changes in fitness beyond the range edge (Prince and Carter 1985, Angert and Schemske 2005, Geber and Eckhart 2005, Samis and Eckert 2009), and how phenotypic selection differs among range locations (Jenkins and Hoffman 1999, Etterson 2004, Griffith and Watson 2005, Angert et al. 2008). These studies demonstrate the extent to which range expansion is constrained in contemporary environments and yield insights into the dependence of further range expansion on adaptation in particular traits (Griffith and Watson 2006), but

do not reveal history of past range expansion or gene flow. Molecular population genetic approaches can contribute to our understanding of range expansion and gene flow (Ross-Ibara et al. 2008, van Heerwaarden et al. 2009, Keller et al. 2010, Moeller et al. 2011). Though ecological genetic and population genetic approaches are often performed separately, a synthetic approach that merges them is necessary for a comprehensive understanding of the factors that limit a species range (van Heerwaarden et al. 2009, Moeller et al. 2011).

Population genetic analyses provide insight into demography, population structure and gene flow that can help us to understand the history of species' range limits (Excoffier et al. 2009, Keller et al. 2010, Moeller et al. 2011). For example, recently founded populations, to the extent that they have undergone population bottlenecks, are expected to have decreased genetic diversity and a deficit of rare variants compared to interior populations (Excoffier *et al.* 2009), which often (Andrewartha and Birch 1954, Brown 1984) but not always (Sagarin and Gaines 2002) contain more individuals. Further, if these recently founded populations are part of an ongoing range expansion, some rare variants may quickly increase to high frequency at the range edge due to genetic drift (Excoffier and Ray 2008, Excoffier et al. 2009). By contrast, populations at a demographic range limit characterized by extinction – colonization metapopulation dynamics are theorized to have even more strongly reduced genetic diversity than recently founded stable populations (Pannell and Charlesworth 1999, Pannell and Charlesworth 2000) and few rare variants as they are expected to be lost to genetic drift (Wakeley and Aliacar 2001, Moeller et al. 2011).

In addition to insights on the demographic dynamics involved in range expansion, population genetics can reveal population structure and patterns of gene flow among populations. This knowledge is important because asymmetric gene flow from interior to edge populations introducing alleles that compromise adaptation at these sites is one of the primary evolutionary explanations for species' range limits (Antonovics 1976, Hoffman and Blows 1994, Kirkpatrick and Barton 1997, Bridle and Vines 2007). Specifically, when the environments are selectively different, high rates of gene flow from interior to edge populations concurrent with limited differentiation among populations is consistent with the hypothesis that gene flow limits range expansion. Directional gene flow from interior to edge populations has been suggested to limit range expansion of *Drosophila serrata* (Jenkins and Hoffman 1999). However, in organisms with limited dispersal and large ranges, the potential for gene flow to limit range expansion is unclear (Barton 2001, Silvertown 2001).

In this study, we examine patterns of population genetic diversity and the directionality of migration among four populations of the annual legume *Chamaecrista fasciculata*. Two of the populations were located within the species' range, one population was located near the western range edge, and the fourth population was near the northern range edge (Fig. 4-1). We used sequence data to examine how geographic range location influences (1) haplotype richness, nucleotide diversity and the number of unique haplotypes, (2) the frequency distribution of genetic variants and (3) asymmetric rates of migration among populations. While population genetic data can be informative for understanding range edge dynamics, similar patterns can result from different

processes (Moeller et al. 2011). Thus, to come to an integrative perspective on range limits in this species, we interpret these results in the context of field experiments examining individual fitness within and beyond the north and western range edges of *C. fasciculata*.

Materials and methods

Study system and sample collection

Chamaecrista fasciculata (Fabaceae) is an insect-pollinated annual legume native to central and eastern North America. The northern range limit for this species occurs from southern Minnesota east to Connecticut, and the western range limit is from western Minnesota south into Mexico (Fig. 4-1). Neither the northern or western range limit coincides with distinct geographical barriers, though they do occur along gradients of mean annual temperature and precipitation, respectively (Hijmans *et al.* 2005). Populations are distinct, vary in size from dozens to thousands of individuals and are typically located in habitats with low competition, such as abandoned agricultural fields, open woodlands, disturbed prairies and roadsides (Foote and Jackobs 1966, Irwin and Barneby 1982; JSG, personal observation). *C. fasciculata* is self-compatible, but highly out-crossing because of the buzz pollination mechanism that requires vibration from bees for pollen to dehisce from the anthers (Fenster 1991a). Seeds are dispersed explosively from the seedpods in the fall, and typically only move a few meters from the plant (Fenster 1991a).

In fall 2007, we collected seed from four populations; two within the interior of *C. fasciculata*'s range (InteriorMO: Cuivre River State Park, Troy, Missouri; InteriorIA: Conard Environmental Research Area, Grinnell College, Kellogg, Iowa), one at the western range edge (W.Edge: Konza Prairie Biological Station, Kansas State Univ., Manhattan, Kansas), and one at the northern range edge (N.Edge: Grey Cloud Dunes Scientific and Natural Area, Cottage Grove, Minnesota) (Fig 4-1., Table S4-1). All populations lie within natural areas and had been present in the area for at least 5 years prior to seed collection (personal comm. with site managers). At each site, seedpods were collected from fifty haphazardly selected maternal plants spaced at least 5 m apart. In January 2008, seeds were sterilized in 10% sodium hypochlorite, scarified with a metal file, and germinated on filter paper before transplanting to pots in the greenhouse. The top-most leaf was collected from 17 randomly selected flowering plants from each population and frozen at -80°C. DNA was extracted using a Qiagen DNeasy Plant Kit (Qiagen Inc, Valencia, CA) following the manufacturer's specifications.

Primers were developed using a *C. fasciculata* EST database obtained by whole transcriptome sequencing (Singer *et al.* 2009). We haphazardly selected 16 contigs that were between 600-900 bp in length and did not align to multiple copies by BLAST to the *Glycine max* and *Medicago truncatula* genomes. We used PCR to screen the 16 primers in one individual from each population. The successful PCR reactions were cleaned with Exosap (Affymetrix Inc, Santa Clara, CA) and directly sequenced at the University of Minnesota Biomedical Genomics Center on an ABI 3730xl capillary electrophoretic DNA analyzer. Sequence data was aligned, trimmed and heterozygous peak calls were

manually checked using Codon Code Aligner (CodonCode Corporation, Dedham, MA). We checked for problematic indels or other features that would prevent sequencing, and reduced our sample to 9 DNA fragments (Table S4-2 for primer pairs) for sequencing with the complete panel of 68 individuals.

In 7 of the 9 chosen fragments, introns present in the EST sequence resulted in amplified fragments considerably longer than 900 bp (Table S4-2). For these PCR products, we developed internal sequencing primers so as to sequence ~600bp in both the forward and reverse directions. Two of the DNA fragments contained indels. The length and locations of these were reconstructed by sequencing in both directions and manually aligning against sequences homozygous for the long allele. However, indel polymorphism is not readily incorporated into current analyses of sequence evolution, so we coded the indel sequence as missing data for all samples in further analysis. The sequences were phased to haplotypes using the program PHASE (Stephens *et al.* 2001) implemented in DnaSP (Librado and Rozas 2009), with 93% of haplotype pairs having posterior probabilities > 90%. The original sequences were deposited in GenBank at accession numbers XXXXX-XXXXX.

Data analyses

We calculated haplotype richness, haplotype diversity, and the mutation – scaled effective population size ($\theta = 4N_e\mu$) based on the average number of segregating sites per site (θ_W), and the average number of nucleotide differences per site ($\theta\pi$) for each locus and population using DNASP V5 (Librado and Rozas 2009). The number of unique

haplotypes and expected heterozygosity (H_e) were calculated in ARLEQUIN (Excoffier *et al.* 2005). Examination with CONTRIB (Petit *et al.* 1998) showed that our estimates of haplotype richness were not affected by different numbers of samples for each population due to missing data. We restricted our analysis of θ_W and $\theta\pi$ to silent sites (synonymous and introns) as these are more likely to be evolving neutrally. Because sequence diversity is heterogeneous across the genome, comparing average values across loci obscures potentially meaningful differences among populations. To account for this, we mean-transformed and standardized the values of haplotype richness, unique haplotypes, H_e, and θ_W and $\theta\pi$ for each locus. We used a non-parametric Kruskal-Wallis rank sum test (kruskal.test in R (R Development Core Team 2009)), which avoids making any assumptions about the underlying distribution of the data, to evaluate the probability that populations differed in the amount of diversity they contained, and performed post-hoc multiple comparison tests (kruskalmc in library pgirmess (Davis *et al.* 2005)) to determine which populations differed significantly.

To characterize the frequency distribution of polymorphic sites in each population, we calculated Tajima's *D* (Tajima 1989) based on sequence data and Fu's *Fs* (Fu 1997) based on haplotype data with DNASP. Tajima's *D* and Fu's *Fs* summarize the frequency distribution of genetic variants, with negative values indicative of an excess of rare variants potentially due to recent population expansion, and positive values suggestive of recently bottlenecked populations. We used Kruskal-Wallis tests as before to evaluate if the *D* and *Fs* differed among populations.

Population structure was examined in two ways. First, we used F_{ST} to evaluate the pairwise genetic divergence among populations. Confidence intervals for F_{ST} were calculated with 1000 random permutation in FSTAT (Goudet 1995). Second, we examined the assignment of individuals to groups based on their multilocus haplotypes using STRUCTURE v2.3.3 (Pritchard et al. 2000, Falush et al. 2003), which is appropriate for this species as it is primarily outcrossing (F_{IS} = 0.15). We used the correlated allele frequency (F) model with admixture, which assumes populations split from a common ancestor at the same time, and diverged through genetic drift at a rate inversely proportional to their effective population size (Falush *et al.* 2003). We ran 10 replicate runs for values of *K* from 1 to 5 for 200000 reps after a burn-in of 50000. To evaluate the most likely value of *K*, we used both the ad-hoc method of Pritchard (2000), which favors the model with the highest probability of the model given *K*, and the ΔK method of Evanno et al. (2005), implemented in the R library *corrsieve* (Campana *et al.* 2011), which favors the model with the greatest second-order change in ln Pr($X|K$).

