ECOLOGY LETTERS

Ecology Letters, (2012) doi: 10.1111/j.1461-0248.2012.01796.x

LETTER

Climate envelope modelling reveals intraspecific relationships among flowering phenology, niche breadth and potential range size in *Arabidopsis thaliana*

Joshua A. Banta,[1*] Ian M. Ehrenreich,[2] Silvia Gerard[3] Lucy Chou,[4] Amity Wilczek,[5] Johanna Schmitt,[6] Paula X. Kover,[7] and Michael D. Purugganan[8]

Abstract

Species often harbour large amounts of phenotypic variation in ecologically important traits, and some of this variation is genetically based. Understanding how this genetic variation is spatially structured can help to understand species' ecological tolerances and range limits. We modelled the climate envelopes of *Arabidopsis thaliana* genotypes, ranging from early- to late-flowering, as a function of several climatic variables. We found that genotypes with contrasting alleles at individual flowering time loci differed significantly in potential range size and niche breadth. We also found that later flowering genotypes had more restricted range potentials and narrower niche breadths than earlier flowering genotypes, indicating that local selection on flowering can constrain or enhance the ability of populations to colonise other areas. Our study demonstrates how climate envelope models that incorporate ecologically important genetic variation can provide insights into the macroecology of a species, which is important to understand its responses to changing environments.

Keywords

Arabidopsis thaliana, climate envelope, climate envelope modelling, climate niche, genetic diversity, geographical distribution, niche breadth, niche modelling, niche width, potential range size.

Ecology Letters (2012)

INTRODUCTION

Understanding species distributions is one of the central goals of ecology (May 1999). This requires an understanding of species' environmental requirements, known as the fundamental niche (Hutchinson 1957), and species' tolerance to varying environments, known as the niche breadth (Levins 1968). A species' habitat tolerances, in turn, determine its potential range size (Svenning & Skov 2004; Paul *et al.* 2009). It is generally expected that a species with a broad niche breadth should have a large potential range, because it can tolerate a wide variety of environments, whereas a species with a narrow niche breadth should have a small potential range, because it can only tolerate a small range of environments (Brown 1984).

The niche breadth can be measured experimentally by determining a species' tolerances to various environmental axes, measured under controlled environmental conditions (e.g. Garbutt & Bazzaz 1987), or inferred from field trials (Fournier-Level *et al.* 2011; Rehfeldt *et al.* 1999). But these approaches are logistically complex: a myriad of environmental axes must be tested, each across a range of values. Such an endeavour is not always feasible. Climate envelope modelling (a form of ecological niche modelling) is an alternative approach to quantify niche breadths and delimit potential ranges that relies on observational data rather than on experimental manipulations (Svenning & Skov 2004; Thomas *et al.* 2004; Hijmans *et al.* 2005; Warren *et al.* 2008). It assumes that climatic factors are an important component of a species' environmental tolerances and preferences across its range (Svenning & Skov 2004). The idea is to identify correlations between the presence of a species (from survey data) and climatic factors, and to use this information to predict other suitable areas where the species could live (Paul *et al.* 2009). Climate envelope modelling has been used to quantify the ranges and potential ranges of hundreds of species across a wide variety of taxa (Thomas *et al.* 2004).

Climate envelope modelling has traditionally been employed at the whole-species level, assuming common intraspecific environmental affinities and tolerances (Bolnick *et al.* 2003). But this assumption is contradicted by local adaptation, which causes habitat affinities and tolerances to diverge; in fact, it has long been recognised that members of a species from different areas are not necessarily ecologically equivalent (Clausen *et al.* 1940). Therefore, climate envelope modelling at the species level masks the known observation that climatic tolerances vary intraspecifically (as in Eckert *et al.* 2009; Fournier-Level *et al.* 2011; Hancock *et al.* 2011). If individuals within species vary in their climate envelopes, then the potential range sizes inferred from the climate envelopes will also vary intraspecifically. This is especially important when considering species'

[1]*Department of Biology, University of Texas at Tyler, 3900 University Blvd., Tyler, TX, 75799, USA*
[2]*Section of Molecular and Computational Biology, University of Southern California, Los Angeles, CA, 90089, USA*
[3]*Department of Systems Biology, Harvard Medical School, Boston, MA, 02115, USA*
[4]*Animal Medical Center, 510 E. 62nd Street, New York, NY, 10065, USA*
[5]*Deep Springs College, Big Pine, CA, 93513, USA*
[6]*Department of Ecology and Evolutionary Biology, Brown University, Providence, RI, 02912, USA*
[7]*Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY, UK*
[8]*Department of Biology and Center for Genomics and Systems Biology, New York University, New York, NY, 10003, USA*
[*]*Correspondence: Email: jbanta@uttyler.edu*

range shifts in response to climate change, as the response is an ensemble of individual responses, or when designing biological reserves, as the suitability of a particular habitat may depend on the populations being considered.

