

## SYNTHESIS

# Are Species' Range Limits Simply Niche Limits Writ Large? A Review of Transplant Experiments beyond the Range

Anna L. Hargreaves,<sup>1,\*</sup> Karen E. Samis,<sup>2</sup> and Christopher G. Eckert<sup>1</sup>

1. Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada; 2. Department of Biology, University of Prince Edward Island, Charlottetown, Prince Edward Island C1A 4P3, Canada

Submitted November 28, 2012; Accepted September 26, 2013; Electronically published December 13, 2013

Online enhancements: appendixes. Dryad data: <http://dx.doi.org/10.5061/dryad.c3287>.

**ABSTRACT:** Many species' range limits (RL) occur across continuous environmental gradients without obvious barriers imposing them. Such RL are expected to reflect niche limits (NL) and thus to occur where populations cease to be self-sustaining. Transplant experiments comparing fitness within and beyond species' ranges can test this hypothesis, but interpretive power depends strongly on experimental design. We first identify often overlooked aspects of transplant design that are critical to establishing the causes of RL, especially incorporating transplant sites at, and source populations from, the range edge. We then conduct a meta-analysis of published beyond-range transplant experiments ( $n = 111$  tests). Most tests (75%) found that performance declined beyond the range, with the strongest declines detected when the measure of performance was lifetime fitness (83%), suggesting that RL commonly involve niche constraints (declining habitat quality). However, only 46% supported range limits occurring at NL; 26% (mostly geographic RL) fell short of NL with self-sustaining transplants beyond the range, and 23% (all elevational RL) exceeded NL with range-edge populations acting as demographic sinks. These data suggest an important but divergent role for dispersal, which may commonly constrain geographic distributions while extending elevational limits. Meta-analysis results also supported the importance of biotic interactions at RL, particularly the long-held assertion of their role in causing low-elevation and equatorial limits.

**Keywords:** range limit, geographic distribution, transplant experiments, meta-analysis, local adaptation, dispersal limitation.

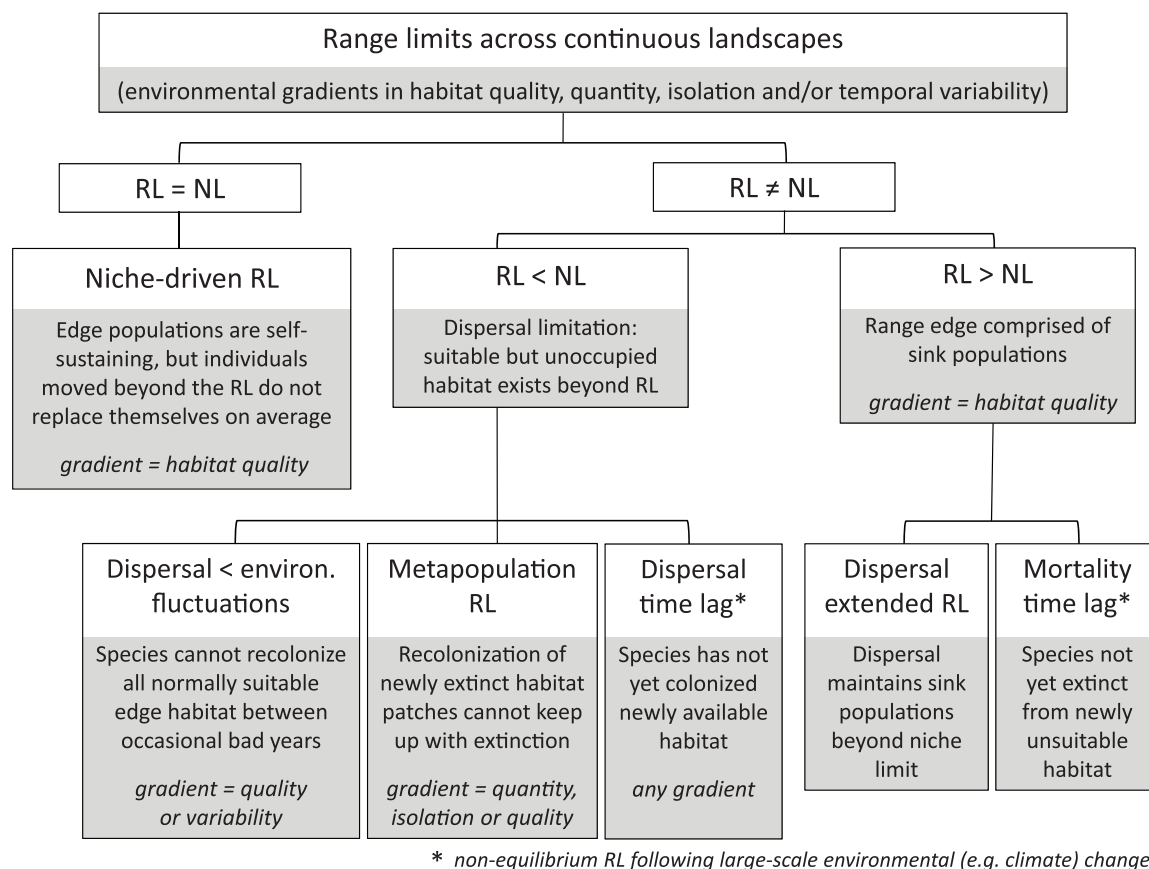
## Introduction

All species are limited in their distributions (Sexton et al. 2009). Although some range limits are imposed by sharp discontinuities in habitat, such as land to ocean, most occur across relatively continuous environments where habitat at the range edge does not differ greatly from that beyond the range (Gaston 2003). Explaining why range

limits arise across continuous landscapes involves fundamental questions in ecology, such as how niches constrain habitat use, while their stability over long time periods raises important evolutionary questions about how local adaptation and niche expansion are constrained (Holt 2003). Understanding range limits is also increasingly topical for conservation, as biologists attempt to predict the range expansions of invasive species (Alexander and Edwards 2010) and the range shifts expected to result from climate change (Parmesan et al. 2005; Sekercioglu et al. 2008; Doak and William 2010). However, despite sustained theoretical interest (Darwin 1859; MacArthur 1972; Antonovics 1976; Holt 2003) and a flush of empirical research in the past 2 decades, the ecological factors that impose range limits in the short term and the evolutionary processes that prevent natural selection from transcending them in the long term remain poorly understood (Gaston 2009). Moreover, the best way to test relevant hypotheses has not been explicitly addressed and perhaps contributes to uncertainty.

A simple hypothesis for range limits across continuous environments is that species stop occurring where abiotic and biotic constraints reduce individual fitness below that required for populations to be self-sustaining (i.e., finite rate of population growth  $\lambda < 1$ ; Holt 2003). In other words, species' range limits (RL) would be the spatial reflection of their niche limits (NL; Sexton et al. 2009; fig. 1, "Niche-driven RL"). If the environment consists of a gradient in ecologically important variables and range limits indeed coincide with niche limits, then a species' optimal habitat should occur in the interior of its range, with habitat quality declining toward (and beyond) the range edges (Brown 1984; Holt 2003). This predicted fitness decline toward range edges in turn suggests a pattern in abundance, where interior populations are expected to be larger and denser than those at the range edge (the "abundant center" model; Brown 1984; Sagarin and Gaines 2002). Although underlying gradients are often envisioned

\* Corresponding author; e-mail: [alhargreaves@gmail.com](mailto:alhargreaves@gmail.com).



**Figure 1:** Causes for range limits (RL) across continuous environments, highlighting the roles of niche constraints and dispersal in each case. The niche limit (NL) is the location beyond which habitat quality is too low for populations to be self-sustaining. We distinguish between equilibrium RL, which remain relatively stable for many generations, and nonequilibrium RL (indicated by an asterisk), which show directional movement over time. Equilibrium RL overlie and are ultimately imposed by one or more gradients of decreasing habitat quality or quantity or of increasing habitat isolation or temporal variability toward range limits. Gradients in habitat quality imply the involvement of niche constraints: fitness and population growth rates decline toward (and beyond) RL. Dispersal of individuals beyond RL caused solely by niche constraints does not result in range expansion. Transplant experiments can distinguish among the  $RL = NL$ ,  $RL < NL$ , and  $RL > NL$  hypotheses by establishing whether populations are self-sustaining at and beyond the range edge.

as abiotic, biotic interactions may also impose or sharpen range limits (Case et al. 2005). It has long been proposed that antagonistic biotic interactions, such as competition, parasitism, or herbivory, are especially important in defining range limits in relatively species-rich parts of the range (Normand et al. 2009), that is, lower elevation and latitudinal limits (Darwin 1859; Dobzhansky 1950; MacArthur 1972; Brown et al. 1996; for a review, see Gaston 2003).

When range limits involve gradients in habitat quality, such that fitness or population growth decline toward and beyond range limits, they are generally expected to occur where populations cease to be self-sustaining (i.e., at the niche limit, where  $\lambda = 1$  but  $\lambda < 1$  beyond it; fig. 1, “Niche-driven RL”). However, even when range limits are at equilibrium and involve niche constraints, they may not

coincide with the average location of the niche limit. Such exceptions have generally been explained by temporal environmental variation and/or dispersal (fig. 1; see Hargreaves and Eckert 2013 for a detailed review). First, high dispersal can maintain demographic sink populations (where  $\lambda < 1$ ) beyond the niche limit (fig. 1, “Dispersal-extended RL”; Pulliam 2000). Second, when range limits reflect constraints that fluctuate temporally (e.g., climate, populations of interacting species), the effect of temporal variation will depend on the relative magnitudes of environmental fluctuations versus dispersal. For example, when range limits reflect constraints in climatic tolerance, rare bad years may extirpate a species from edge habitat. Species with low dispersal may not recolonize edge habitat before the next bad year, such that range limits fall short of average niche limits (fig. 1, “Dispersal < environ. fluc-

tuations”). In contrast, a species with high dispersal may promptly recolonize edge habitats after bad years and even expand its range into normally unsuitable habitat during unusually good years (e.g., Bowman et al. 2005). The effect of dispersal on range limits depends not only on the species’ intrinsic dispersal potential but also on the steepness of the environmental gradient (Phillips 2012). A given dispersal distance will traverse a greater “ecological distance” on steep versus shallow gradients (Kirkpatrick and Barton 1997). Thus, as the steepness of the environmental gradient increases, there should be fewer cases of dispersal limitation where range limits fall short of niche limits (fig. 1, “ $RL < NL$ ”) and more cases where range limits coincide with or exceed niche limits (fig. 1, “ $RL = NL$ ” and “ $RL > NL$ ”).