F_{ST} and STRUCTURE reveal patterns of population structure but not the evolutionary process responsible for genetic structure (Beerli and Palczewski 2010). In order to gain further insight into gene flow among populations, we used Bayesian Markov Chain Monte Carlo coalescent models implemented in MIGRATE-N (Beerli 2010). MIGRATE-N gives estimates of the mutation-scaled effective population size ($\theta = 4N_e\mu$) and pairwise mutation-scaled migration rate between populations ($M = m/\mu$). Default settings were used for the DNA sequence model (Felsenstein 84 model of evolution, transition/transversion ratio = 2). We set uniform priors for θ [0, 0.05] and M

[0, 10000] divided into 3000 bins, and ran four chains with static heating (temperatures of 1, 1.5, 3, 10000) for a single long run 1×10^7 steps (sampling every 100 steps) with a burn-in of 1×10^4. To explicitly compare among models of population structure, we used the approximate marginal likelihood method of Beerli and Palczewski (2010). We fit five models to the data; a model with no population structure (Panmictic), a model with no migration among populations (No migration), a model with migration only from the interior to the edge populations (Edge only), a model with only migration northward (North only), and a final model assuming asymmetric migration among all populations (Full). We used the Bezier approximation score to calculate the Bayes Factor and select the most probable model from among these five models.

Results

Sequence diversity

Haplotype richness differed significantly among populations, with significantly more total haplotypes (Kruskal-Wallis $\chi^2_{3,32} = 16.9$, $P < 0.001$, Fig. 4-2a, Table S4-3) and unique haplotypes (Kruskal-Wallis $\chi^2_{3,32} = 14.1$, $P = 0.003$, Fig. 4-2b) in the InteriorMO population than in any of the other populations (Fig. 4-2a). Species-wide estimates of nucleotide diversity at silent sites were similar based on segregating sites ($\theta_W = 0.008$) and pair-wise nucleotide differences ($\theta\pi = 0.009$; Table S4-3). Among populations, $\theta\pi$ did not differ among any populations (Kruskal-Wallis $\chi^2_{3,32} = 4.0$, $P = 0.26$, Fig. 4-2c). However, θ_W was significantly greater in the InteriorMO population than the others (Kruskal-Wallis $\chi^2_{3,32} = 16.5$, $P < 0.001$, Fig. 4-2d), implying that differences were

largely due to rare variants in the InteriorMO population. H_e did not differ among populations (Kruskal-Wallis $\chi^2_{3,32} = 0.46$, P = 0.93), indicating that while there are fewer rare variants in edge populations, allele frequencies are more even (Fig. 4-3, Table S4-3).

Demographic history

Tajima's *D* and Fu's *Fs* had distributions that overlapped zero for all populations (Fig. 4-4, Table S4-4) and did not significantly differ from each other (Kruskal-Wallis $\chi^2_{3,32} = 0.81$, P = 0.84). However, there was a pattern for higher values in the edge populations than the interior populations (Fig. 4-4). This could be indicative of recent bottlenecks or population contraction in the edge populations. The high positive outliers and greater variance in *D* and *Fs* in the W.Edge population compared to the other population are consistent with population contraction and metapopulation dynamics (Wakeley and Aliacar 2001, Moeller et al. 2011).

Population structure

Overall, populations were significantly diverged from each other (mean $F_{st} = 0.16$, 99% confidence interval 0.057 – 0.258), mostly due to the W.Edge population being highly diverged from the remaining populations (Fig. 4-5, Table S4-5). STRUCTURE runs showed a peak in model likelihood at $K = 3$, with both the highest Pr $(X|K)$ and ΔK at this value (Table 3). With $K = 3$, the InteriorMO, W.Edge and N.Edge populations all contained different demes at high frequency (> 80% of individuals), and few individuals had coancestry coefficients <80% with the deme they were in (Fig. 4-1). In contrast,

individuals in the InteriorIA population had mixed coancestry coefficients that primarily grouped with either the InteriorMO (~45%) or N.Edge demes (~52%) (Fig. 4-1). This pattern, as well as low levels of F_{ST} among these populations, suggests northward range expansion from the InteriorMO population to InteriorIA and then the NEdge population (Fig. 4-1), with selection and drift at the expanding range edge causing population differentiation (Excoffier *et al.* 2009). Both F_{ST} and haplotype clustering showed that the W.Edge population was largely independent from the remaining populations (Fig. 4-1, 4-5), which were more similar to each other. This strong differentiation is not likely due to the peripheral status of this population, as it is similar to the InteriorIA population in levels of molecular diversity. A likely explanation is the W.Edge region was colonized at the same time as the InteriorMO region, and before the regions further north (InteriorIA, N.Edge) regions. However, the low molecular diversity (Fig. 4-2) and effective population size (Fig. 4-6) imply that the W.Edge population has been chronically smaller than the other populations. Thus, it was both more likely to have diverged through genetic drift and less likely to have contributed propagules to northward range expansion than the InteriorMO population.

Migration

The most probable MIGRATE-N coalescent model of population structure was the Full model assuming asymmetric migration among all populations (Table 4-2). From the Full model, migration rates were high among all populations, with estimates of *M* between 4500 – 7300 (Table S4-6). There was some evidence of asymmetric migration

rate towards the northern range edge, with greater migration into the InteriorIA and N.Edge populations than back into the more southern InteriorMO population (Fig. 4-6). Migration estimates into the W.Edge population were low, especially from the InteriorIA and N.Edge populations, consistent with evidence from population structure that this population is significantly diverged and has contributed less to northward range expansion.

Discussion

Theoretical models of species' range limits make testable predictions about the magnitude and structure of genetic diversity among populations at different geographic range locations. Population genetics can be used to evaluate which of these models are likely to be important for determining species' range limits (e.g. Moeller et al. 2011). In this study, we found that molecular diversity is lower in both northern and western range edge populations compared to an interior population of the annual native plant species, *C. fasciculata*. Population structure inferred from F_{ST} and haplotype clustering show considerable differentiation of the population at the western range edge from the remaining populations, which are less, but significantly, differentiated from each other (Fig. 4-1, 4-5). Thus, there is little evidence for recent migration among populations, though coalescent analyses did indicate that there was migration among populations on a deeper timescale. In a companion study, transplant experiments show that populations are not expected to persist beyond the range edge, even if dispersal is substantial (Chapter 1).

In light of these results we propose that both the western and northern range edges are limited by ecological-genetic processes (Antonovics 1976).

Population size and range location

Our molecular data support the "abundant center" model of species distributions (Andrewartha and Birch 1954, Brown 1984). Specifically, sequence and haplotype diversity were greater in the InteriorMO population compared to both the north and west edge populations (Fig. 4-2). The estimated effective population size (assuming standard neutral model, $Ne = \theta/4\mu$ with $\mu = 7^{-9}$,(Ossowski *et al.* 2010), with θ from the MIGRATE model) for the InteriorMO population (156000) is twice as great as for the InteriorIA population (76100), four times larger than the N.Edge population (38600) and ten times larger than the W.Edge population (16500; Fig. 4-6). These results are consistent with the 64% of studies that found a decrease in genetic diversity from range center to edge (Eckert *et al.* 2008). However, Eckert et al. (2008) focused on studies reporting H_e, which we did not find to differ among populations (Fig. 4-2). A potential reason for this is high gene flow (Fig. 4-6), which may elevate estimates of H_e. The use of more sensitive markers of diversity (i.e. θ_W) thus allowed us to detect a pattern that may otherwise have been missed.

The data also suggest that the InteriorIA population is more appropriately considered an intermediate-edge population. We originally selected the InteriorIA population to be an "interior" population based on geographic location (~1000km south of the northern range edge in Minnesota), large population size and individual condition

in the field (Chapter 1). However, measures of molecular diversity in this population are reduced compared to InteriorMO and more similar to the edge populations, suggesting that it has experienced similar population history. Further, haplotype structuring shows that this population has mixed coancestry with the InteriorMO and N.Edge populations. This pattern indicates that the InteriorIA population may be an intermediate step of ongoing northward range expansion (Fig. 4-1). This finding underscores how molecular data can provide novel insights on history of range expansion, independent of geographic location.

Implications for range limits

The W.Edge population appears to have experienced population contraction, as indicated by the reduced molecular diversity at all loci and absence of rare variants. One potential explanation for this reduction in diversity is recent arrival. However, given geographic location and historical patterns of species range shifts in this region (Williams et al. 2000), this explanation is unlikely. A second potential explanation is that there has been frequent extinction and recolonization of populations in this region. Field observations support these findings. Over three years of observing this population, it ranged in size from 500 to near 10000 flowering individuals, and other populations of *C. fasciculata* in the area had recently gone extinct (J. Etterson, pers. comm.). Further, as *C. fasciculata* is sensitive to reduced precipitation in this region (Craine *et al.* 2011), populations are likely to have increased extinction rates in dry years, as suggested by the small effective population size (Fig. 4-6). Thus, we conclude that frequent extinction

likely limits population growth, and low genetic diversity slows adaptation, such that populations fail to colonize conditions beyond this range edge. In fact, with greater aridity in this region due to climate change (Etterson 2004, Craine et al. 2011), it is likely that *C. fasciculata* is undergoing range retraction in this region as its potential to adapt to new climate conditions may limited by antagonistic genetic correlations (Etterson and Shaw 2001).

At the northern edge, we also found reduced molecular diversity, but evidence for northward range expansion from an interior population (Fig. 4-1, 5) suggesting that this region was more recently colonized, likely following warming since the Last Glacial Maximum. We emphasize that adaptation to local conditions (i.e. phenology) has occurred as well as range shifts. Specifically, in a greenhouse study using the genotypes sequenced, we found that the populations were significantly diverged in date of first flower ($F_{3,202} = 39.7$, $P < 0.0001$), with the N.Edge population flowering first and the InteriorIA populations intermediate to the two southern populations (Fig. S4-2). While this adaptation has enable populations to persist at the range edge, field experiments show that fitness is strongly reduced beyond the northern range edge, indicating that it has not adapted to novel conditions in this region (Chapter 1, 3). Given the history of past range expansion and adaptation (Fig. 4-1), it is unclear what prevents *C. fasciculata* from expanding its range edge by adaptation to novel environmental conditions. Competitors can cause range limitation with (Case and Taper 2000) or without (Price and Kirkpatrick 2009) gene flow by reducing population size, and thus adaptive potential. In a transplant study, estimates of population growth rate were greater with neighbors removed at the

northern range edge (Chapter 1), and second transplant study suggested that *C. fasciculata* may be competitively restricted to sub-optimal habitats at its range edge because competitors reduce seed production (Chapter 3). Thus, the adaptation of *C. fasciculata* to conditions at and beyond the range edge may be constrained by both climate and biotic factors preventing population growth in this region. Intriguingly, mutualism apparently also plays a role; as we found that appropriate rhizobia are nearly absent beyond *C. fasciculata*'s range edge than within its range, and that individual fitness is severely reduced in their absence (Chapter 2). This conclusion contrasts with the common expectation that climatic conditions, not biotic factors, will limit a species' pole-ward range edge.

Summary and prospects for future range expansion

Our study demonstrates that integration of findings from population genetic and ecological genetic studies yields a more complete understanding of the causes of species range limits can be achieved (van Heerwaarden et al. 2009, Moeller et al. 2011). We find that molecular diversity is reduced in both the northern and western range edges, but that gene flow is high between populations. Given field work revealing that population growth is below replacement beyond both these range edges, we conclude that extinction-recolonization dynamics and ecological-genetic constraints are likely to be key processes limiting both of these range edges. However, what limits the southern edge of this species is still an unexplored question.