Climate envelope models at the intraspecific level are more challenging because the unit of analysis is not as clear. Populations, for instance, lack replication. Fortunately, recent studies (Fournier-Level *et al.* 2011; Eckert *et al.* 2009; Hancock *et al.* 2011) have demonstrated that genotypes are a relevant unit for climate envelope modelling. Unlike populations, the same genotypes can be found in multiple locations. Furthermore, the genotypes can be selected based on their ecological differentiation from one another, maximising the variation present in the dataset and allowing for intraspecific comparisons between macroecological variables and important phenotypes. Thus, large-scale ecological properties can be linked to phenotypic variation within a species.

In this study, we modelled the climate envelopes of the model plant species *Arabidopsis thaliana* to understand the intraspecific relationships between niche breadth, potential range size, and an important plant phenotype (flowering time). The relationship between the niche breadth and potential range size has been studied on an interspecific level (e.g. Hoffmann 2005; Köckemann *et al.* 2009), but the degree to which these patterns may be labile to genetic variation within species has rarely been explored, and general axioms are lacking (Angert *et al.* 2011). *A. thaliana* provides an ideal opportunity to study intraspecific macroecological variation. In addition to having a long and productive history as a model genetic organism (Koorneef & Meinke 2010), *A. thaliana* is extremely ecologically diverse, occupying climates ranging from hot, sunny, semi-arid areas in the Iberian peninsula to cold, cloudy and wet environments above the Arctic Circle in Scandanavia, and found at a variety of altitudes (Hoffmann 2002; Montesinos-Navarro *et al.* 2011; Figs 1 and 2). Furthermore, because of *A. thaliana*'s genetic tractability, it is possible to identify individuals with genetically based ecological differences. We focused on flowering timing in *A. thaliana*, because this phenotype has a genetic basis (Ehrenreich *et al.* 2009), shows evidence of adaptation to climatic conditions (Montesinos-Navarro *et al.* 2011), and is associated with fitness under field conditions (Korves *et al.* 2007).

We sought to understand the separate environmental tolerances, and potential geographical ranges, of individual genotypes of *A. thaliana* that are associated with differences in flowering time, an important life history trait that serves as a marker for ecological differentiation. Our study specifically addresses the following questions: (1) Do single-locus genotypes that are ecologically differentiated also differ in their niche breadths and potential range sizes, and if so, which genotypes have the larger niche breadths and potential range sizes? (2) Are niche breadth and potential range size correlated with each other on an intraspecific level (i.e. using single-

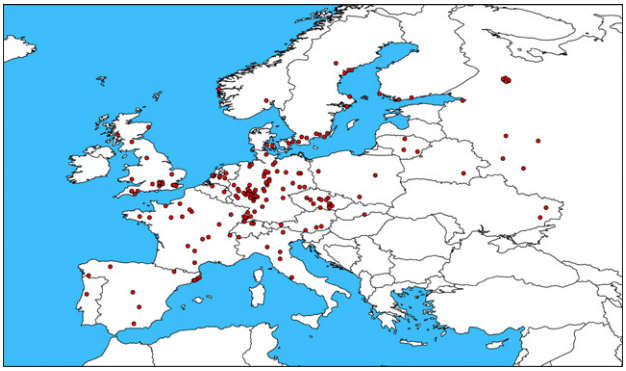

Figure 1 Map showing the collection locations of the natural accessions used for the landscape genetics analyses (313 total).