Theoretically, range limits can also occur across continuous landscapes without underlying niche constraints, that is, no gradient in habitat quality per se (Holt et al. 2005; Bridle and Vines 2007; Kawecki 2008). If species exist as metapopulations toward the edge of their range, then equilibrium range limits may be imposed by gradients where the size of habitat patches, their proximity, or the ease of dispersal through the matrix between them declines toward limits (Holt et al. 2005). Metapopulation range limits then occur where the recolonization of vacant patches fails to keep pace with the extinction of occupied patches (Carter and Prince 1981; Holt and Keitt 2000; fig. 1). Metapopulation models can account for both sharp range limits despite a gradual underlying gradient and stable range limits despite high-quality habitat patches beyond the range edge (Holt and Keitt 2000).

The hypotheses above consider range limits at equilibrium, which remain in more or less the same location for many generations. However, ranges may expand or contract in response to large-scale environmental change, for example, climate warming, glacial advance, and conversion of land to agriculture. Such movements are distinguished from fluctuations of equilibrium range limits by their directional nature. If dispersal cannot keep pace with the opening of suitable habitat at expanding range edges (e.g., the frequent time lag in habitat colonization after glacial retreat; Svenning et al. 2008), then the range will be temporarily dispersal limited, regardless of the underlying environmental gradient (fig. 1, “Dispersal time lag”). Similarly, if extinction from deteriorating habitat at contracting range edges takes several generations, then ranges will be temporarily extended by sink populations (Jump et al. 2009; fig. 1, “Mortality time lag”).

Although there is little doubt that gradients in habitat quality and consequent niche constraints are involved in many range limits, the frequency with which range limits reflect the location at which populations cease to be self-sustaining (fig. 1, “Niche-driven RL”) is far from clear

(Griggs 1914; Carter and Prince 1981; Bozinovic et al. 2011). The strongest tests of whether range limits coincide with niche limits and of their proximate causes come from transplant experiments that compare fitness within and beyond species’ ranges (Gaston 2003). By standardizing density and thereby controlling for density-dependent effects on fitness, transplant experiments provide a powerful tool to assess whether habitat quality declines toward and beyond range limits and thus whether niche constraints are involved. Transplant experiments are the only way to determine whether populations would be self-sustaining beyond the current range limit and can therefore identify range limits that are (at least partially) dispersal limited (fig. 1, “ $RL < NL$ ”). Transplant sites at range edges can determine whether edge populations are demographic sinks, revealing cases where range limits extend beyond niche limits (fig. 1, “ $RL > NL$ ”). When combined with experimental manipulations, transplant experiments can also test the importance of possible range-limiting factors (e.g., Battisti et al. 2005). Finally, transplants can help uncover the evolutionary constraints maintaining stable range limits (Sexton et al. 2009; because transplants must be combined with quantitative- or population-genetic techniques to fully distinguish among evolutionary hypotheses, we do not discuss this application further). Despite their potential power, surprisingly few transplant experiments aimed at testing range-limit theory are optimally designed, often failing to include the best combination of source populations and transplant sites, replication in time and space, or adequate assessment of lifetime fitness.

The goals of this review are twofold. First, we consider the theoretical strengths and weaknesses of various transplant experiment designs (“Methodological Considerations in Transplant Design”), with the hope of encouraging future studies to maximize the explanatory power of their experiments. Second, we review all published studies that transplanted species or subspecies beyond their geographic (latitudinal or longitudinal) or elevational range to assess the role of niche constraints in determining range limits. We conduct a meta-analysis to: (a) assess how often elevational and/or geographic range limits involve niche constraints (i.e., declining habitat quality) by assessing how often and how much fitness declines beyond range margins; (b) assess how often range limits coincide with the location of the niche limit (i.e., transplants self-sustaining at the range edge but not beyond it; fig. 1); (c) test whether range limits across steeper environmental gradients more frequently coincide with niche limits by comparing range limits across elevational (putatively steeper) and geographic (shallower) gradients; (d) test whether biotic interactions are more important at low-elevation versus high-elevation limits, as commonly predicted; and (e) test whether experimental design (including selection of

transplant sites and sources, replication, measurement of fitness, and site alteration) influences experimental results.

### Methodological Considerations in Transplant Design

Four key aspects govern the explanatory power of transplant-experiment designs: (1) the location of sites to which individuals are transplanted (within the range interior [*I*], at the range edge [*E*], or beyond the range [*B*]; see table 1 for abbreviations); (2) the location of source populations that provide the transplant material (range interior [*i*] or edge [*e*]); (3) the replication of sites and sources in both space and time; and (4) the assessment of fitness. Table 2 shows the various combinations of the first two components and how often each design has been used in the literature (see “Meta-Analysis Methods” below). For brevity, we use the notation “*source populations*” (lowercase italics) → “*TRANSPLANT SITES*” (uppercase italics) to describe experimental designs (table 1).

#### *Transplant Sites*

The most basic transplant experiment involves a two-site design, in which a species is transplanted beyond its current range limit and its fitness is compared to that of a control transplant population within the range (experiments without a within-range control are inadequate as they cannot distinguish range-limiting niche constraints from either a bad year that lowers fitness throughout the range or negative transplant effects). Reduced fitness beyond the range suggests that conditions there are indeed less favorable than those within the range, implying that niche constraints (i.e., declining habitat quality) are in-

involved. Alternatively, if transplant populations are self-sustaining beyond the range, then it would seem that the range limit falls short of the niche limit (RL < NL). In this case, mechanisms such as dispersal limitation or meta-population dynamics must be inferred to explain the range limit.

Although two-site designs can identify when niche constraints are involved, those that place the control site in the range interior cannot determine whether edge populations are self-sustaining or demographic sinks. Better inference can be obtained from a three-site design that includes sites at the range edge and further toward the range center. For example, consider figure 2*A* and 2*B*; in both cases, populations are not self-sustaining beyond the range ( $\lambda < 1$ ; i.e., population growth is negative). Without an edge site, one might conclude that range and niche limits coincide (RL = NL), but the addition of an edge site reveals contrasting scenarios. In figure 2*B*, transplant populations at the range edge are less successful than those in the interior but are self-sustaining ( $\lambda = 1$ ), providing stronger evidence that range and niche limits coincide. In figure 2*A*, however, transplants at the range edge are unsustainable ( $\lambda < 1$ ), suggesting that edge populations are demographic sinks and consequently that the range limit exceeds the niche limit (RL > NL; Pulliam 2000). Range extension via sink populations may commonly occur in species with high dispersal abilities, as long as conditions beyond the range permit initial establishment (Pulliam 2000). Note that edge populations of long-lived species may experience many years where individuals do not replace themselves without  $\lambda$  falling below 1, and thus identification of true sink populations may be challenging.

Inclusion of an edge site is critical to correctly inter-

**Table 1:** Abbreviations used in the text

Type, abbreviation	Meaning
Site and source codes:	
<i>I, i</i>	In the range interior (distance from range edge varied greatly among tests)
<i>E, e</i>	At the range edge (as defined by the authors or verified by distribution information they provided)
<i>B</i>	Beyond the range edge (distance from the range edge varied widely among tests)
Transplant designs:	
<i>sources</i> → <i>SITES</i>	Source populations used → sites to which they were transplanted
For example, <i>ie</i> → <i>IEB</i>	For example, individuals from both interior ( <i>i</i> ) and edge ( <i>e</i> ) source populations were transplanted to interior, edge, and beyond sites ( <i>I, E, and B, respectively</i> )
Performance parameters:	
<i>source.SITE</i>	Value of a given performance parameter for that source at that transplant site
For example, <i>i.B</i>	For example, performance of interior ( <i>i</i> ) source populations transplanted to beyond ( <i>B</i> ) sites
Performance comparisons:	
<i>source.SITES</i>	Proportional difference in performance for a source population between two sites
For example, <i>i.BI</i>	For example, $\frac{i.B - i.I}{(i.I + i.B)/2}$

Note: Source populations are always lowercase, and transplant sites are always capitalized. See “Performance Declines beyond the Range” under “Meta-Analysis Methods” in the text for a description of how the best performance parameters and comparisons were selected.

Table 2: Transplant experiment designs

		Transplant sites used in the experiment			
		<i>B</i> <sup>a</sup>	<i>I</i> and <i>B</i>	<i>E</i> and <i>B</i>	<i>I</i> , <i>E</i> , and <i>B</i>
Source populations	Theoretical advantages	Most basic test of $\lambda^b$ beyond the range	Controls for transplant effects and bad years, enabling interpretation of $\lambda < 1$ beyond the range <sup>b</sup>	Tests for environmental gradient and whether niche and range limits coincide	
<i>i</i>	High genetic diversity, no negative maternal effects	$i \rightarrow B$	$i \rightarrow IB$ (66) <sup>c</sup>	$i \rightarrow EB$ (2)	$i \rightarrow IEB$ (34) <sup>d</sup>
<i>e</i>	Simulates biologically realistic colonization scenario	$e \rightarrow B$	$e \rightarrow IB$ (0)	$e \rightarrow EB$ (2)	$e \rightarrow IEB$ (0)
<i>i</i> and <i>e</i>	Enables tests of local adaptation and maternal effects	$ie \rightarrow B$	$ie \rightarrow IB$ (0)	$ie \rightarrow EB$ (0)	$ie \rightarrow IEB$ (7) <sup>e</sup>

Note: Transplant sites are designated by uppercase letters: range interior (*I*), at the range edge (*E*), beyond the range limit (*B*). Source populations are designated by lowercase letters (*i* and *e*). The inferential power of a design increases from top left (weakest) to bottom right (strongest). Cell numbers in brackets indicate the tests that used each design.