Funding and Assistance

We thank the managers of Conard Environmental Research Area (Grinnell College), Konza Prairie Biological Station (Kansas State University), Cuivre River State Park (Missouri Dept. of Natural Resources) and Grey Cloud Dunes Scientific and Natural Area (Minnesota Dept. of Natural Resources) for permission to collect seed. Susan Singer generously provided early access to EST data. Jen Sellner assisted with sequencing, and Steve Keller and Dave Moeller provided discussion and comments.

Tables and Figures

Table 4-1. STRUCTURE results and model selection based on haplotype data for the number of populations (*K*) present in the data. Results are based on mean of 10 replicate runs.

K	Mean ln Pr(*X*|*K*)	SD ln Pr(*X*|*K*)	Pr (*K*)	Δ*K*
1	-1577.4	0.63	0.00	-
2	-1506.4	248.36	0.00	0.25
3	-1373.1	2.23	**1.00**	**119.23**
4	-1505.1	27.11	0.00	4.72

Table 4-2. MIGRATE-N results and model selection based on haplotype data for the number of populations (*K*) present in the data. Results are based on mean of 10 replicate runs.

Model	# Parameters	Bezier lmL	LBF	Model probability
Full	16	-9247.8	0	1.00
Edge only	9	-9341.8	-93.9	0.00
North only	8	-9700.9	-453.0	0.00
No migration	4	-10502.8	-1255.0	0.00
Panmictic	1	-9370.7	-122.9	0.00

Figure 4-1. Map of *C. fasciculata* study populations with inferred clustering to three genetic demes for each population from STRUCTURE. The population locations are marked with the yellow circles. For each population, every individual is represented by a single line, with the proportional assignment of the individual to each genetic deme given by the three colors. The dotted line represents the approximate range edge of *C. fasciculata* based maximum entropy species modeling of 1506 observations on current climate data (Supplemental Information).

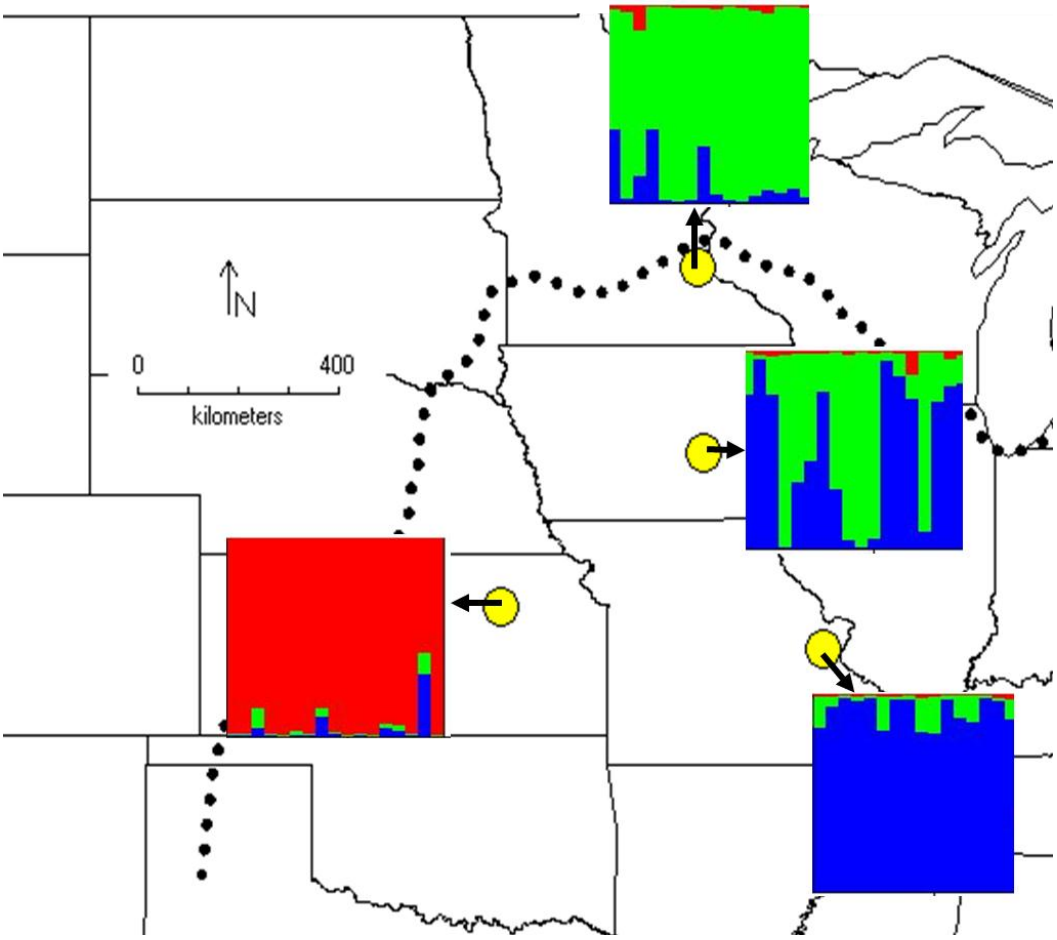


Figure 4-2. (a) Total haplotypes, (b) unique haplotypes, (c) pairwise nucleotide diversity ($\theta\pi$) and (d) segregating sites per site (θ_W) in the four study populations of *C. fasciculata*. Standard errors are given for estimators of θ but not haplotypes which are sums across loci.

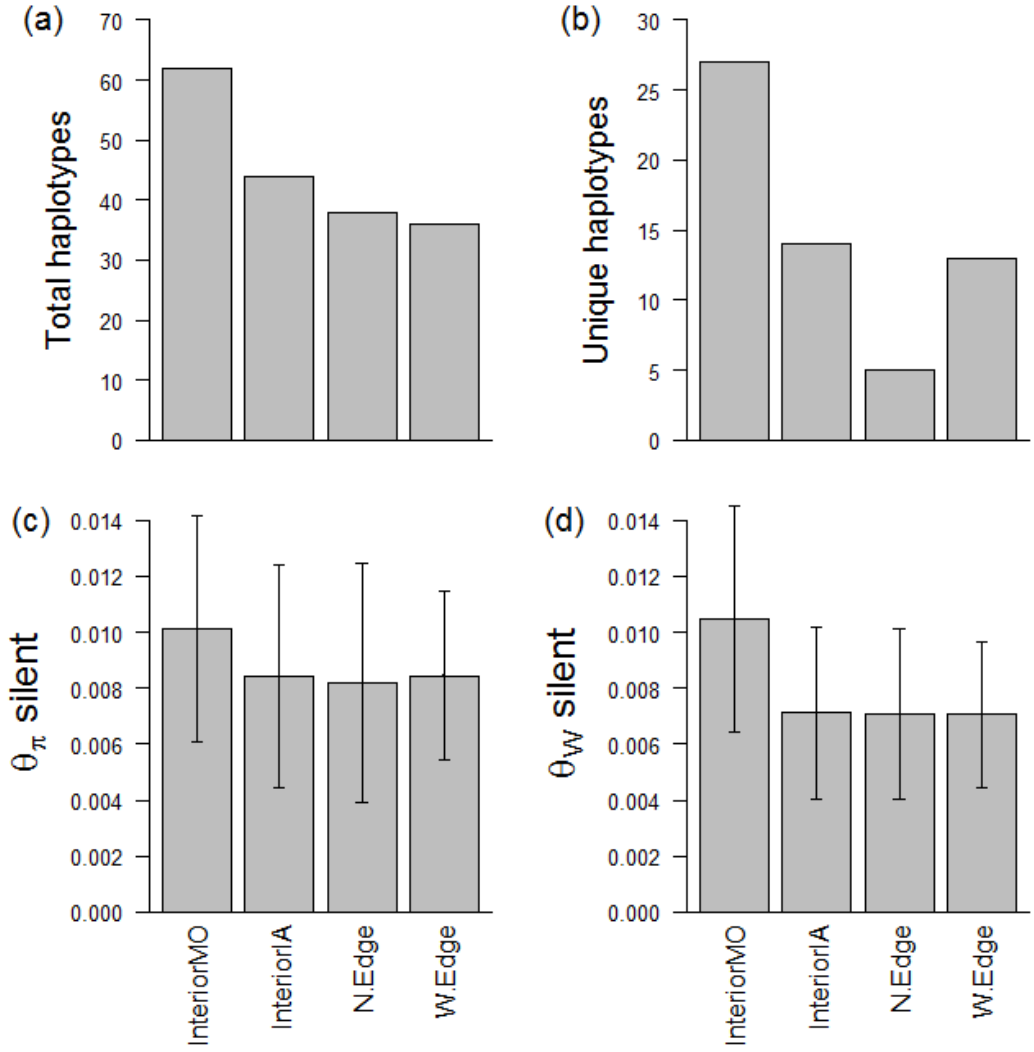


Figure 4-3. Box-and-whisker plot of the observed distribution of expected heterozygosity for 9 loci in the four study populations of *C. fasciculata*.

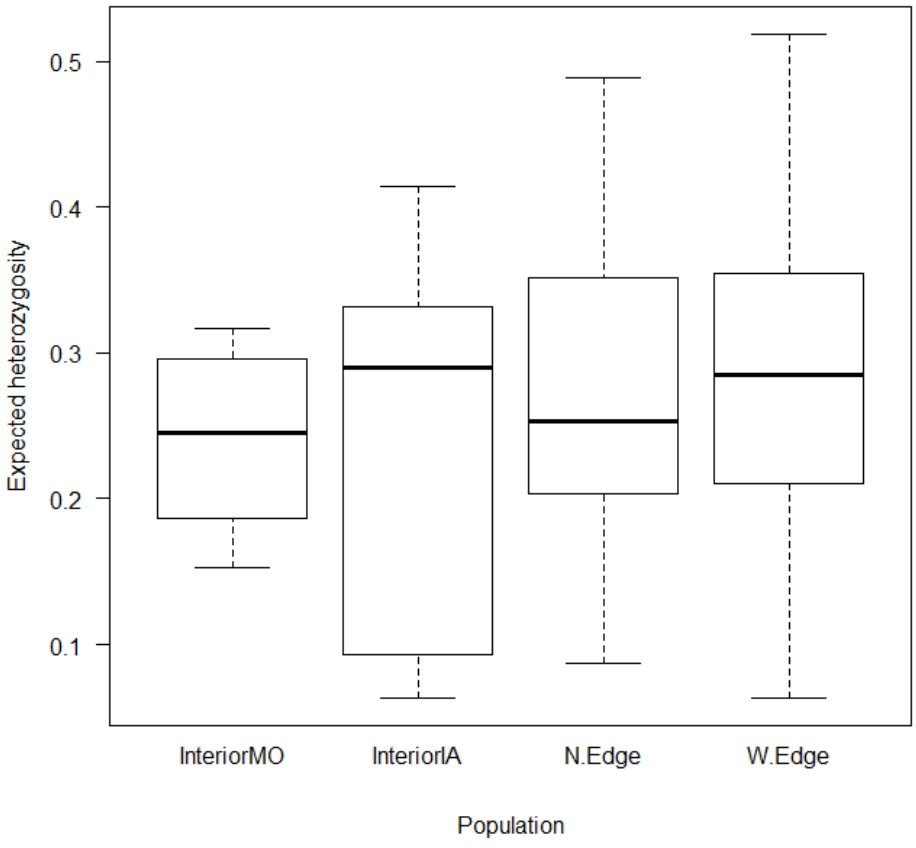


Figure 4-4. Box-and-whisker plot of the observed distribution of Tajima's *D* and Fu's *Fs* for 9 loci in the four study populations of *C. fasciculata*.