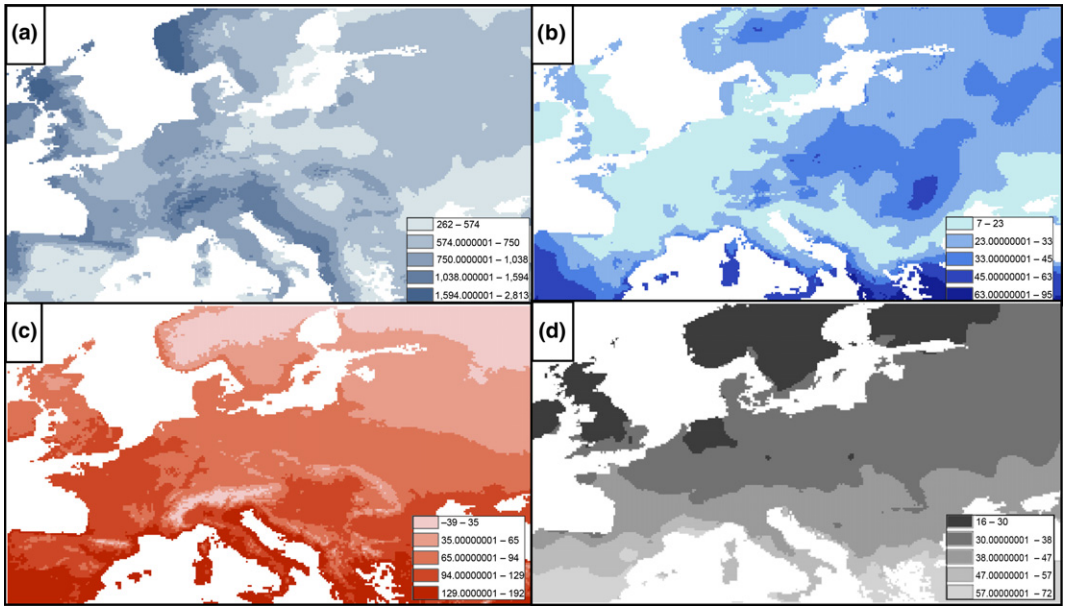

Figure 2 The climatic layers used in this study: (a) total annual precipitation (mm), (b) precipitation seasonality, (c) average annual temperature (in °C multiplied by 10) and (d) average annual cloud cover (percentage of total possible sunlight). Darker shades represent more precipitation, greater precipitation seasonality, higher temperatures and more cloud cover.

locus genotypes as the unit of analysis)? and (3) Does the way potential range size is measured affect the patterns observed?

MATERIALS AND METHODS

Plant material

We used two groups of plants in this study. One group was an artificial population that was a genetic mosaic of 19 naturally occurring individuals from throughout *A. thaliana*'s broad natural range in Europe (multi-parent advance generation intercross lines, or MAGIC lines; described in Kover *et al.* 2009). This population was used to identify (map) naturally occurring single-locus genotypes that differ in flowering phenology. The other group of plants were naturally collected individuals (referred to as natural accessions) from across this same range (Fig. 1), and spanning many different climatic regimes (Fig. 2). The natural accessions were subdivided based on their single-locus genotypes at the loci that were associated with flowering time variation in the mapping study; these genotypes were used to model intraspecific variation in the niche breadths and potential range sizes of *A. thaliana* on the landscape. There were a total of 327 distinct MAGIC lines and 313 natural accessions used in this study (Table S1).

Single-locus genotypes

We used haplotype-tagging SNPs (htSNPs) developed by Ehrenreich *et al.* (2009) to identify loci with alleles causing significant differences in flowering time. These loci were in 47 candidate flowering time genes (see Tables S2 and S5). See the supplementary information for more details.

Flowering Phenotypes

We grew 10 replicates of each MAGIC line under long-day conditions (14-h light, 10-h dark) at 20 °C in EGC walk-in environmental chambers using a fully randomised design. We rotated the plants weekly between two different growth chambers to average out any systemic whole-chamber effects. There were an average of 9.05 replicates (SD = 1.20) that survived until bolting and were included in subsequent analyses. We phenotyped the MAGIC lines for bolting time (the initiation of the elongation of the flowering shoot, marked by the presence of flower buds) and rosette leaf number (a well-established morphological proxy for flowering time) under long-day conditions.