<sup>a</sup> Studies without an *I* or *E* control transplant site were not included in the meta-analysis.

<sup>b</sup> The  $\lambda$  is population growth rate, where populations are sustainable at  $\lambda \geq 1$ .

<sup>c</sup> Sixty-six tests from 20 studies.

<sup>d</sup> Thirty-four tests from 18 studies.

<sup>e</sup> Seven tests from four studies.

preting a fitness decline beyond the range, but identifying the range edge may be challenging for sparsely or patchily distributed species (Fortin et al. 2005). To facilitate the incorporation of individual case studies into the broader theoretical discussions of range limits, authors should describe the nature of the range limit studied (e.g., sharp or diffuse, continuous or patchy, whether there is evidence to support metapopulation dynamics). Studies should also specify how the limit's location was determined (e.g., whether outlying populations were included or excluded), and the distance from all transplant sites to the range edge should be provided to improve interpretation.

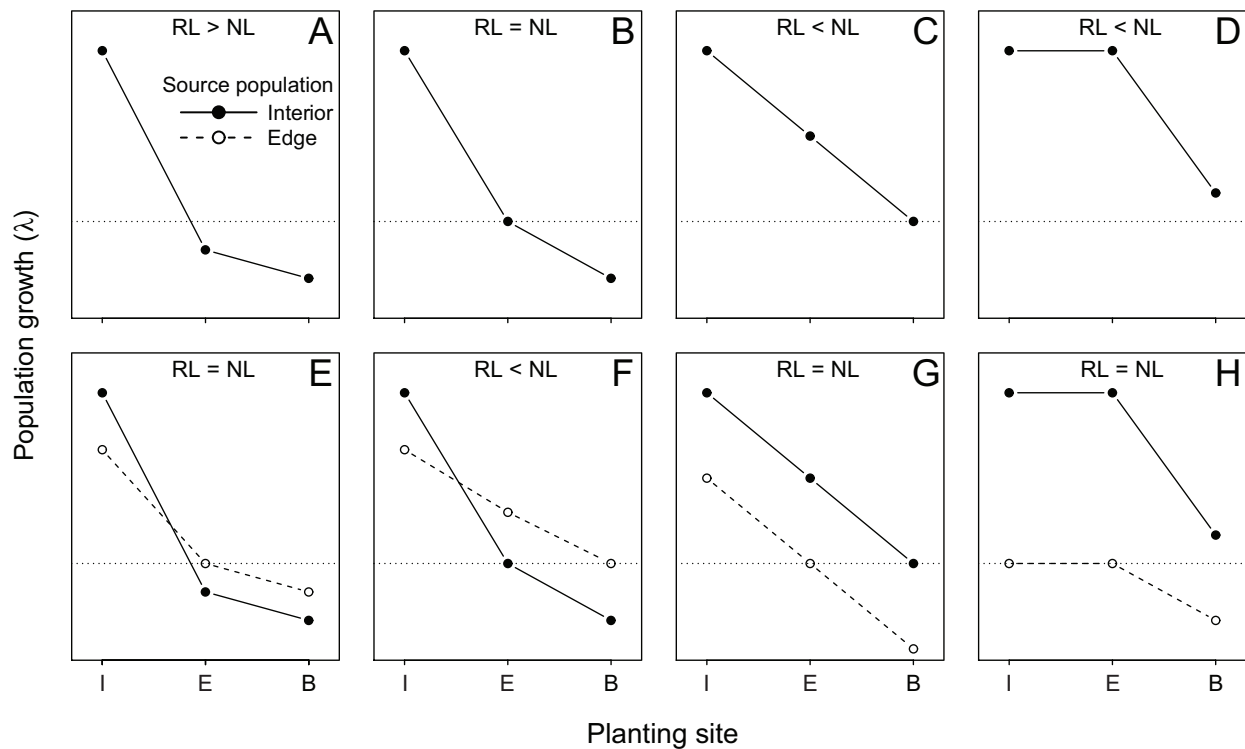
### Source Populations

Regardless of the number and location of transplant sites, inference is limited if only source populations from the range interior are used. Individuals from the range edge are the most likely to disperse beyond the range due simply to their proximity. Thus, edge sources (i.e., from populations close enough to the range edge that their progeny could disperse beyond it) provide the most realistic test of natural range-expansion scenarios. Further, if ranges overlie environmental gradients, then edge populations may (i) be better adapted to conditions at the edge and beyond, (ii) experience different developmental environments (e.g., maternal effects), or (iii) be smaller and more prone to genetic drift than interior populations. Because all of these factors can affect fitness (discussed below), performance of interior sources beyond the range reveals little about the actual mechanisms imposing range limits. To illustrate this point, figure 2 presents four scenarios in

which including an edge source changes the conclusion otherwise reached about the range limit, as data from interior sources overestimate (fig. 2E, 2F) or underestimate (fig. 2G, 2H) fitness declines at and beyond the range limit.

Given an underlying environmental gradient, interior-source populations may be poorly suited to conditions at and beyond the range limit if they are locally adapted to interior conditions (Leimu and Fischer 2008; Vergeer and Kunin 2013) or receive maternal provisioning or developmental cues more suitable for interior environments (Mousseau and Fox 1998; Donohue 2009). For example, in figure 2E and 2F, interior and edge sources are reciprocally transplanted and local populations outperform “foreign” ones in their home environment, a pattern diagnostic of local adaptation (Kawecki and Ebert 2004) but also consistent with nongenetic, adaptive maternal effects (Mousseau and Fox 1998). In figure 2E, edge-source populations are self-sustaining at the range limit, whereas interior sources are not. As edge sites no longer appear to be demographic sinks, this changes the interpretation of transplant results from range limits exceeding niche limits (fig. 2A) to range limits coinciding with niche limits. In figure 2F, edge-source populations also have positive growth rates beyond the range limit, suggesting that it occurs before niche limits are reached rather than coinciding with niche limits, as suggested by transplants of interior individuals alone (fig. 2B).

Edge sources may not always outperform interior sources at the range edge. If edge populations are small and isolated (as predicted by the abundant center model), they may suffer reduced genetic quality due to the fixation of deleterious alleles via genetic drift (Eckert et al. 2008;



**Figure 2:** Hypothetical results for transplant experiments using three types of transplant sites: within the range interior (*I*), at the range edge (*E*), and beyond the range (*B*). Dotted horizontal reference lines show the niche limit ( $\lambda = 1$ ), below which populations are not self-sustaining. Above each panel, the resulting conclusion about whether range limits (RL) coincide with (=), exceed (>), or fall short of (<) niche limits (NL) is noted. Top panels (A–D) illustrate cases where the growth of source populations from the range interior (filled circles) declines beyond the range and the resulting (sometimes erroneous) conclusions suggested by transplants of interior sources alone. Bottom panels (E–H) illustrate how inclusion of source populations from the range edge (open circles) alters conclusions of whether RL coincide with NL. *E, F*, Populations are locally adapted such that edge individuals outperform interior individuals at and beyond the RL. *G, H*, Edge populations suffer from poor genetic quality and/or poor provisioning and are outperformed at all sites. For illustration, we assume hypothetical results represent biological reality (accurate assessment of average lifetime fitness in each location; see “Methodological Considerations in Transplant Design”).

Hargreaves and Eckert 2013; fig. 2*G* or 2*H*). Further, if range limits partially reflect deteriorating habitat quality (e.g., fig. 2*G*), then offspring produced at range edges may suffer from poor maternal provisioning (i.e., negative non-genetic maternal effects). In such cases, interior sources may temporarily outperform edge sources at or beyond the range limit, but this advantage would disappear as interior genotypes produce offspring under poor edge conditions and/or succumb to genetic deterioration due to drift. Whether from poor maternal provisioning or low genetic quality, overall low performance of edge individuals suggests that range limits are caused by environmental gradients. Low quality of edge sources accompanied by fitness declines toward the limit (fig. 2*G*) suggest that range limits arise from a gradient in habitat quality, such that edge habitat is too poor to produce offspring that could survive in the even harsher conditions beyond the range. Alternatively, if habitat quality does not decline toward the

range edge (fig. 2*H*), then poor quality of edge individuals likely results from genetic drift in small populations resulting from reduced size and/or increased isolation of habitat patches. In either case, because edge individuals reflect edge conditions and are those capable of dispersing beyond the range, their failure beyond the range provides the definitive test of how range and niche limits relate.

Our goal in figure 2 is not to suggest that all scenarios are equally probable but to point out that correct biological interpretation of transplant experiments requires careful transplant design and that without edge sites and sources, interpretive power is very limited. The response of interior individuals can suggest whether niche constraints are involved, but without a concurrent edge source for comparison, interior sources reveal little about the processes imposing range limits as they are not subject to many of the limitations in play at range edges. A full transplant design that reciprocally transplants interior and edge-

source populations within the range and beyond it is the most powerful for understanding how range limits relate to niche limits and the underlying mechanisms imposing them.