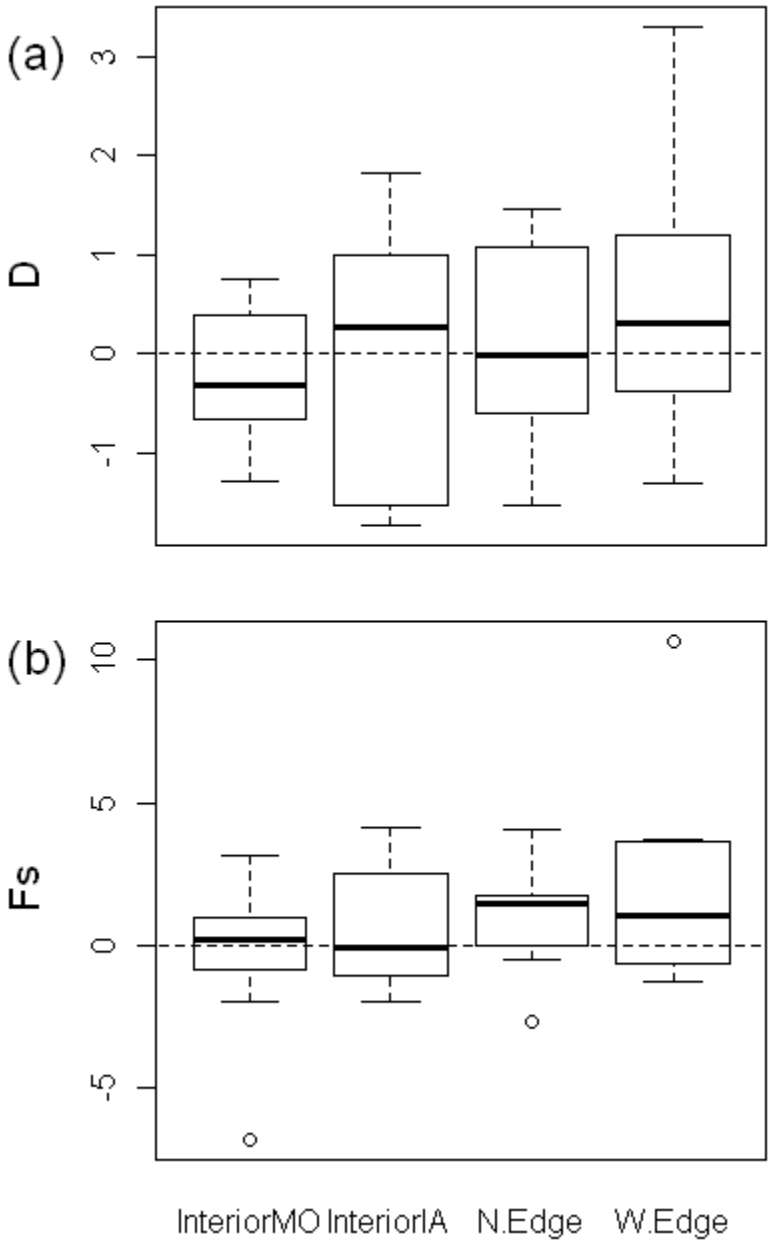


Figure 4-5. Pairwise F_{ST} values for the study populations of *C. fasciculata*.

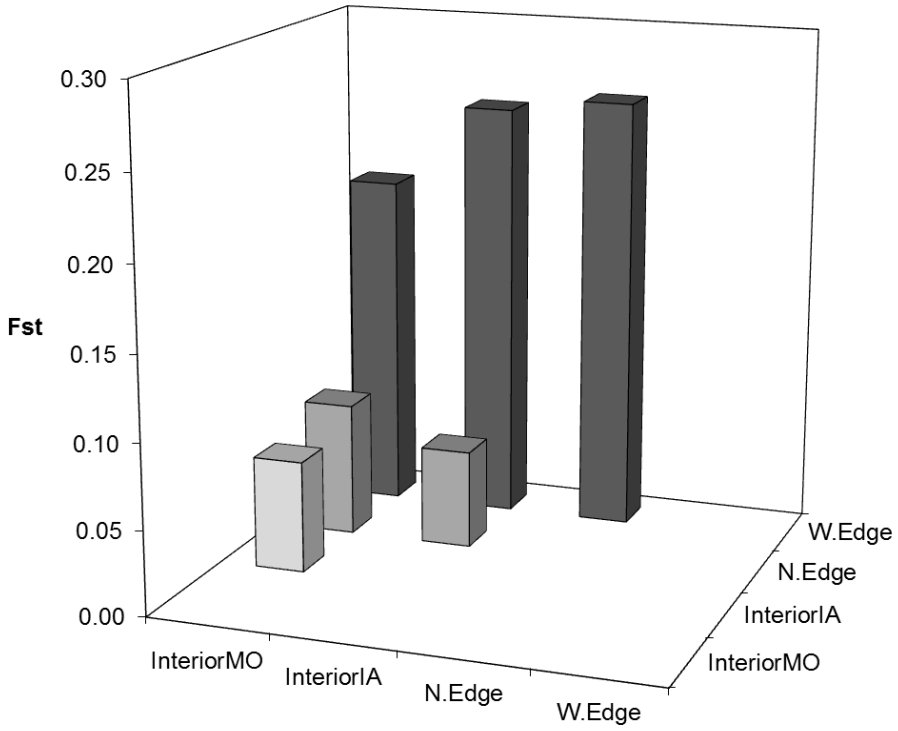
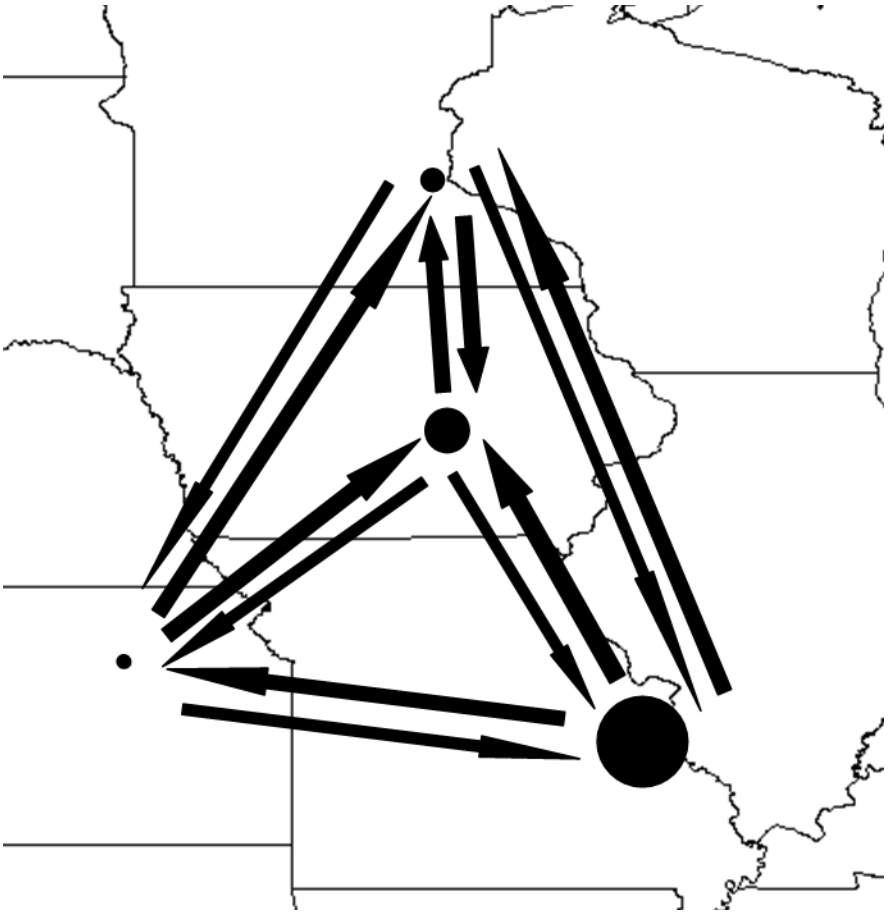


Figure 4-6. Map showing estimated migration rates (*M*) among populations and mutation-scaled effective population sizes (θ) from Full MIGRATE-N model. The thickness of the arrows is proportional to the migration rate, and the size of the circle representing each population is proportional to θ. Absolute values are given in Table S4-6.



Supplemental Information

Ecological niche modeling

C. fasciculata's potential distribution under current conditions was modeled with the program MAXENT (PHILLIPS ET AL. 2006B), using 1506 observations downloaded from the Global Biodiversity Information Facility (http://data.gbif.org). To train the model, we used 75% of the data, with 25% of the data used to test the models. Climate data were downloaded from the WorldClim database (http://worldclim.org) using interpolations of 1950 - 2000 weather data for current climate (Hijmans *et al.* 2005). Model validation was performed using area under the receiver operating characteristic curve (AUC), with observed values of 0.925/0.923 for the training and test data respectively, suggesting excellent fit of the model to the data (Araújo *et al.* 2005). Mapping was done using DIVA-GIS (Hijmans *et al.* 2004).

Table S4-1. *C. fasciculata* population source information. Mean annual temperature (MAT℃) and annual precipitation (PPT mm) were collected from the WorldClim data set implemented in DIVA-GIS (Hijmans *et al*. 2004).

Population	Location	Lat/Long	MAT	PPT
InteriorMO	Cuivre River State Park Troy, MO	39°01'00N 90°55'16W	12.3	963
InteriorIA	Conard Env. Research Area (Grinnell College), Kellogg, IA	41°41'14N 92°52'16W	8.9	882
N.Edge	Grey Cloud Dunes SNA Cottage Grove, MN	44°47'19N 92°57'29W	7.5	754
W.Edge	Konza Prairie (KSU), Manhattan, KS	39°07'07N 96°32'15W	11.9	876

Table S4-2. Polymerase chain reaction (PCR) primers for the nine loci amplified. PCR conditions were denaturing at 95° for 2 minutes, annealing at 56-58° for one minute, extension at 72° for 2 minutes, repeated 34 times. For DNA fragments longer than 900 bp, a separate primer was used for sequencing, either in the forward or reverse direction.