Association mapping

We tested separately for an effect of each locus on flowering time, using the model $\boldsymbol{y} = \boldsymbol{X}\alpha + \varepsilon$, where $\boldsymbol{y}$ is a vector of mean flowering times for each MAGIC line, $\boldsymbol{X}$ is a matrix of single-locus genotypes, α is a vector of allele effects to be estimated and ε is a vector of residual errors. The data were standardised to a mean of zero and a standard deviation of one, and analysed using SAS version 9.1.3 PROC GLM (SAS Institute, Inc., Cary, NC, USA). We did not test for non-additive effects among loci. To control the false discovery rate (FDR) introduced by conducting multiple simultaneous tests, we calculated *Q*-values (FDR-corrected *P*-values) using the QVALUE package (Dabney & Storey 2010) in R version 2.11.1 (R Development Core Team 2010).

Climate envelope modelling

The geographical coordinates of the accessions' locations in the wild were determined from the information at the Arabidopsis Seed Stock Center (http://www.arabidopsis.org), and are presented in Table S1. The extent of our analysis was an area between −8.4° to 38.4° E and 37.2° and 63.7° N (Fig. 2).

Climatic data were based on 30–40 year averages at the 10-min scale, and came from either the Climate Research Unit (New *et al.* 2002) or WorldClim (Hijmans *et al.* 2005). We chose environmental variables likely to affect flowering time, specifically, variables related to the number of frost days, cloud cover, temperature and precipitation (Table S3 and Fig. 1). We chose temperature and precipitation because of their documented effects on flowering time (Balasubramanian *et al.* 2006) and because of other evidence of their importance to *A. thaliana*'s distribution (Hoffmann 2005); we used the number of frost days as a proxy for the strength of the floral-inducing vernalisation cue (Wilczek *et al.* 2009), and also because frost can select against early flowering (Inouye 2000); and we chose cloud cover because it affects light availability and the red:far-red light ratio (Reinhardt *et al.* 2010), which affects flowering time (Callahan & Pigliucci 2002). After filtering the list of climatic variables to include only those correlated less than 0.65 (Table S3), the final variables included in the climate envelope models were: average annual temperature, average annual precipitation, precipitation seasonality and average annual cloud cover (Fig. 2).

To predict the habitat suitability of different areas on the landscape for each single-locus genotype, we employed a climate envelope modelling approach known as maximum entropy distribution using the software Maxent (Dudik *et al.* 2010). Maximum entropy distribution modelling predicts the specific environmental requirements of a taxon, and scores the suitability of every grid cell (quadrat) for that taxon in the defined area; the precision vs. coarseness of the analysis depends on the grain size of the environmental information. Habitat suitability scores (logistic scores) range from zero to one, with one being the highest, and rate the favourability of the habitat for the taxon, based on the climatic values in that grid cell (Phillips & Dudik 2008). This approach is used for species, but has only recently been used to look at intraspecific niche variation (Fournier-Level *et al.* 2011). Maximum entropy distribution modelling outperforms a variety of alternative methods and performs well at small sample sizes (Wisz *et al.* 2008).

We binned individuals together based on their contrasting genotypes at loci affecting flowering time, presumably reflecting adaptive differentiation to different environments. Thus, we modelled individuals belonging to every single-locus genotype separately; we did this by using the information about where individuals with specific single-locus genotypes occur on the landscape. To correct for bias in the sampling intensity of accessions (and, therefore, of single-locus genotypes), we created a continuous (kernel) density function, based on the collections of our samples, using ArcMap (ESRI Inc 2008) (Fig. S1), and included this layer as a 'bias file' in the Maxent options.

We used two metrics to evaluate model fit: AUC and gain. AUC, the area under the operator receiving curve, measures the probability that a randomly chosen presence site will be ranked above a randomly chosen pseudoabsence site (Phillips & Dudik 2008). Gain is the average log probability of the presence samples, minus a constant that makes the uniform distribution have zero gain (Phillips

2005). Models with AUC > 0.75 are traditionally thought of as useful (Elith 2002). To determine the importance of each climatic variable's contribution to the model, we compared model gains with and without each variable included.

Intraspecific relationships

We tested whether homozygous genotypes with contrasting alleles at the same locus had different niche breadths and potential range sizes from each other. We calculated niche breadths using a standardised version of Levins (1968) 'inverse concentration' metric, described by Warren *et al.* (2008):

$$\frac{\frac{1}{\sum_{i=1}^{n}\left(\frac{p_i}{\sum_{i=1}^{n} p_i}\right)^2} - 1}{n},$$

where p_i is the habitat suitability score predicted by Maxent for grid cell i, and n is the total number of grid cells, given the extent and grain size of the climatic layers used in the analysis. The niche breadth values range from 0 to 1, with 1 indicating that different habitats are equally suitable (the maximum possible niche breadth), and values of 0 indicating that some habitats are much more suitable than others (the narrowest possible niche breadth).