### *Replication*

Interpretation of the panels in figure 2 assumes that the single points accurately reflect average lifetime fitness at different locations relative to the range limit. However, transplant experiments are labor intensive, and each region (interior, range edge, beyond range) is often represented by a single transplant site assumed to represent habitat quality in that region, thus confounding site effects with regional effects. Further, whereas putatively suitable habitat within the range can be identified by the presence of reproductive individuals, identifying habitat beyond the range may be difficult unless the species has well-known habitat requirements. Using a single beyond-range site increases the risk of choosing an inappropriate site and mistakenly concluding that conditions beyond the range are unsuitable. An especially powerful approach is to create a transect of sites across the range margin. The pattern of fitness variation can then reveal the steepness of underlying environmental gradients (Woodward and Pigott 1975; Levin and Clay 1984) and, if multiple source populations are used, patterns of local adaptation.

Transplants should also be replicated in time to increase the likelihood that experimental conditions reflect those in average years. Transplants conducted in an abnormally “good” year may conclude that conditions beyond the range (or at the range edge) are suitable, whereas during average years they are not. In contrast, those conducted in “bad” years may conclude that edge populations are sinks when they are normally self-sustaining or that habitat beyond the range is unsuitable when in fact the range is dispersal limited. Replication in time is especially important for long-lived organisms, which need only replace themselves once, on average, for population growth to be stable; transplants in most years might find range-edge conditions unfavorable even if there are actually enough good years for populations to persist. Realistically, even replicated transplants are unlikely to sample more than a few years. Studies should therefore measure climatic conditions during the experiment and report these in the context of long-term climate data from the area so that the normality of experimental conditions can be assessed.

### *Assessing Fitness*

Lifetime fitness is the cumulative result of traits expressed throughout an organism’s life and will not necessarily be accurately predicted by performance at a subset of life

stages. For example, 300 km north of its range, annual cocklebur survived and grew as well as it did within its range but failed to produce fertile seeds (Griffith and Watson 2005, 2006). Measuring performance at a subset of life stages may miss the critical stage at which persistence beyond the range is limited. Although measuring lifetime fitness will often be impossible for long-lived perennials (but see Campbell and Waser 2007 for a notable exception), it can be estimated by transplanting multiple life stages and incorporating results into a life table analysis (Woodward 1990; Latimer et al. 2009).

Studies may also fail to detect fitness declines beyond the range if potentially limiting factors are standardized across transplant sites (e.g., removing natural vegetation, watering transplanted plants throughout the experiment). Although standardizing some variables enables more controlled tests of how other factors influence range limits, it impedes assessment of whether range and niche limits coincide and of which factors ultimately limit the range.

## **Meta-Analysis Methods**

### *Literature Survey*

To assess whether the existing experimental evidence suggests overall patterns in the role of niche constraints in determining range limits, we conducted a meta-analysis of published transplant experiments. We exhaustively searched the literature (up to April 2013) for studies that transplanted species or subspecies to sites beyond their geographic or elevational range and compared performance to concurrent transplants within the range. We excluded studies of invasive species in their nonnative range, studies of taxa below the subspecies level (i.e., races, varieties, morphs), two studies whose presentation of results prevented extraction of the relevant data, and studies of transplants of  $F_2$  hybrids between sister species. We considered each range limit (high or low elevation and geographic cardinal direction) for each taxon (species or subspecies) to constitute a single test of whether range limits reflect niche limits (total 111 tests from 42 studies; full references available in the Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.c3287>; Hargreaves et al. 2014).

### *Data Compilation and Summary*

We compiled data on five performance parameters: emergence (e.g., hatching in insects, germination plus very early seedling survival in plants), survival (typically measured from the seedling or juvenile stage onward), growth, reproduction, and lifetime fitness. Lifetime fitness is the best measure of performance, followed by direct multiplicative components of lifetime fitness (emergence, survival, and

reproduction) and then growth, which may correlate with fitness components but is not a direct component itself. When lifetime fitness was not available but at least two multiplicative fitness components (emergence, survival, reproduction) were, we calculated their product to obtain an overall fitness parameter. We restrict some of the analyses below to tests that measured lifetime or overall fitness.

We classified transplant sites and source populations into three geographic regions: range interior, range edge, and beyond the range. For each test, we recorded the mean value of each performance parameter for each source population (interior [*i*] or edge [*e*]) at each transplant site (interior [*I*], edge [*E*], or beyond range [*B*]). When there were multiple sites per region, we averaged performance across sites within regions. If a performance parameter was zero in all regions, we assumed this was due to a constraint in study design or implementation, and data for that parameter were excluded. When there were multiple source populations from a given region (edge or interior), we averaged their performance at each site.

To facilitate comparisons among taxa and studies, we calculated the relative change in performance (performance comparison) between sites for each source as the difference in performance of a source between two sites divided by the mean performance of that source across both sites (table 1). This was done for each performance parameter available for each test. For measures based on counts or proportions, which cannot be negative, relative performance varies from  $-2$  (positive performance within the range [*I* or *E*] and zero fitness beyond the range [*B*]) to  $+2$  (zero fitness within the range but positive performance beyond it). If a performance parameter was zero at both sites in the comparison, the performance comparison was undefined (denominator = 0) and excluded from analyses. When data were available for multiple years or multiple measurements of the same parameter (e.g., height and leaf length as correlates of growth), we used the average of their performance comparisons. Transplants using multiple life stages (e.g., seedlings and adults) could have multiple measurements for growth and/or survival (emergence and reproduction only occur at a single life stage). In these cases, we averaged performance comparisons across life stages for growth and calculated the product of performance comparisons across life stages for survival, because survival is cumulative.

#### *Performance Declines beyond the Range*

To facilitate an overall synthesis of performance beyond the range, we selected the single measure for each test that best assessed the fate of potential dispersers beyond the range. The ultimate measure of how natural dispersers would fare is to compare lifetime fitness of edge individuals

at the range edge and beyond the range (see “Methodological Considerations in Transplant Design”). However, studies varied greatly in the sites and sources they included, and few measured lifetime fitness. Therefore, for each test, we identified the best available performance parameter and performance comparison as follows. When multiple parameters were available, we chose the best parameter according to the ranking: lifetime fitness > overall fitness > a single fitness component (i.e., emergence, survival, or reproduction) > growth. Multiple performance comparisons are only possible for three-site designs that incorporate interior, edge, and beyond-range transplant sites ( $ie \rightarrow IEB$  or  $i \rightarrow IEB$ ; no studies planted only edge sources at all three sites; table 2). Since performance at the range edge indicates whether edge populations are self-sustaining, and since comparing performance at the range edge and beyond is least likely to underestimate the suitability of beyond-range habitat, we selected the comparison “edge versus beyond” over “interior versus beyond” for three-site designs. For studies that transplanted both edge and interior sources ( $ie \rightarrow IEB$ ), we selected edge sources as the best comparison ( $e.BE$ ; table 1). For studies that used only interior sources ( $i \rightarrow IEB$ ), we selected  $i.BE$ , except for nine tests in which  $i.BE$  was undefined (zero fitness at both beyond and edge sites), for which we used the “beyond versus interior” comparison ( $i.BI$ ). Meta-analysis data are deposited in the Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.c3287> (Hargreaves et al. 2014).

All statistical analyses were conducted using R, version 2.15.1 (R Core Team 2013). The distribution of performance comparisons between within- and beyond-range sites among the 111 tests was not normal and had a pronounced peak at  $-2$  (fig. A1; figs. A1–A3, B1 available online). Accordingly, we calculated both mean and median performance comparisons across the range and calculated their 95% confidence intervals by bootstrapping (i.e., drawing random samples of the same number of tests with replacement; Sokal and Rohlf 1995), using “test” as the unit of replication and 10,000 resamplings. Confidence intervals for means and medians always overlapped broadly, so we report only mean performance comparisons  $\pm 1$  SD.

*Comparing Performance Declines.* We explored how performance declined beyond the range and how declines differed among contrasting range limits. We compared the best performance comparisons among categories of tests (elevational vs. geographic limits and high-elevation vs. low-elevation limits; too few tests of equatorial limits prevent the equivalent analysis across latitudes) by testing for differences in (a) the frequency of tests where performance declined beyond the range, using contingency table, Pearson  $\chi^2$  tests; and (b) the magnitude of performance de-



clines, using randomization tests to account for departure of data and residuals from normality. In randomization tests, the  $F$  values from a linear model (with type of range limit as a categorical predictor) were compared to a null distribution of 10,000  $F$  values generated by randomizing the response variable among test categories (Manly 1997). Randomization  $P$  values represent the proportion of  $F$  values from randomized data that are at least as great as the  $F$  calculated from observed data. We report them as “ $P \cong$ ” because values vary slightly among replicate analyses due to the randomization process. We further explored beyond-range performance declines using tests that placed multiple beyond-range sites in a transect of increasing distance from the range limit. We included only studies where at least one  $B$  site had nonzero fitness (once performance is zero, further declines in habitat quality are undetectable:  $n = 15$  tests, 4 geographic and 11 elevational). For these tests, we calculated the relative change in performance to the closest beyond-range site and to the farthest beyond-range site and tested whether these differed using a two-tailed paired  $t$ -test.

#### *Coincidence of Range Limits and Niche Limits*

We determined whether the evidence suggested that range and niche limits coincided using a subset of appropriate tests. To be included, a test must have included an edge site (see “Transplant Sites” above). Ideally, studies should also use an edge source, but so few did (9 of 111 tests) that we included tests that used only interior sources. We excluded studies where transplant populations failed ( $\lambda < 1$ ) at all sites, which may reflect methodological problems or a bad year throughout the range, and studies that altered potentially limiting factors across sites. We included only tests that assessed lifetime fitness or overall fitness (defined above). If average overall fitness was high enough that individuals would at least replace themselves at a given site, we considered this to indicate that  $\lambda \geq 1$ . Of 111 tests, 45 included an edge site, 88 measured at least two fitness components (from which we could calculate overall fitness), and 74 used natural transplant sites, but only 26 met all three criteria. We interpreted the results of the 26 tests as follows: if transplants beyond the range were self-sustaining ( $\lambda \geq 1$ ), then the range limit fell short of the niche limit (RL < NL). If transplants in the range interior were self-sustaining ( $\lambda \geq 1$ ) but those at the range edge and beyond were not ( $\lambda < 1$ ; i.e., edge populations were sinks), then the range limit exceeded the niche limit (RL > NL). If transplants at the range edge were self-sustaining but those beyond were not, then we accepted that range and niche limits coincided (RL = NL).