Locus	Forward primer	Reverse primer	Approx. length (bp)
cf1044	5'-TTTGGGTCTGTGGGCTCTGG	5'-CCAGGAAATCCACCTTGTGC	1000
cf1184	5'-TGAGCTTCTACAAGGTCCC	5'-ACCTGCACCAACTGTTCTCC	1500
cf10036	5'-TATGACTATCAGTTAGTGG	5'-AGGAATATCAGTGCCCTCAGC	1500
cf10104	5'-ATCTCTGAAATGACAACGC	5'-AATGTCGCTATCGACAGACG	1200
cf10117	5'-ACTCTGTCTAGCTCCC	5'-TGCTTGAAACCAAACCTGCC	2000
cf10463	5'-AGGATGTTGGAGAAGCGACC	5'-ATTACCTGGTAACTCTAGC	700
cf10639	5'-TTCGTCCTCGCAACAACCC	5'-ACAGTCTGTCTAGTCTCTGG	1600
cf11490	5'-AACCATTCAATGCTGAACCG	5'-TCCCAATTCACAGATGGTGG	1200
cf11621	5'-TGTTCCCTTAGACGTATGCC	5'-ATCTCGTGTCTTGGTTCTCC	600

Locus	Internal sequencing primer	Direction	Indel?
cf1044	5'-TTCGGTGCTGGACGAGG	F	
cf1184	5'-ATCACAGCATTAAGACC	R	
cf10036	5'-ACTCATCTCCTCAATGCC	R	10bp
cf10104	5'-TAGAGTGATCGAGTCG	R	
cf10117	5'-CATATAAGTGAAGGG	R	
cf10463	na		
cf10639	5'-ACCCTAAGACTTTACCAC	R	33bp
cf11490	5'-ACCAGTGTCTAGTAGCC	R	
cf11621	na		

Table S4-3. Molecular diversity at nine nuclear loci for four populations. N* is the number of samples, H is the number of haplotypes, S is the number of segregating sites, H div is haplotype diversity, and He is expected heterozygosity.

Pop	Locus	Length (bp)	N*	H	S	H div	*θw* total	*θw* silent
InteriorMO	**cf10036**	710	32	5	7	0.629	0.00248	0.00309
	cf10104	459	28	18	22	0.96	0.01288	0.03955
	cf10117	590	24	7	18	0.605	0.00817	0.01563
	cf1044	556	30	7	10	0.798	0.00454	0.00448
	cf10463	525	32	5	5	0.639	0.00236	0.00678
	cf10639	585	28	8	11	0.762	0.0051	0.00659
	cf11490	762	32	3	5	0.486	0.00163	0.00213
	cf11621	510	32	6	7	0.73	0.00341	0.01509
	cf1184	713	30	3	2	0.297	0.00071	0.00091
InteriorIA	**cf10036**	710	34	4	5	0.677	0.00175	0.00217
	cf10104	459	28	13	15	0.934	0.0084	0.02965
	cf10117	590	32	2	3	0.063	0.00126	0.00207
	cf1044	556	32	4	10	0.587	0.00447	0.00441
	cf10463	525	28	3	3	0.204	0.00147	0.00467
	cf10639	585	20	5	5	0.716	0.00254	0.00401
	cf11490	762	30	4	5	0.628	0.00166	0.00217
	cf11621	510	26	6	6	0.738	0.00308	0.01365
	cf1184	713	32	3	3	0.123	0.00104	0.00134
N.Edge	**cf10036**	710	30	4	5	0.522	0.0018	0.00224
	cf10104	459	32	8	14	0.865	0.00811	0.02865
	cf10117	590	32	4	14	0.383	0.00589	0.01139
	cf1044	556	32	7	10	0.815	0.00447	0.00441
	cf10463	525	30	4	3	0.193	0.00144	0.00459
	cf10639	585	24	2	1	0.489	0.00048	0.00076
	cf11490	762	28	2	1	0.254	0.00034	0.00044
	cf11621	510	32	5	5	0.685	0.00243	0.01078
	cf1184	713	32	2	1	0.121	0.00035	0.00045
W.Edge	**cf10036**	710	32	4	3	0.381	0.00106	0.00132
	cf10104	459	26	10	14	0.871	0.00799	0.02527
	cf10117	590	26	3	12	0.588	0.00533	0.00983
	cf1044	556	30	5	9	0.607	0.00409	0.00448
	cf10463	525	32	3	4	0.232	0.00189	0.00678
	cf10639	585	28	2	3	0.349	0.0014	0.00221
	cf11490	762	34	3	5	0.64	0.0016	0.0021
	cf11621	510	28	4	5	0.648	0.00252	0.01116
	cf1184	713	32	2	1	0.063	0.00035	0.00045

Pop	Locus	Length (bp)	N*	H	S	H div	*θw* total	*θw* silent
species-wide	**cf10036**	710	128	6	7	0.724	0.00184	0.00229
	cf10104	459	114	35	26	0.939	0.01108	0.03444
	cf10117	590	114	9	18	0.436	0.005747	0.01099
	cf1044	556	124	14	15	0.837	0.005002	0.00512
	cf10463	525	122	5	5	0.347	0.001771	0.0009
	cf10639	585	100	12	15	0.802	0.005249	0.00719
	cf11490	762	124	5	6	0.55	0.00146	0.00191
	cf11621	510	118	7	7	0.719	0.002569	0.01137
	cf1184	713	126	6	5	0.152	0.001296	0.00167

Table S4-3 continued.

Pop	Locus	*θπ* total	θπ silent	He
InteriorMO	**cf10036**	0.00217	0.0027	0.217
	cf10104	0.01173	0.03848	0.245
	cf10117	0.00904	0.01686	0.296
	cf1044	0.00559	0.0054	0.311
	cf10463	0.00178	0.00453	0.187
	cf10639	0.00309	0.00411	0.156
	cf11490	0.00208	0.00273	0.317
	cf11621	0.00356	0.01577	0.260
	cf1184	0.00043	0.00055	0.153
InteriorIA	**cf10036**	0.00295	0.00367	0.414
	cf10104	0.01084	0.03839	0.332
	cf10117	0.00032	0.00052	0.063
	cf1044	0.00521	0.00527	0.29
	cf10463	0.00053	0.0019	0.093
	cf10639	0.0027	0.00426	0.299
	cf11490	0.00269	0.00353	0.411
	cf11621	0.00337	0.01491	0.286
	cf1184	0.00026	0.0034	0.063
N.Edge	**cf10036**	0.00209	0.0026	0.293
	cf10104	0.01072	0.04065	0.352
	cf10117	0.00483	0.00904	0.203
	cf1044	0.00656	0.00685	0.365
	cf10463	0.0005	0.00178	0.087
	cf10639	0.00088	0.00139	0.489

Pop	Locus	$\theta\pi$ total	θπ silent	Pop
	cf11490	0.00033	0.00044	0.253
	cf11621	0.00242	0.01073	0.247
	cf1184	0.00017	0.00022	0.121
W.Edge	**cf10036**	0.00088	0.0011	0.21
	cf10104	0.00868	0.02624	0.285
	cf10117	0.01056	0.01946	0.519
	cf1044	0.00655	0.00718	0.405
	cf10463	0.00091	0.00463	0.119
	cf10639	0.0019	0.003	0.349
	cf11490	0.00233	0.00305	0.355
	cf11621	0.00255	0.01128	0.26
	cf1184	0.00009	0.00011	0.063
species-wide	**cf10036**	0.00286	0.00355	
	cf10104	0.010952	0.03763	
	cf10117	0.006959	0.0129	
	cf1044	0.006439	0.00666	
	cf10463	0.000974	0.00049	
	cf10639	0.003304	0.005	
	cf11490	0.002052	0.00269	
	cf11621	0.003062	0.01356	
	cf1184	0.000241	0.00031	

Table S4-4. Tajima's *D* and Fu's *Fs* calculated for each population and locus.

	InteriorMO		InteriorIA		N.Edge		W.Edge		Overall	
Locus	***D***	***Fs***	***D***	***Fs***	***D***	***Fs***	***D***	***Fs***	***D***	***Fs***
1044	0.733	0.989	0.523	4.150	1.464	1.726	1.874	3.701	0.7733	-0.16
1184	-0.808	-0.863	-1.730	-1.708	-0.783	-0.495	-1.142	-1.265	-1.671	-6.451
10036	-0.362	0.586	1.827	2.775	0.437	1.431	-0.391	-0.672	1.241	2.427
10104	-0.319	-6.819	0.992	-1.984	1.066	2.101	0.296	-0.799	-0.034	-14.42
10117	0.384	2.493	-1.730	0.071	-0.596	4.055	3.301	10.656	0.592	3.658
10463	-0.666	-0.742	-1.527	-1.059	-1.539	-2.716	-1.315	0.006	-0.928	-1.263
10639	-1.286	-1.967	0.185	-0.088	1.391	1.462	0.862	3.439	-1.026	-2.634
11490	0.746	3.123	1.709	2.541	-0.019	0.448	1.195	3.647	0.881	2.262
11621	0.131	0.200	0.272	-0.272	-0.012	0.005	0.030	1.011	0.439	0.501
Mean	-0.161	-0.333	0.058	0.492	0.157	0.891	0.523	2.192	0.030	-1.787
StdDev	0.708	2.915	1.410	2.158	1.032	1.883	1.474	3.756	1.010	5.635