#### *The Effect of Gradient Steepness: Elevational versus Geographic Limits*

We tested whether the proportion of tests that found each relation between range and niche limits (RL < NL, RL = NL, RL > NL) differed between elevational and geographic limits using a  $2 \times 3$  contingency table,  $\chi^2$  test. Because a given distance “as the crow flies” covers a greater climatic change (and corresponding ecological change) along an elevational versus geographic gradient, we suspected that elevational studies might tend to cover greater climatic gradients and thus detect more severe fitness declines beyond the range and more cases where RL = NL versus RL < NL due to this methodological bias. We tested this using several approaches detailed in appendix B (appendixes A and B available online). In brief, we tested (1) whether elevational studies tended to place beyond-range sites climatically farther from (a) the range edge and (b) the within-range control site ( $I$  or  $E$ ) used in the best comparison (in fact, they did; app. B); (2) whether the distance between beyond-range and within-range control sites correlated negatively with the change in performance beyond the range (i.e., greater distances associated with greater declines), as might be expected given a continuously declining gradient in habitat quality; and (3) whether such a correlation might account for steeper performance declines and/or more frequent occurrence of RL  $\geq$  NL detected across elevational limits. To include geographic and elevational studies in the same analyses, we converted elevational changes (meters above sea level) to an equivalent change in latitudinal distance (km) based on how mean temperature changes with both latitude and altitude (Colwell et al. 2008; app. B).

#### *Importance of Biotic Interactions*

We tested the expectation that biotic interactions are more important in determining low versus high range limits. For each test, we classified whether the range limit was imposed (a) by abiotic factors alone or (b) partially or completely by biotic interactions, based on the authors’ assessment and data presented. We excluded tests where the range limit appeared to be dispersal limited, either as determined by the authors or because  $\lambda \geq 1$  beyond the range under natural conditions, and tests with insufficient data to draw a conclusion. Only one test of equatorial limits remained, so we restricted the analysis to elevational limits ( $n = 58$  tests). We tested whether the likelihood that ranges were abiotically controlled versus partially or completely biotically controlled varied among tests of high-elevation versus low-elevation limits using a  $2 \times 2$  contingency table,  $\chi^2$  test.

### Importance of Study Design

**Sites and Sources.** We tested the importance of simulating a realistic colonization scenario using tests that employed the full, reciprocal, three-site design ( $ie \rightarrow IEB$ ; table 2), per the predictions in figure 2. First, we tested whether performance changed more from home to beyond-range sites for interior versus edge sources (fig. 2E–2H). This expectation is based on the premises that given an environmental gradient, (a) beyond-range sites should differ more from interior sites than edge sites and (b) edge sources might be better adapted to beyond-range conditions. We used a paired  $t$ -test to compare the relative performance change of interior sources between beyond-range and interior sites ( $i.BI$ ) to that of edge sources between beyond-range and edge sites ( $e.BE$ ) with the expectation that  $i.BI < e.BE$ . Second, we tested the prediction that edge sources are better suited to beyond-range conditions (i.e., relative performance change from  $E$  to  $B$  sites is more negative for interior vs. edge sources,  $e.BE > i.BE$ ; fig. 2E, 2F) using a paired  $t$ -test. Third, we assessed the importance of including a range-edge transplant site using data from all three-site studies ( $i \rightarrow IEB$  and  $ie \rightarrow IEB$  designs). We tested the prediction that performance of interior sources should decline more from interior to beyond sites than from edge to beyond sites (i.e.,  $i.BI < i.BE$ ; fig. 2E–2G).

Unlike sites in the range interior and edge, suitable habitat beyond the range cannot be identified by the presence of the study species, resulting in a potential bias toward poor performance beyond the range simply due to poor site selection. We tested whether studies that used a single beyond-range site detected more frequent and/or greater

performance declines beyond the range than those with multiple beyond-range sites using  $\chi^2$  and randomization tests, respectively (details in app. B).

**Assessing Fitness.** We compared the frequency and magnitude of performance declines among tests that varied in how well they assessed fitness, per “Comparing Performance Declines” above. First, because measuring performance at a subset of life stages may miss the critical stage at which persistence is limited, we expected that tests measuring lifetime fitness would reveal the greatest beyond-range performance declines, followed by tests measuring overall fitness and then by tests measuring single performance parameters. We compared the relative change in performance beyond the range between these three categories using a  $3 \times 2 \chi^2$  test and a randomization test. Second, studies may underestimate performance declines toward and beyond range limits if potentially limiting factors are standardized across transplant sites. Accordingly, we compared relative performance change beyond the range between tests that altered sites and tests where individuals were transplanted into natural habitat using a  $2 \times 2 \chi^2$  test and a randomization test.

### Meta-Analysis Results and Discussion

Forty-two transplant studies met our criteria (summarized in tables 2, 3), yielding 111 tests of range limits involving 93 taxa (88 plant species or subspecies from 29 families and 5 invertebrate species from 5 families). In most cases only 1 range limit was tested per taxon, but 14 taxa (all at the species level) had 2 range limits tested, and 2 species

**Table 3:** Support for the hypothesis that range limits (RL) and niche limits (NL) coincide (i.e., RL = NL)

Outcome: study inclusion criteria	All limits	Geographic limits	Elevational limits	Geographic vs. elevational $\chi^2$ <sup>a</sup>
Fitness declined beyond the RL:				
All included	75% (111)	70% (33)	77% (78)	.642, $P = .47$
Estimated LTF	83% (23)	86% (14)	78% (9)	.240, $P = .99$
Natural conditions	84% (74)	75% (24)	88% (50)	2.017, $P = .19$
RL =, <, > NL, respectively:				
Included an $E$ site; natural conditions; <sup>b</sup> assessed at least two of Em, S, R; $\lambda$ not <1 at all sites	46%, 23%, 31% (26)	17%, 83%, 0% (6)	55%, 5%, 40% (20)	16.142, $P = .0010$ <sup>c</sup>
As above, plus LTF	62%, 38%, 0% (8)	25%, 75%, 0% (4)	100% (4) <sup>d</sup>	4.800, $P = .15$ <sup>c</sup>

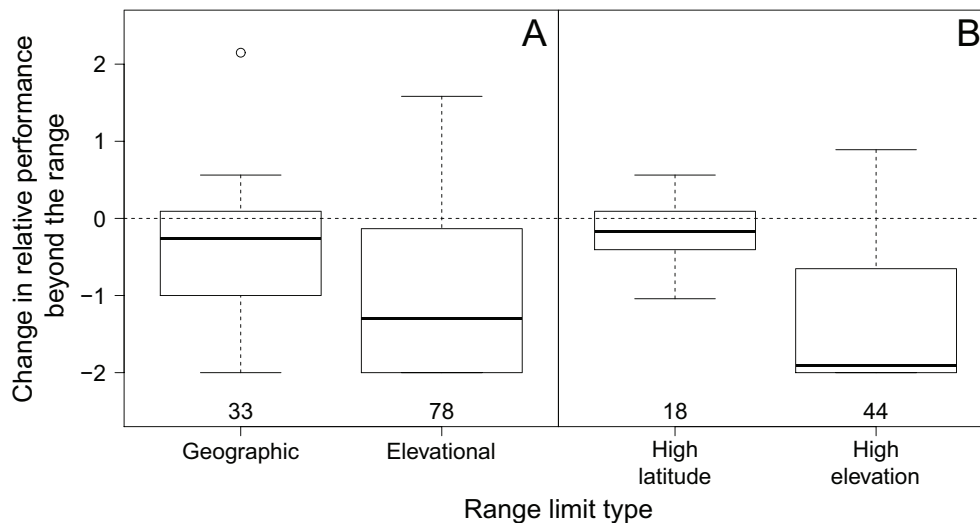
Note: RL < NL suggests the range is limited by dispersal, RL > NL suggests edge ( $E$ ) populations are demographic sinks. Cell contents are % of tests supporting each outcome (total tests), where each test is a unique RL for a single taxon. LTF = lifetime fitness, Em = emergence, S = survival, R = reproduction.

<sup>a</sup>  $\chi^2$  tests compare geographic and elevational limits using  $2 \times 2$  (df = 1) and  $2 \times 3$  (df = 2) contingency tables for the frequency of fitness declines beyond the RL and RL versus NL outcomes, respectively.

<sup>b</sup> “Natural conditions” means neither biotic nor abiotic factors were standardized across transplant sites.

<sup>c</sup> Only the  $\chi^2$  test outcomes for RL = NL comparisons are shown.

<sup>d</sup> Only the outcome for RL = NL is shown, as tests were not available for RL < or > NL within elevational limits.



**Figure 3:** Performance of transplants declined more strongly beyond elevational than geographic range limits. *A*, All geographic and elevational limits (randomization  $P \cong .0023$ ). *B*, Only high-latitude geographic and high-elevation limits (randomization  $P < .00001$ ). The dashed line at zero indicates no difference in performance between individuals transplanted within the range (interior or edge sites) and those transplanted beyond it, negative values mean fitness declined beyond the range, and  $-2$  indicates zero fitness beyond the range. Numbers below boxes are the tests in each category. For boxes, the center line is the median, and the top and bottom lines represent the twenty-fifth and seventy-fifth percentiles, respectively. Whiskers extend to the maximum and minimum values or 1.5 interquartile ranges from the median (whichever is less), with the open circle lying beyond this range.

had 3 tested. There were more tests of elevational than geographic limits, and in both cases, the cold range limit was tested more often (44 vs. 34 tests of high-elevation vs. low-elevation limits and 18 vs. 5 tests of polar vs. equatorial limits). Another 10 tests considered western or eastern (longitudinal) geographic limits. Ten tests involved range limits that coincided with interspecific hybrid zones. Most tests were modestly replicated through time: 83 tests were conducted for  $\geq 2$  years, but only 8 spanned  $\geq 1$  generation (i.e., replicated lifetime fitness). We assume that the number of tests in abnormally good years (where normally unsustainable edge and beyond-range sites appear sustainable) and abnormally bad years (where normally sustainable edge and beyond-range sites appear unsustainable) are more or less equal and balance each other out. A two-site design using interior-source populations and transplant sites was by far the most common ( $i \rightarrow IB$ ; table 2), even though inferential power for the same effort would be much greater if sources came from the range edge (fig. 2). Only seven tests used the optimal, fully reciprocal three-site design ( $ie \rightarrow IEB$ ; table 2).

#### *Performance Declines beyond the Range*

Performance declined beyond the range in 75% of 111 tests (considering the best parameter and best comparison for each test; table 3). This suggests that range limits are

often associated with a decline in habitat quality (i.e., an underlying niche constraint) but that declining habitat quality alone may not explain a sizeable proportion (25%) of current limits. When more than one range limit was tested for the same species, performance always declined across at least one limit and declined across all limits for 8 of 16 species. Mean relative change in performance from within-range to beyond-range sites was  $-0.825$ , and the 95% confidence intervals ( $-1.01$ – $-0.634$ ) show that this is significantly below zero, indicating that performance declined beyond the range. Fewer than half the tests (46) replicated beyond-range sites, but this did not seem to influence the frequency or magnitude of fitness declines detected (app. B). When tests included a transect of beyond-range sites, performance declined significantly more from within-range control sites to the farthest site beyond the range than to the closest beyond-range site (paired  $t_{14} = 2.49$ ,  $P \cong .0259$ ,  $n = 15$  tests), further supporting the common association between range limits and declining habitat quality. Performance declines were more frequent and greater in magnitude across elevation limits than geographic limits (table 3; fig. 3; app. B). Comparing high-elevation versus low-elevation limits (too few tests of equatorial limits prevent this comparison across latitudes), performance declines were more frequent across high limits (86% vs. 64% of tests;  $\chi_1^2 = 5.07$ ,  $P = .031$ ) and greater in magnitude (mean  $\pm 1$  SD: high  $-1.288 \pm 0.903$ ,

$n = 42$  tests; low  $-0.622 \pm 1.00$ ,  $n = 34$ ; randomization  $P \cong .0032$ ).

#### *Coincidence of Range Limits and Niche Limits*

Of the 26 tests that included an edge site, measured at least two fitness components, and used natural transplant sites, only 46% suggested that range and niche limits coincided (table 3). Of the 14 range limits that did not coincide with niche limits, 8 seemed to exceed niche limits ( $RL > NL$ , indicated by zero fitness at edge sites), suggesting that dispersal maintains sink populations beyond niche limits. Six range limits seemed to occur before niche limits ( $RL < NL$ , indicated by good performance beyond the range), suggesting dispersal limitation (fig. 1). Unfortunately, lack of information on dispersal abilities, historical range-limit locations, and habitat patchiness prevents us from untangling chronic dispersal limitation across stable range limits (i.e., failure to recolonize normally good edge habitat after bad years or ephemeral edge patches in metapopulations) from dispersal lags imposing temporary range limits after directional climate change (e.g., postglaciation recolonization or following contemporary climate warming).

#### *The Effect of Gradient Steepness: Elevational versus Geographic Limits*

Given the shorter spatial distances needed to traverse elevational climatic gradients, dispersal should limit elevational ranges less often and maintain range limits beyond niche limits more often than for geographic ranges. In other words, we expected more cases where  $RL > NL$  (edge populations are sinks) and  $RL = NL$ . Indeed, only one of the six (17%) geographic limits coincided with niche limits, whereas 83% fell short, and none exceeded niche limits (table 3). This suggests that although 70% of geographic ranges occurred across a gradient of declining fitness (table 3), the actual range limit may often be constrained by dispersal. Notably, all six geographic tests came from North America and Europe, where postglaciation migration lags are relatively common (Qian and Ricklefs 2007; Normand et al. 2011). In contrast, 55% of the 20 elevational limits coincided with niche limits, and 40% exceeded niche limits, suggesting that elevational edge populations are often demographic sinks (table 3). All tests where  $RL > NL$  came from a Himalayan study conducted at the maximum elevations at which vascular plants persist (Klimeš and Doležal 2010), illustrating how steep elevational gradients might pave the way for dispersal to maintain sink populations beyond niche limits. The transplant evidence to date thus suggests that dispersal plays contrasting roles in geographic and elevational limits, often

limiting geographic ranges while maintaining sink populations beyond elevational niche limits. Further, fitness declines were significantly steeper across elevational limits (fig. 3). Although these results are based on relatively small sample sizes (especially for geographic RL), the pattern suggests researchers should be cautious about using elevational gradients as proxies for latitudinal ones.

We tested whether the steepness of elevational range limits resulted in a methodological bias, whereby elevational studies placed sites farther apart climatically due to the ease of covering large climatic differences along steep gradients. Such a bias might account for the steeper fitness declines and more frequent occurrence of  $RL = NL$  versus  $RL < NL$  (table 3; app. B). Although, elevational studies did tend to place beyond-range transplant sites climatically farther from the range limit and within-range control sites, this did not account for the steeper performance declines or more frequent occurrence of  $RL = NL$  detected across elevational limits (app. B). If we compare only equivalent limits, that is, high-elevation versus polar limits (too few tests at equatorial limits prevented comparison to low-elevation limits), while controlling for climatic distance between planting sites in the comparison, results are even more pronounced (fig. 3); significantly greater beyond-range performance declines were detected across high-elevation versus polar limits ( $F_{1,56} = 18.79$ ,  $P \cong .0001$ ; app. B). Given a linear environmental gradient in habitat quality, one might expect an equal performance decline between any two points separated by an equal distance along the gradient, in this case a temperature gradient. The consistently greater fitness declines across elevation limits, even while controlling for temperature-related distance along the gradient, suggests that mean temperature change may not capture the overall ecological steepness of elevation gradients (i.e., other ecological factors change more quickly for a given change in temperature across elevation vs. geographic gradients). Studies that compared equivalent elevational and latitudinal limits for the same species would be invaluable in exploring these issues.

#### *Importance of Biotic Interactions*

As predicted, high-elevation range limits were more likely to be governed by abiotic factors alone (82% of 38 tests) than were low-elevation limits (45% of 20 tests;  $\chi_1^2 = 8.91$ ,  $P = .004$ ). Latitudinal limits showed the same pattern (8 of 8 high limits and 0 of 1 low limits were caused solely by abiotic factors), but there were too few studies of equatorial limits to test this statistically. These results support an increasing role of biotic interactions at warmer and putatively more-species-rich range limits. More convincing evidence would come from tests that contrasted different limits for the same species, thereby controlling

for variation among species and for transplant methods among studies. To date, however, most of the few transplants across multiple limits per species have standardized sites in a way that affects biotic interactions, preventing an overall comparison. Interestingly, transplants that ameliorated potentially limiting biotic interactions detected fewer and smaller fitness declines beyond the range than those transplanting into unaltered natural habitat (see “Assessing Fitness” below), supporting the general importance of biotic interactions in limiting species’ distributions. Clearly, studies that test both high and low range limits for the same species without standardizing biotic interactions would make a much-needed contribution to the literature, as would controlled and fully crossed experimental manipulations of biotic and abiotic factors. Information on whether high and low limits truly represent species-poor and species-rich environments would help tease apart pattern and process (Schemske 2009).