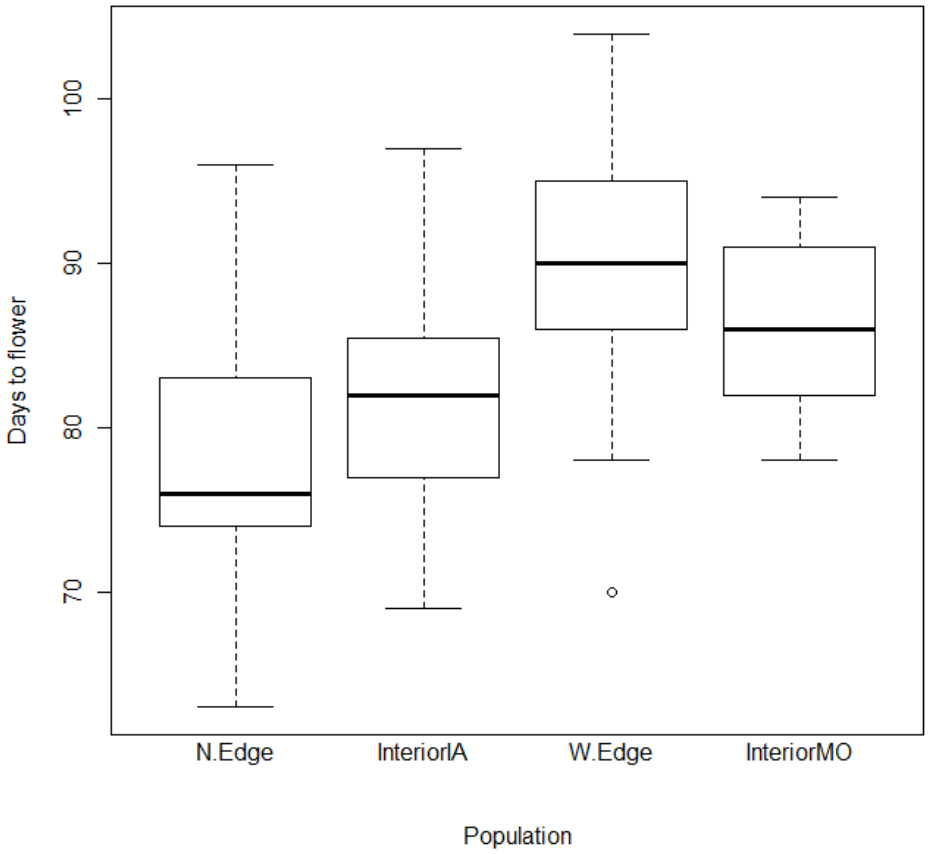
Table S4-5. Pairwise population F_{ST} and 95% confidence intervals calculated from 1000 random permutation in FSTAT (Goudet 1995).

95% confidence interval				
	InteriorMO	InteriorMO	N.Edge	W.Edge
InteriorMO	0			
InteriorIA	0.022 - 0.117	0		
N.Edge	0.041 - 0.126	0.014 - 0.11	0	
W.Edge	0.097 - 0.298	0.101 - 0.369	0.091 - 0.401	0

Table S4-6. Results from full MIGRATE-N including source and recipient population, the mode of the posterior distribution of the migration parameter *M* and bounds of 95% confidence intervals, the recipient population's θ and *4Nm* (product of *M* and θ).

Source	**Recipient**	**M mode**	**M 2.5%**	**M 97.5%**	**Recipient θ**	**4*Nm***
InteriorIA	InteriorMO	4428	2673	6697	0.00437	19.4
N.Edge	InteriorMO	4422	2800	7010	0.00437	19.3
W.Edge	InteriorMO	4528	2197	6980	0.00437	19.8
	avg.	4459				19.5
InteriorMO	InteriorIA	7585	5067	9610	0.00213	16.2
N.Edge	InteriorIA	6498	3397	8323	0.00213	13.8
W.Edge	InteriorIA	6515	3490	8790	0.00213	13.9
	avg.	6866				14.6
InteriorMO	N.Edge	6425	4387	9563	0.00108	6.9
InteriorIA	N.Edge	6462	4303	9497	0.00108	7.0
W.Edge	N.Edge	6318	3947	8663	0.00108	6.8
	avg.	6401.7				6.9
InteriorMO	W.Edge	5702	3160	8677	0.00046	2.6
InteriorIA	W.Edge	4765	2797	7077	0.00046	2.2
N.Edge	W.Edge	4172	2510	7183	0.00046	1.9
	avg.	4880				2.2

Figure S4-1. Distribution of date to first flower for the 17 genotypes from each population that were sequenced.



Bibliography

Ackerly, D. D. 2003. Community assembly, niche conservatism, and adaptive evolution in changing environments. International Journal of Plant Sciences **164 (3 Suppl.)**:S165-S184.

Ackerly, D. D., C. A. Knight, S. B. Weiss, K. Barton, and K. P. Starmer. 2002. Leaf size, specific leaf area and microhabitat distribution of chaparral woody plants: contrasting patterns in species level and community level analyses. Oecologia **130**:449-457.

Alexander, H. M., S. Price, R. Houser, D. Finch, and M. Tourtellot. 2007. Is there reduction in disease and pre-dispersal seed predation at the border of a host plant's range? Field and herbarium studies of Carex blanda. Journal of Ecology **95**:446-457.

Andrewartha, H. G. and L. C. Birch. 1954. The Distribution and Abundance of Animals. University of Chicago Press, Chicago, IL.

Angert, A. and D. Schemske. 2005. The evolution of species' distributions: reciprocal transplants across the elevation ranges of *Mimulus cardinalis* and *M. lewisii*. Evolution **59**:1671-1684.

Angert, A. L. 2006. Demography of central and marginal populations of monkeyflowers (*Mimulus cardinalis* and *M. lewisii*). Ecology **87**:2014-2025.

Angert, A. L., H. D. Bradshaw Jr, and D. W. Schemske. 2008. Using experimental evolution to investigate geographic range limits in monkeyflowers. Evolution **62**:2660-2675.

Antonovics, J. 1976. The nature of limits to natural selection. Annals of the Missouri Botanical Garden **63**:224-247.

Antonovics, J. 2009. The effect of sterilizing diseases on host abundance and distribution along environmental gradients. Proceedings of the Royal Society B: Biological Sciences **276**:1443-1448.

Antonovics, J. and A. D. Bradshaw. 1970. Evolution in closely adjacent plant populations. VIII. Clinal patterns at a mine boundary. Heredity **25**:349-362.

Araújo, M. B., R. G. Pearson, W. Thuiller, and M. Erhard. 2005. Validation of species - climate impact models under climage change. Global Change Biology **11**:1504-1513.

Baack, E. J., N. C. Emery, and M. L. Stanton. 2006. Ecological factors limiting the distribution of *Gilia tricolor* in a California grassland mosaic. Ecology **87**:2736-2745.

Barton, N. 2001. Adaptation at the edge of a species' range. Pages 365-392 *in* J. Silvertown and J. Antonovics, editors. Integrating Ecology and Evolution in a Spatial Context. Blackwell Science Ltd, Oxford, England.

Beerli, P. 2010. Migrate version 3.2 - a maximum likelihood and Bayesian estimator of gene flow using the coalescent.

Beerli, P. and M. Palczewski. 2010. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. Genetics **185**:313-326.

Béna, G., A. Lyet, T. Huguet, and I. Olivieri. 2005. *Medicago* - *Sinorhizobium* symbiotic specificity evolution and the geographic expansion of *Medicago*. Journal of Evolutionary Biology **18**:1547-1558.

Bischoff, A., L. Crémieux, M. Smilauerova, C. S. Lawson, S. R. Mortimer, J. Dolezal, V. Lanta, A. R. Edwards, A. J. Brook, M. Macel, J. A. N. Leps, T. Steinger, and H. Müller-Schärer. 2006. Detecting local adaptation in widespread grassland species - the importance of scale and local plant community. Journal of Ecology **94**:1130-1142.

Bissett, A., A. E. Richardson, G. Baker, S. Wakelin, and P. H. Thrall. 2010. Life history determines biogeographical patterns of soil bacterial communities over multiple spatial scales. Molecular Ecology **19**:4315-4327.

Both, C., S. Bouwhuis, C. M. Lessells, and M. E. Visser. 2006. Climate change and population declines in a long-distance migratory bird. Nature **441**:81-83.

Bridle, J. R. and T. H. Vines. 2007. Limits to evolution at range margins: when and why does adaptation fail? Trends in Ecology & Evolution **22**:140-147.

Brooker, R. W., F. T. Maestre, R. M. Callaway, C. L. Lortie, L. A. Cavieres, G. Kunstler, P. Liancourt, K. TielbÖRger, J. M. J. Travis, F. Anthelme, C. Armas, L. Coll, E. Corcket, S. Delzon, E. Forey, Z. Kikvidze, J. Olofsson, F. Pugnaire, C. L. Quiroz, P. Saccone, K. Schiffers, M. Seifan, B. Touzard, and R. Michalet. 2008. Facilitation in plant communities: the past, the present, and the future. Journal of Ecology **96**:18-34.

Brown, J. H. 1984. On the relationship between abundance and distribution of a species. American Naturalist **124**:255-279.

Burdon, J., A. Gibson, S. D. Searle, M. Woods, and J. Brockwell. 1999. Variation in the effectiveness of symbiotic associations between native rhizobia and temperate Australian Acacia: within species interactions. Journal of Applied Ecology **36**:398-408.

Bushnell, O. and W. Sarles. 1937. Studies on the root-nodule bacteria of wild leguminous plants in Wisconsin. Soil Science **44**:409-423.

Caldwell, M. M., T. E. Dawson, and J. H. Richards. 1998. Hydraulic lift: consequences of water efflux from the roots of plants. Oecologia **113**:151-161.

Callaway, R. M., R. W. Brooker, P. Choler, Z. Kikvidze, C. J. Lortie, R. Michalet, L. Paolini, F. I. Pugnaire, B. Newingham, E. T. Aschehoug, C. Armas, D. Kikodze, and B. J. Cook. 2002. Positive interactions among plants increase with stress. Nature **417**:844-848.

Campana, M. G., H. V. Hunt, H. Jones, and J. White. 2011. CorrSieve: software for summarizing and evaluating Structure output. Molecular ecology resources **11**:349-352.

Carter, R. N. and S. D. Prince. 1985. The geographical distribution of prickly lettuce (*Lactuca serriola*): I. A general survey of its habitats and performance in Britain. Journal of Ecology **73**:27-38.

Case, T. and M. Taper. 2000. Interspecific competition, environmental gradients, gene flow, and the coevolution of species' borders. American Naturalist **155**:583-605.
Case, T. J., R. D. Holt, M. A. McPeek, and T. H. Keitt. 2005. The community context of species' borders: ecological and evolutionary perspectives. Oikos **108**:28-46.
Caughley, G., D. Grice, R. Barker, and B. Brown. 1988. The edge of the range. Journal of Animal Ecology **57**:771-785.
Cornelissen, J. H. C., S. Lavorel, E. Garnier, S. Díaz, N. Buchman, D. E. Gurvich, P. B. Reich, H. ter Steege, H. D. Morgan, M. G. A. van der Heijden, J. G. Pausas, and H. Poorter. 2003. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. Australian Journal of Botany **51**:335-380.
Craine, J. M., J. B. Nippert, E. G. Towne, S. Tucker, S. W. Kembel, A. Skibbe, and K. K. McLauchlan. 2011. Functional consequences of climate change-induced plant species loss in a tallgrass prairie. Oecologia **165**:1109-1117.
Crawley, M. J. 2007. The R Book. John Wiley & Sons, Chichester, England.
Crémieux, L., A. Bischoff, M. Šmilauerová, C. S. Lawson, S. R. Mortimer, J. Dolezal, V. Lanta, A. R. Edwards, A. J. Brook, T. Tscheulin, M. Macel, J. Leps, H. Muller-Scharer, and T. Steinger. 2008. Potential contribution of natural enemies to patterns of local adaptation in plants. New Phytologist **180**:524-533.
Darwin, C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London.
Davis, M. B. and R. G. Shaw. 2001. Range shifts and adaptive responses to Quaternary climate change. Science **292**:673-679.
Davis, M. B., R. G. Shaw, and J. R. Etterson. 2005. Evolutionary response to changing climate. Ecology **86**:1704-1714.
Doyle, J. J. and M. A. Luckow. 2003. The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiology **131**:900-910.
Eckert, C. G., K. E. Samis, and S. C. Lougheed. 2008. Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. Molecular Ecology **17**:1170-1188.
Eckhart, V. M., M. A. Geber, W. F. Morris, E. S. Fabio, P. Tiffin, and D. A. Moeller. 2011. The geography of demography: long-term demographic studies and species distribution models reveal a species border limited by adaptation. American Naturalist **in press**.
Eckhart, V. M., I. Singh, A. M. Louthan, A. J. Keledjian, A. Chu, D. A. Moeller, and M. A. Geber. 2010. Plant - soil water relations and species border of *Clarkia xantiana* ssp. *xantiana* (Onagraceae). International Journal of Plant Sciences **171**:749-760.
Ehlers, B. K. and J. Thompson. 2004. Do co-occurring plant species adapt to one another? The response of *Bromus erectus* to the presence of different *Thymus vulgaris* chemotypes. Oecologia **141**:511-518.
Etterson, J. R. 2004. Evolutionary potential of *Chamaecrista fasciculata* in relation to climate change. I. Clinal patterns of selection along an environmental gradient in the Great Plains. Evolution **58**:1446-1458.

Etterson, J. R. and R. G. Shaw. 2001. Constraint to adaptive evolution in response to global warming. Science **294**:151-154.
Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology **14**:2611-2620.
Excoffier, L., M. Foll, and R. J. Petit. 2009. Genetic consequences of range expansions. Annual Review of Ecology, Evolution, and Systematics **40**:481-501.
Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin ver 3.0: an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online **2**:47-50.
Excoffier, L. and N. Ray. 2008. Surfing during population expansions promotes genetic revolutions and structuration. Trends in Ecology & Evolution **23**:347-351.
Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164**:1567-1587.
Fargione, J., C. S. Brown, and D. Tilman. 2003. Community assembly and invasion: An experimental test of neutral versus niche processes. Proceedings of the National Academy of Sciences **100**:8916-8920.
Fenster, C. 1991a. Gene flow in *Chamaecrista fasciculata* (Leguminosae) I. Gene dispersal. Evolution **45**:398-409.
Fenster, C. B. 1991b. Gene flow in *Chamaecrista fasciculata* (Leguminosae) II. Gene establishment. Evolution:410-422.
Fine, P. V. A., Z. J. Miller, I. Mesones, S. Irazuzta, H. M. Appel, M. H. H. Stevens, I. Sääksjärvi, J. C. Schultz, and P. D. Coley. 2006. The growth–defense trade-off and habitat specialization by plants in Amazonian forests. Ecology **87 (7 Suppl)**:S150-S162.
Foote, L. E. and J. A. Jackobs. 1966. Soil factors and the occurrence of partridge pea (*Cassia fasciculata* Michx.) in Illinois. Ecology **47**:968-975.
Fu, Y.-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics **147**:915-925.
Galloway, L. F. and C. Fenster. 2000. Population differentiation in an annual legume: local adaptation. Evolution **54**:1173-1181.
Gaston, K. J. 2003. The Structure and Dynamics of Geographic Ranges. Oxford University Press, Oxford, UK.
Gaston, K. J. 2009. Geographic range limits: achieving synthesis. Proceedings of the Royal Society B: Biological Sciences **276**:1395-1406.
Geber, M. A. and V. M. Eckhart. 2005. Experimental studies of adaptation in *Clarkia xantiana*. II. Fitness variation across a subspecies border. Evolution **59**:521-533.
Geyer, C. J. and R. G. Shaw. 2010. Aster models and Lande-Arnold beta, Technical Report 675. Sch. of Statistics, Univ. of Minnesota.
Geyer, C. J., S. Wagenius, and R. G. Shaw. 2007. Aster models for life history analysis. Biometrika **94**:415-426.

Gilman, S. E. 2006. The northern geographic limit of the intertidal limpet *Collisella scabra*: a test of performance, recruitment, and temperature hypotheses. Ecography **29**:709-720.
Goudet, J. 1995. Fstat version 1.2: a computer program to calculate F statistics. Journal of Heredity **86**:485-486.
Griffith, T. M. and M. A. Watson. 2005. Stress avoidance in a common annual: reproductive timing is important for local adaptation and geographic distribution. Journal of Evolutionary Biology **18**:1601-1612.
Griffith, T. M. and M. A. Watson. 2006. Is evolution necessary for range expansion? Manipulating reproductive timing of a weedy annual transplanted beyond its range. American Naturalist **167**:153-164.
Haldane, J. B. 1956. The relation between density regulation and natural selection. Proceedings of the Royal Society of London, Series B: Biological Sciences **145**:306-308.
He, F. and S. P. Hubbell. 2011. Species-area relationships always overestimate extinction rates from habitat loss. Nature **473**:368-371.
Heath, K. D. 2010. Intergenomic epistasis and coevolutionary constraint in plants and rhizobia. Evolution **64**:1446-1458.
Heath, K. D. and P. Tiffin. 2007. Context dependence in the coevolution of plant and rhizobial mutualists. Proceedings of the Royal Society B: Biological Sciences **274**:1905.
Henne, P. D., F. S. Hu, and D. T. Cleland. 2007. Lake-effect snow as the dominant control of mesic-forest distribution in Michigan, USA. Journal of Ecology **95**:517-529.
Hereford, J. 2009. A quantitative survey of local adaptation and fitness trade-offs. American Naturalist **173**:579-588.
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. International Journal of Climatology **25**:1965-1978.
Hijmans, R. J., C. Guarino, P. Bussink, M. Marthur, I. Cruz, I. Barrentes, and E. Rojas. 2004. A geographic information system for analysis of species distribution data. See www.diva-gis.org.
Hochberg, M. E. and A. R. Ives. 1999. Can natural enemies enforce geographical range limits? Ecography **22**:268-276.
Hoffman, A. A. and M. W. Blows. 1994. Species borders: ecological and evolutionary perspectives. Trends in Ecology & Evolution **9**:223-227.
Holt, R. D. 2003. On the evolutionary ecology of species' ranges. Evolutionary Ecology Research **5**:159-178.
Holt, R. D. and M. Barfield. 2009. Trophic interactions and range limits: the diverse roles of predation. Proceedings of the Royal Society B: Biological Sciences **276**:1435-1442.

Holt, R. D. and R. Gomulkiewicz. 1997. How does immigration influence local adaptation? A reexamination of a familiar paradigm. American Naturalist **149**:563-572.
Holt, R. D. and T. H. Keitt. 2000. Alternative causes for range limits: a metapopulation perspective. Ecology Letters **3**:41-47.
Huntley, B. 1991. How plants respond to climate change: migration rates, individualism and the consequences for plant communities. Annals of Botany **67**:15-22.
Irwin, H. S. and R. C. Barneby. 1982. The American Cassiinae. A synoptical revision of the Leguminosae tribe Cassieae subtribe Cassiinae in the New World. New York Botanical Garden Press, New York, NY.
Jackson, S. T., J. L. Betancourt, R. K. Booth, and S. T. Gray. 2009. Ecology and the ratchet of events: Climate variability, niche dimensions, and species distributions. Proceedings of the National Academy of Sciences **106**:19685-19692.
Jenkins, N. L. and A. A. Hoffman. 1999. Limits to the southern border of *Drosophia serrata*: cold resistance, heritable variation, and trade-offs. Evolution **53**:1823-1834.
Jump, A. S., C. Matyas, and J. Penuelas. 2009. The altitude-for-latitude disparity in the range retractions of woody species. Trends in Ecology & Evolution **24**:694-701.
Kartesz, J. T. and Biota of North America Program (BONAP). 2011. North American Plant Atlas (http://www.bonap.org). Chapel Hill, N.C.
Kavanagh, K. and M. Kellman. 1986. Performance of *Tsuga canadensis* (L.) Carr. at the centre and northern edge of its range: a comparison. Journal of Biogeography **13**:145-157.
Keller, S. R., M. S. Olson, S. Silim, W. Schroeder, and P. Tiffin. 2010. Genomic diversity, population structure, and migration following rapid range expansion in the Balsam poplar, *Populus balsamifera*. Molecular Ecology **19**:1212-1226.
Kirkpatrick, M. and N. Barton. 1997. Evolution of a species' range. American Naturalist **150**:1-23.
Lande, R. and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution **37**:1210-1226.
Leimu, R. and M. Fischer. 2008. A meta-analysis of local adaptation in plants. PLoS ONE **3**:e4010.
Librado, P. and J. Rozas. 2009. DnaSP v5: A software for comprhensive analysis of DNA polymorphism data. Bioinformatics **25**:1451-1452.
Loehle, C. 1998. Height growth rate tradeoffs determine northern and southern range limits for trees. Journal of Biogeography **25**:735-742.
Lopez, S., F. Rousset, F. H. Shaw, R. G. Shaw, and O. Ronce. 2009. Joint effects of inbreeding and local adaptation on the evolution of genetic load after fragmentation. Conservation Biology **23**:1618-1627.
MacArthur, R. H. 1972. Geographical ecology: patterns in the distribution of species. Princeton University Press, Princeton, New Jersey.
Macel, M., C. S. Lawson, S. R. Mortimer, M. Šmilauerova, A. Bischoff, L. Crémieux, J. Dolezal, A. R. Edwards, V. Lanta, T. Martijn Bezemer, W. H. Van der Putten, J.