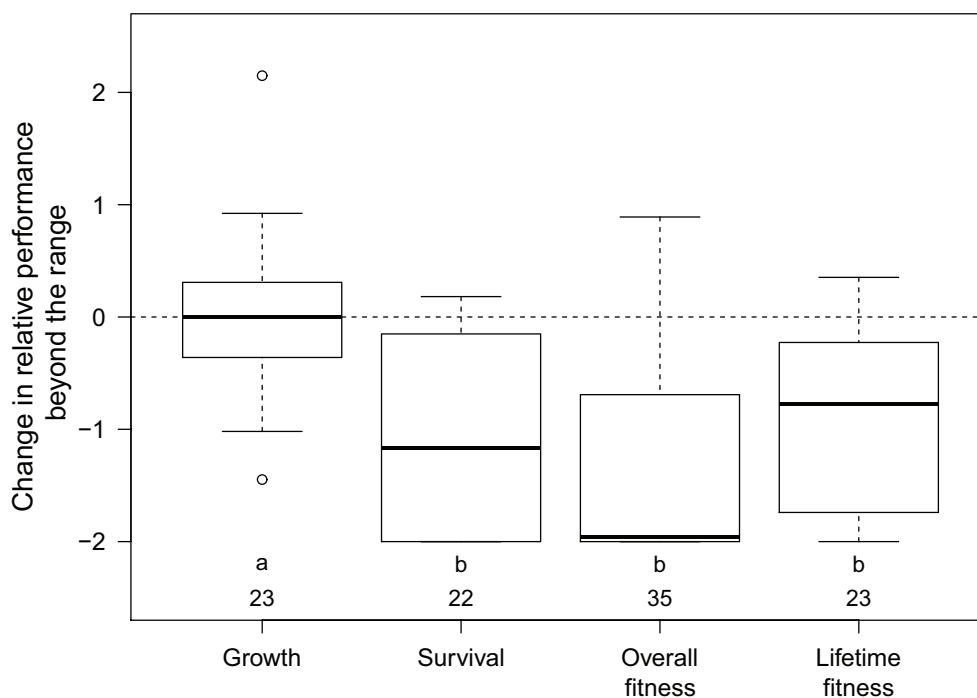
#### Importance of Study Design

*Sites and Sources.* We tested the importance of including edge sites and sources and of replicating beyond-range sites, according to the predictions in “Methodological Considerations in Transplant Design.” Only seven tests used the full reciprocal transplant design (*ie* → *IEB*; Gauthier et al. 1998; Geber and Eckhart 2005; Poll et al. 2009; Samis and Eckert 2009), with which the effect of including edge sites and sources is evaluated. Among these, we did not find support for the prediction that performance would decline more sharply beyond the range for interior versus edge sources (paired  $t_6 = 0.64$ ,  $P = .95$ ). Not all of the seven tests provided adequate data to assess whether  $RL = NL$  (see below), so we could not evaluate how often use of an edge source changed the conclusions reached using interior sources, as predicted in figure 2. We could test the expectation that edge sources might be better adapted to beyond-range conditions by testing whether their performance declined less than that of interior sources when transplanted beyond the range (i.e., *i.BE* is more negative than *e.BE*) but found no difference among the seven tests available (paired  $t_6 = 0.54$ ,  $P = .60$ ). Contrary to expectations, performance of interior sources did not decline more from range-interior to beyond-range sites than from edge to beyond sites (i.e., *i.BI* vs. *i.BE*;  $n = 30$ , paired  $t_2 = 0.058$ ,  $P = .95$ ). Thus, although most studies detected a decline in habitat quality beyond the range (table 3), and most with multiple beyond-range sites found that habitat quality declined with increasing distance from the range edge (see above), there is less evidence for declining habitat quality from the range center to the range edge.

*Assessing Fitness.* Performance declines beyond the range

for each parameter (emergence, survival, growth, reproduction, overall fitness, lifetime fitness) are summarized in figure A2. Considering the best parameter for each test, the proportion of tests in which performance declined beyond the range was slightly higher when assessing lifetime fitness (83%,  $n = 23$  tests) and overall fitness (83%,  $n = 35$  tests) than single performance parameters (66%,  $n = 53$  tests), but the difference was not significant ( $3 \times 2 \chi_2^2 = 4.10$ ,  $P = .14$ ). However, performance declined more strongly for tests of lifetime or overall fitness than tests of single parameters (randomization  $P \cong .0004$ ). Among individual parameters, only growth and survival had sample sizes large enough to compare statistically. Tests that measured survival showed declines similar to those measuring overall and lifetime fitness (fig. 4). In contrast, tests that measured only growth revealed no mean change in performance beyond the range (fig. 4), emphasizing the inadequacy of growth alone for assessing fitness beyond range limits. Rerunning the analyses excluding the 23 tests that measured only growth increased the frequency and magnitude of beyond-range performance declines and reduced the significance of several statistical comparisons (by reducing the sample size) but did not alter the qualitative nature of any comparison (results not shown).

Studies may underestimate fitness declines toward and beyond range limits if potentially limiting factors are standardized or otherwise altered across transplant sites. Almost half the studies we reviewed (42% of studies, 33% of tests) altered one or more potentially limiting factors across sites. The most commonly altered was competition; 33 tests (from 13 plant studies) reduced or eliminated belowground and/or aboveground competition, either deliberately or as a side effect of transplanting into gardens or pots. Fifteen tests standardized soil conditions, 15 watered transplants throughout the experiment, 3 reduced herbivory or predation, and 1 applied fungicide to control disease. Biotic factors were always among those altered; thus, 33% of 111 tests did not adequately assess their role in limiting ranges. Notably, transplants into unaltered, natural environments were significantly more likely to detect a performance decline beyond the range than those into ameliorated environments ( $\chi_1^2 = 9.55$ ,  $P = .002$ ; table 3). Similarly, the mean performance decline beyond the range was much greater for transplants into unaltered ( $-0.997 \pm 0.996$ ) than altered ( $-0.482 \pm 0.889$ ) environments (randomization  $P \cong .0097$ ; fig. A3). Because site alterations predominantly ameliorated biotic factors, these results suggest that biotic interactions, especially competition, may often be important in limiting species’ distributions, a conclusion increasingly supported by both theoretical (Case et al. 2005; Gilman et al. 2010) and biogeographical (Araujo and Luoto 2007) studies.



**Figure 4:** Different performance parameters vary in detecting fitness declines beyond the range, considering the best comparison and best parameter for each test (see text). Overall performance is the product of  $\geq 2$  of emergence, survival, or reproduction but not lifetime fitness. Contrasting letters indicate significant differences in pairwise tests (where Type I error rates were controlled by Bonferroni correction), and numbers indicate number of tests. The dashed line at zero indicates no difference in performance between individuals transplanted within the range (interior or edge sites) and those transplanted beyond the range, whereas negative values indicate that fitness declined beyond the range. Boxes are as in figure 3.

### Conclusions and Recommendations for Future Research

Transplant experiments beyond the range are a cornerstone of research on the limits to species' distributions. Despite great potential and increasing use, transplant experiments vary in their ability to assess whether range limits involve niche constraints, how closely range and niche limits coincide, and which ecological variables impose range limits when niche constraints are involved. We end with three main conclusions and six recommendations intended to help future studies make the strongest possible contribution to the growing literature on range limits, drawing attention to exemplary studies.

#### Conclusions

Fitness declined beyond the range limit in 75% of 111 cases reviewed, and declines were more severe at transplants farther outside a given range limit, suggesting that declining habitat quality and resulting niche constraints generally play a role in imposing range limits. However, the evidence that range and niche limits coincide is much

weaker (46% of tests) and could only be obtained from 26 tests. Cases where they do not coincide are largely explained by dispersal, which plays different roles in generating elevational versus geographic range limits. Dispersal limitation seems to commonly constrain the location of geographic limits, whereas elevational limits more often coincided with niche limits, or exceeded them, as dispersal maintained sink populations beyond niche limits. This pattern likely reflects the shorter distances over which climate and related factors change across elevational versus geographic scales and means that care should be taken in using elevation range limits as proxies for latitudinal ones. Finally, biotic interactions seem important in limiting ranges, especially at warmer, putatively more-species-rich range limits. However, this remains poorly tested due to the large number of studies that alter biotic interactions (especially competition) at transplant sites.

#### Recommendations

To provide the best experimental tests of why range limits exist, all transplant experiments should strive to achieve the following methodological practices: (1) Clearly state

where the range limit is and where transplant sites are relative to it (ideally with a map) and how the range limit was identified, including sources of uncertainty, for example, zones of continuous versus discontinuous distribution. (2) Include a range-edge site so that the full spectrum of experimental outcomes is possible and some conclusion can be made about whether range and niche limits coincide. (3) Include an edge-source population to mimic realistic colonization scenarios. Ideally, incorporate edge- and interior-source populations to assess how local adaptation and/or maternal effects are involved in range-limit formation, which in turn can begin to illuminate evolutionary processes governing range limits (Sexton et al. 2009). (4) Include multiple beyond-range transplant sites to better capture a gradient in habitat quality beyond the range and to avoid potentially confusing poor site selection with inhospitable habitat. (5) Measure or at least estimate lifetime fitness, as no other metric can really answer whether populations could persist beyond the range. If this is not possible, measure components of lifetime fitness (emergence, survival, or reproductive success) and not just growth, which our results suggest is an unreliable proxy for fitness. (6) Conduct at least one set of transplants under natural conditions so that all possible causes of range limits, including biotic interactions, can be evaluated.

The studies that make the greatest contribution to building and testing range-limit theory in the future will be those that go beyond simple transplant experiments. Studies that experimentally manipulate a putatively limiting factor (with a control treatment under natural conditions) are especially valuable in determining the proximate ecological causes of range limits (e.g., Stinson 2005; Griffith and Watson 2006). Comparing high-elevation and low-elevation range limits, polar and equatorial geographic range limits, or equivalent elevational and latitudinal range limits of the same species would help determine how often and to what extent the mechanisms that cause these range limits differ (e.g., Barton 1993). In all cases, studies that reciprocally transplant edge and interior sources within the range and plant both beyond the range are especially valuable (e.g., Geber and Eckhart 2005), as they can illuminate not only how range limits relate to niche limits but also the underlying ecological and evolutionary mechanisms imposing them.

### Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada (Discovery Grants to C.G.E. and K.E.S. and an Alexander Graham Bell Canada Graduate Scholarship–Doctoral [CGS D] to

A.L.H.) and Queen's University and IODE Canada (doctoral scholarships to A.L.H.). We thank two anonymous reviewers and O. Schmitz for helpful comments on early drafts.

### Literature Cited

- Alexander, J. M., and P. J. Edwards. 2010. Limits to the niche and range margins of alien species. *Oikos* 119:1377–1386.
- Antonovics, J. 1976. The nature of limits to natural selection. *Annals of the Missouri Botanical Garden* 63:224–247.
- Araujo, M. B., and M. Luoto. 2007. The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography* 16:743–753.
- Barton, A. M. 1993. Factors controlling plant distributions: drought, competition, and fire in montane pines in Arizona. *Ecological Monographs* 63:367–397.
- Battisti, A., M. Stastny, S. Netherer, C. Robinet, A. Schopf, A. Roques, and S. Larsson. 2005. Expansion of geographic range in the pine processionary moth caused by increased winter temperatures. *Ecological Applications* 15:2084–2096.
- Bowman, J., G. L. Holloway, J. R. Malcolm, K. R. Middel, and P. J. Wilson. 2005. Northern range boundary dynamics of southern flying squirrels: evidence of an energetic bottleneck. *Canadian Journal of Zoology* 83:1486–1494.
- Bozinovic, F., P. Calosi, and J. Spicer. 2011. Physiological correlates of geographic range in animals. *Annual Review of Ecology, Evolution, and Systematics* 42:155–179.
- Bridle, J., and T. Vines. 2007. Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology and Evolution* 22:317–344.
- Brown, J. H. 1984. On the relationship between abundance and distribution of species. *American Naturalist* 124:255–279.
- Brown, J. H., G. Stevens, and D. M. Kaufman. 1996. The geographic range: size, shape, boundaries and internal structure. *Annual Review of Ecology and Systematics* 27:597–623.
- Campbell, D. R., and N. M. Waser. 2007. Evolutionary dynamics of an *Ipomopsis* hybrid zone: confronting models with lifetime fitness data. *American Naturalist* 169:298–310.
- Carter, R. N., and S. D. Prince. 1981. Epidemic models used to explain biogeographical distribution limits. *Nature* 293:644–645.
- Case, T. J., R. D. Holt, M. A. McPeck, and T. H. Keitt. 2005. The community context of species' borders: ecological and evolutionary perspectives. *Oikos* 108:28–46.
- Colwell, R. K., G. Brehm, C. Cardelús, A. Gilman, and J. T. Longino. 2008. Global warming, elevational range shifts, and lowland biotic attrition in the wet tropics. *Science* 322:258–261.
- Darwin, C. 1859. *On the origins of species by means of natural selection*. J. Murray, London.
- Doak, D. F., and F. William. 2010. Demographic compensation and tipping points in climate-induced range shifts. *Nature* 467:959–962.
- Dobzhansky, T. 1950. Evolution in the tropics. *American Scientist* 38:209–221.
- Donohue, K. 2009. Completing the cycle: maternal effects as the missing link in plant life histories. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364:1059–1074.
- Eckert, C. G., K. E. Samis, and S. Lougheed. 2008. Genetic variation

- across species' geographic ranges: the central-margin hypothesis and beyond. *Molecular Ecology* 17:1170–1188.
- Fortin, M. J., T. H. Keitt, B. J. Maurer, M. L. Taper, D. M. Kaufman, and T. M. Blackburn. 2005. Species' geographic ranges and distributional limits: pattern analysis and statistical issues. *Oikos* 108: 7–17.
- Gaston, K. J. 2003. *The structure and dynamics of geographic ranges*. Oxford University Press, Oxford.
- . 2009. Geographic range limits: achieving synthesis. *Proceedings of the Royal Society B: Biological Sciences* 276:1395–1406.
- Gauthier, P., R. Lumaret, and A. Bédécarrats. 1998. Ecotype differentiation and coexistence of two parapatric tetraploid subspecies of cocksfoot (*Dactylis glomerata*) in the Alps. *New Phytologist* 139: 741–750.
- Geber, M. A., and V. M. Eckhart. 2005. Experimental studies of adaptation in *Clarkia xantiana*. II. Fitness variation across a subspecies border. *Evolution* 59:521–531.
- Gilman, S. E., M. C. Urban, J. Tewksbury, G. W. Gilchrist, and R. D. Holt. 2010. A framework for community interactions under climate change. *Trends in Ecology and Evolution* 25:325–331.
- 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.
- . 2006. Is evolution necessary for range expansion? manipulating reproductive timing of a weedy annual transplanted beyond its range. *American Naturalist* 167:153–164.
- Griggs, R. F. 1914. Observations on the behavior of some species at the edges of their ranges. *Bulletin of the Torrey Botanical Club* 41:25–49.
- Hargreaves, A. L., and C. G. Eckert. 2013. Evolution of dispersal and mating systems along geographic gradients: implications for shifting ranges. *Functional Ecology* (forthcoming), doi:10.1111/1365-2435.12170.
- Hargreaves, A. L., K. E. Samis, and C. G. Eckert. 2014. Data from: Are species' range limits simply niche limits writ large? a review of transplant experiments beyond the range. *American Naturalist*, Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.c3287>.
- Holt, R. D. 2003. On the evolutionary ecology of species' ranges. *Evolutionary Ecology Research* 5:159–178.
- Holt, R. D., and T. H. Keitt. 2000. Alternative causes for range limits: a metapopulation perspective. *Ecology Letters* 3:41–47.
- Holt, R. D., T. H. Keitt, M. A. Lewis, B. A. Maurer, and M. L. Taper. 2005. Theoretical models of species' borders: single species approaches. *Oikos* 108:18–27.
- Jump, A., C. Mátyás, and J. Peñuelas. 2009. The altitude-for-latitude disparity in the range retractions of woody species. *Trends in Ecology and Evolution* 24:694–701.
- Kawecki, T. J. 2008. Adaptation to marginal habitats. *Annual Review of Ecology, Evolution, and Systematics* 39:321–342.
- Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. *Ecology Letters* 7:1225–1241.
- Kirkpatrick, M., and N. H. Barton. 1997. Evolution of a species' range. *American Naturalist* 150:1–23.
- Klimeš, L., and J. Doležal. 2010. An experimental assessment of the upper elevational limit of flowering plants in the western Himalayas. *Ecography* 33:590–596.
- Latimer, A. M., J. A. Silander, A. G. Rebelo, and G. F. Midgley. 2009. Experimental biogeography: the role of environmental gradients in high geographic diversity in Cape Proteaceae. *Oecologia* (Berlin) 160:151–162.
- Leimu, R., and M. Fischer. 2008. A meta-analysis of local adaptation in plants. *PLoS ONE* 3:1–8.
- Levin, D. A., and K. Clay. 1984. Dynamics of synthetic *Phlox drummondii* populations at the species margin. *American Journal of Botany* 71:1040–1050.
- MacArthur, R. 1972. *Geographical ecology: patterns in the distribution of species*. Princeton University Press, Princeton, NJ.
- Manly, B. 1997. *Randomization, bootstrap and Monte Carlo methods in biology*. Chapman & Hall, London.
- Mousseau, T., and C. Fox. 1998. The adaptive significance of maternal effects. *Trends in Ecology and Evolution* 13:403–407.
- Normand, S., R. E. Ricklefs, F. Skov, J. Bladt, O. Tackenberg, and J. C. Svenning. 2011. Postglacial migration supplements climate in determining plant species ranges in Europe. *Proceedings of the Royal Society B: Biological Sciences* 278:3622–3653.
- Normand, S., U. A. Treier, C. Randin, P. Vittoz, A. Guisan, and J. C. Svenning. 2009. Importance of abiotic stress as a range-limit determinant for European plants: insights from species responses to climatic gradients. *Global Ecology and Biogeography* 18:437–449.
- Parmesan, C., S. Gaines, L. Gonzalez, D. M. Kaufman, J. Kingsolver, A. Townsend Peterson, and R. Sagarin. 2005. Empirical perspectives on species borders: from traditional biogeography to global change. *Oikos* 108:58–75.
- Phillips, B. L. 2012. Range shift promotes the formation of stable range edges. *Journal of Biogeography* 39:153–161.
- Poll, M., B. J. Naylor, J. M. Alexander, P. J. Edwards, and H. Dietz. 2009. Seedling establishment of Asteraceae forbs along altitudinal gradients: a comparison of transplant experiments in the native and introduced ranges. *Diversity and Distributions* 15:254–265.
- Pulliam, H. R. 2000. On the relationship between niche and distribution. *Ecology Letters* 3:349–361.
- Qian, H., and R. E. Ricklefs. 2007. A latitudinal gradient in large-scale beta diversity for vascular plants in North America. *Ecology Letters* 10:737–744.
- R Core Team 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.r-project.org>.
- Rohde, R. 2012. Global Warming Art, accessed February 2013. <http://www.globalwarmingart.com>.
- 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.
- 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.
- Schemske, D. 2009. Biotic interactions and speciation in the tropics. Pages 219–239 in R. K. Butlin, J. R. Bridle, and D. Schluter, eds. *Speciation and patterns of diversity*. Cambridge University Press, Cambridge.
- Sekercioglu, C. H., S. H. Schneider, J. P. Fay, and S. R. Loarie. 2008. Climate change, elevational range shifts, and bird extinctions. *Conservation Biology* 22:140–150.
- 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.
- Sokal, R., and F. Rohlf. 1995. *Biometry: the principles and practice of statistics in biological research*. W. H. Freeman, New York.



- Stinson, K. A. 2005. Effects of snowmelt timing and neighbor density on the altitudinal distribution of *Potentilla diversifolia* in western Colorado, USA. *Arctic Antarctic and Alpine Research* 37:379–386.
- Svenning, J. C., S. Normand, and F. Skov. 2008. Postglacial dispersal limitation of widespread forest plant species in nemoral Europe. *Ecography* 31:316–326.
- Vergeer, P., and W. E. Kunin. 2013. Adaptation at range margins: common garden trials and the performance of *Arabidopsis lyrata* across its northwestern European range. *New Phytologist* 197:989–1001.
- Woodward, F. I. 1990. The impact of low temperatures in controlling the geographic distribution of plants. *Philosophical Transactions of the Royal Society B: Biological Sciences* 326:585–593.
- Woodward, F. I., and C. D. Pigott. 1975. Climatic control of altitudinal distribution of *Sedum rosea* (L) and *Sedum telephium* L. I. Field observations. *New Phytologist* 74:323–334.

Associate Editor: Oswald J. Schmitz  
Editor: Judith L. Bronstein



Upper elevational range limit of beech trees in Patagonia, Argentina, clearly showing the influence of climate. Photograph by Anna Hargreaves.