M. Igual, C. Rodriguez-Barrueco, H. Müller-Schärer, and T. Steinger. 2007. Climate vs. soil factors in local adaptation of two common plant species. Ecology **88**:424-433.
Macnair, M. 1987. Heavy metal tolerance in plants: A model evolutionary system. Trends in Ecology & Evolution **2**:354-359.
Marsico, T. D. and J. J. Hellmann. 2009. Dispersal limitation inferred from an experimental translocation ofLomatium(Apiaceae) species outside their geographic ranges. Oikos **118**:1783-1792.
Mattson, W. J. and R. A. Haack. 1987. The role of drought in outbreaks of plant-eating insects. Bioscience **27**:110-118.
Mayr, E. 1963. Animal species and evolution. Harvard University Press, Cambridge, Mass.
Milliken, G. A. and D. E. Johnson. 2009. Analysis of Messy Data. Vol 1. Designed Experiments. Second edition. CRC Press, New York.
Mitchell-Olds, T. and R. G. Shaw. 1987. Regression analysis of natural selection: statistical inference and biological interpretation. Evolution **41**:1149-1161.
Moeller, D. A., M. A. Geber, and P. Tiffin. 2011. Population genetics and the evolution of geographic range limits in an annual plant. American Naturalist **in press**.
Morris, W. F., R. A. Hufbauer, A. A. Agrawal, J. D. Bever, V. A. Borowicz, G. S. Gilbert, J. L. Maron, C. E. Mitchell, I. M. Parker, and A. G. Power. 2007. Direct and interactive effects of enemies and mutualists on plant performance: a meta-analysis. Ecology **88**:1021-1029.
Naisbitt, T. and J. I. Sprent. 1993. The long-term effects of nitrate on the growth and nodule structure of the caesalpinioid herbaceous legume *Chamaecrista fasciculata* Michaux. Journal of Experimental Botany **44**:829-836.
Norton, L. R., L. G. Firbank, A. Scott, and A. R. Watkinson. 2005. Characterising spatial and temporal variation in the finite rate of population increase across the northern range boundary of the annual grass Vulpia fasciculata. Oecologia **144**:407-415.
Nuñez, M. A., T. R. Horton, and D. Simberloff. 2009. Lack of belowground mutualisms hinders Pinaceae invasions. Ecology **90**:2352-2359.
Ossowski, S., K. Schneeberger, J. I. Lucas-Lledo, N. Warthmann, R. M. Clark, R. G. Shaw, D. Weigel, and M. Lynch. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science **327**:92-94.
Ownbey, G. B. and T. Morley. 1993. Vascular Plants of Minnesota: A Checklist and Atlas. University of Minnesota Press, Minneapolis, MN.
Pannell, J. and B. Charlesworth. 1999. Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. Evolution **53**:664-676.
Pannell, J. R. and B. Charlesworth. 2000. Effects of metapopulation processes on measures of genetic diversity. Philosophical transactions of the Royal Society of London. Series B, Biological sciences **355**:1851-1864.
Parker, M. A. 2001. Mutualism as a constraint on invasion success for legumes and rhizobia. Diversity and Distributions **7**:125-136.

Parker, M. A. and D. A. Kennedy. 2006. Diversity and relationships of bradyrhizobia from legumes native to eastern North America. Canadian Journal of Microbiology **52**:1148-1157.

Parker, M. A., W. Malek, and I. M. Parker. 2006. Growth of an invasive legume is symbiont limited in newly occupied habitats. Diversity and Distributions **12**:563-571.

Parmesan, C. and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature **421**:37-42.

Pearson, R. G. and T. E. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography **12**:361-371.

Petit, R., A. El Mousadik, and O. Pons. 1998. Identifying populations for conservation on the basis of genetic markers. Conservation Biology **12**:844-855.

Phillips, B. L., G. P. Brown, J. K. Webb, and R. Shine. 2006a. Invasion and the evolution of speed in toads. Nature **439**:803.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006b. Maximum entropy modeling of species geographic distributions. Ecological Modelling **190**:231-259.

Price, T. D. and M. Kirkpatrick. 2009. Evolutionarily stable range limits set by interspecific competition. Proceedings of the Royal Society B: Biological Sciences **276**:1429.

Prince, S. D. and R. N. Carter. 1985. The geographical distribution of prickly lettuce (*Lactuca serriola*): III. Its performance in transplant sites beyond its distribution limit in Britain. Journal of Ecology **73**:49-64.

Pringle, A., J. D. Bever, M. Gardes, J. L. Parrent, M. C. Rillig, and J. N. Klironomos. 2009. Mycorrhizal Symbioses and Plant Invasions. Annual Review of Ecology, Evolution, and Systematics **40**:699-715.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945-959.

Pujol, B. and J. Pannell. 2008. Reduced responses to selection after species range expansion. Science **321**:96.

Purves, D. W. 2009. The demography of range boundaries versus range cores in eastern US tree species. Proceedings of the Royal Society B: Biological Sciences **276**:1477.

R Development Core Team. 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ratcliff, W. C., S. V. Kadam, and R. F. Denison. 2008. Poly-3-hydroxybutyrate (PHB) supports survival and reproduction in starving rhizobia. FEMS Microbiology Ecology **65**:391-399.

Reed, D. H. and R. Frankham. 2001. How closely correlated are molecular and quantitative measures of genetic variation? a meta analysis. Evolution **55**:1095-1103.

Rehfeldt, G. E., D. E. Ferguson, and N. L. Crookston. 2008. Quantifying the abundance of co-occurring conifers along inland Northwest (USA) climate gradients. Ecology **89**:2127-2139.

Reich, P. B., I. J. Wright, J. Cavender-Bares, J. Craine, J. Oleksyn, M. Westoby, and M. B. Walters. 2003. The evolution of plant functional variation: traits, spectra, and strategies. International Journal of Plant Sciences **164 (3 Suppl.)**:S143-S164.

Roach, D. and R. Wulff. 1987. Maternal effects in plants. Annual Review of Ecology and Systematics **18**:209-235.

Ross-Ibara, J., S. I. Wright, J. P. Foxe, A. Kawabe, L. DeRose-Wilson, G. Gos, D. Charlesworth, and B. S. Gaut. 2008. Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lytrata*. PLoS ONE **3**:e2411.

Sagarin, R. D. and S. D. Gaines. 2002. The 'abundant centre' distribution: to what extent is it a biogeographical rule? Ecology Letters **5**:137-147.

Sambatti, J. and K. J. Rice. 2006. Local adaptation, patterns of selection, and gene flow in the Californian serpentine sunflower (*Helianthus exilis*). Evolution **60**:696-710.

Samis, K. E. and C. G. Eckert. 2009. Ecological correlates of fitness across the northern geographic range limit of a Pacific Coast dune plant. Ecology **90**:3051-3061.

Sexton, J. P., P. J. McIntyre, A. L. Angert, and K. J. Rice. 2009. Evolution and Ecology of Species Range Limits. Annual Review of Ecology, Evolution, and Systematics **40**:415-436.

Shaw, R. G. and C. J. Geyer. 2010. Inferring fitness landscapes. Evolution **64**:2510-2520.

Shaw, R. G., C. J. Geyer, S. Wagenius, H. H. Hangelbroek, and J. R. Etterson. 2007a. More supporting data analysis for "Unifying life history analysis for inference of fitness and population growth" Technical report 661.

Shaw, R. G., C. J. Geyer, S. Wagenius, H. H. Hangelbroek, and J. R. Etterson. 2007b. Supporting data analysis for "Unifying life history analysis for inference of fitness and population growth" Technical report 658. Sch. of Statistics, Univ. of Minnesota.

Shaw, R. G., C. J. Geyer, S. Wagenius, H. H. Hangelbroek, and J. R. Etterson. 2008. Unifying life-history analyses for inference of fitness and population growth. American Naturalist **172**:E35-47.

Silvertown, J. 2001. Plants stand still, but their genes don't: non-trivial consequences of the obvious. Pages 3-20 *in* J. Silvertown and J. Antonovics, editors. Integrating Ecology and Evolution in a Spatial Context. Blackwell Science, Ltd, Oxford, England.

Singer, S. R., S. L. Maki, A. D. Farmer, D. Ilut, G. D. May, S. B. Cannon, and J. J. Doyle. 2009. Venturing beyond beans and peas: what can we learn from *Chamaecrista*? Plant Physiology **151**:1041-1047.

Somasegaran, P. and H. J. Hoben. 1994. Handbook for rhizobia: methods in legume - Rhizobium technology. Springer, Heidelberg, Germany.

Spoerke, J. M., H. H. Wilkinson, and M. A. Parker. 1996. Nonrandom genotypic associations in a legume-*Bradyrhizobium* mutualism. Evolution **50**:146-154.

Stanton-Geddes, J. 2011. Limits to range expansion in the native annual legume *Chamaecrista fasciculata*. Ph.D thesis. University of Minnesota, St. Paul, MN.
Stanton-Geddes, J. and C. G. Anderson. 2011. Does a facultative mutualism limit species range expansion? Oecologia.
Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. American Journal of Human Genetics **68**:978-989.
Svenning, J. C. and F. Skov. 2007. Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? Ecology Letters **10**:453-460.
Tajima, F. 1989. The effect of change in population size on DNA polymorphism. Genetics **123**:597-601.
Thomas, C. D., A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F. N. Erasmus, M. Ferreira de Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. van Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. Townsend Peterson, O. L. Phillips, and S. E. Williams. 2004. Extinction risk from climate change. Nature Australia **427**:145-148.
Thompson, J. N. 2005. The Geographic Mosaic of Coevolution. University of Chicago Press.
Tilman, D. 1994. Competition and Biodiversity in Spatially Structured Habitats. Ecology **75**:2-16.
Tlusty, B., J. M. Grossman, and P. H. Graham. 2004. Selection of rhizobia for prairie legumes used in restoration and reconstruction programs in Minnesota. Canadian Journal of Microbiology **50**:977-983.
van der Veken, S., J. Rogister, K. Verheyen, M. Hermy, and R. A. N. Nathan. 2007. Over the (range) edge: a 45-year transplant experiment with the perennial forest herb *Hyacinthoides non-scripta*. Journal of Ecology **95**:343-351.
van Heerwaarden, B., V. Kellermann, M. Schiffer, M. Blacket, C. M. Sgrò, and A. A. Hoffmann. 2009. Testing evolutionary hypotheses about species borders: patterns of genetic variation towards the southern borders of two rainforest Drosophila and a related habitat generalist. Proceedings of the Royal Society B: Biological Sciences **276**:1517-1526.
Vellend, M., J. A. Myers, S. Gardescu, and P. Marks. 2003. Dispersal of *Trillium* seeds by deer: implications for long-distance migration of forest herbs. Ecology **84**:1067-1072.
Venables, W. N. and B. D. Ripley. 2002. Modern Applied Statistics with S. Fourth edition. Springer, New York, NY.
Wakeley, J. and N. Aliacar. 2001. Gene geneaologies in a metapopulation. Genetics **159**:893-905.
Warren, R. J. 2010. An experimental test of well-described vegetation patterns across slope aspects using woodland herb transplants and manipulated abiotic drivers. New Phytologist **185**:1038-1049.
Went, F. W. 1942. The dependence of certain annual plants on shrubs in southern California deserts. Bulletin of the Torrey Botanical Club **69**:100-114.

Williams, J. W., W. Thompson, III, P. H. Richard, and P. Newby. 2000. Late quaternary biomes of Canada and the Eastern United States. Journal of Biogeography **27**:585-607.

Wolf, J. B. and M. J. Wade. 2009. What are maternal effects (and what are they not)? Philosophical Transactions of the Royal Society B: Biological Sciences **364**:1107-1115.