The estimation of potential geographical ranges is complicated (Gaston 1994; Paul *et al.* 2009). It involves evaluating habitats where the species is not yet found (Paul *et al.* 2009). The typical approach is to make decisions about whether individual habitat patches are potentially habitable or not (Svenning & Skov 2004; Paul *et al.* 2009; Elith *et al.* 2011). But range margins tend to be gradational rather than sharp (Gaston 1994). Therefore, we introduce a new approach for estimating the potential range size, which more accurately conforms to the gradational nature of species boundaries. We compared this to the more traditional approach based on a discrete habitability/inhabitability.

The traditional approach is to define potential range size as either its areas of potential occupancy or as its overall extent of potential occurrence. The former approach tallies the locations where the species could potentially live, whereas the latter approach measures the total area within which the species could live, ignoring uninhabitable spaces in the interior (Gaston 1994). Because it excludes uninhabitable patches, the 'area of potential occupancy' approach is more discriminating than the 'extent of potential occurrence' approach, but at the same time it also more sensitive to the grain size of the analysis (Gaston 1994; Welk & Bruelheide 2006). At finer grain sizes, however, the estimates should converge (Gaston 1994; Welk & Bruelheide 2006). Therefore, because the scale of our study was fine-grained (10′ × 10′ grid resolution; see Welk & Bruelheide 2006), we used an 'area of potential occupancy' approach to determine ranges.

To measure potential range size as the area of potential occupancy, we chose a threshold of habitat suitability that we deemed habitable. While Elith *et al.* (2011) recommend a default threshold of 0.5 in the absence of better information, we used 0.4 instead, because it is possible that poorer quality habitat is still sufficient (Phillips & Dudik 2008). We calculated potential range size by simply tallying the number of grid cells with habitat suitability scores greater than the threshold value (0.4).

The new approach to delimiting potential range size, which we introduce here for the first time, uses the Maxent output to calculate the median habitat suitability score for each single-locus genotype across the landscape. As all of our climate envelope models had exactly the same number of grid cells (quadrats), it was possible to directly compare these scores to one another and compare potential range sizes. However, if the number of grid cells had been different among the different models, the potential range size estimates could have been made comparable to one another by multiplying the median scores by the total area of the models; thus the climate envelope models covering a smaller area would be discounted relative to ones that cover a larger area, all else being equal.

To test whether single-locus genotypes with contrasting alleles at the same locus had different niche breadths and potential range sizes from each other, we used permutation tests to calculate the 95% confidence interval for the null expectation of no difference in the statistic between the allelic genotypes; we then compared whether the observed difference in the statistic lay outside the permuted 95% CI.

We tested whether flowering phenology is associated with niche breadth or potential range size, and whether niche breadth is associated with potential range size. In separate analyses, we regressed potential range size (measured as either the median habitat suitability score or the number of quadrats meeting the minimum habitat quality score) on flowering time (measured as either days to bolting or number of rosette leaves), potential range size and niche breadth. To compare flowering time with niche breadth, and to compare the two potential range size estimates to one another, we used correlation instead of regression analysis.

RESULTS

Association mapping

Using the artificial MAGIC lines, we identified 23 pairs of homozygous genotypes that significantly differ in their flowering times from their allelic counterparts at the same genetic locus; thus there were 46 genotypes total. We then filtered this list of genotypes after looking at patterns of linkage disequilibrium among the genetic loci (see the supplementary information for more details), leaving 15 single-locus genotype pairs (30 genotypes total). These single-locus genotypes were based on allelic polymorphisms within one of 12 different flowering time genes (Tables S2 and S5; Fig. S2). These same single-locus genotypes are found within the naturally collected accessions from Europe (Fig. 1; Table S1), and they were used as the basis for subdividing the accessions for intraspecific climate envelope modelling.

Climate envelope modelling

All climate envelope models had training AUC values ranging from 0.75 to 0.92, indicating that they conferred substantially higher habitat suitabilities to locations where the respective genotypes occurred than where they were not observed (Table 1). The relative contributions of the different climatic variables to the climate envelope models (as measured by test gain when the model only included that particular environmental variable) varied depending on the particular model. For instance, precipitation seasonality, average annual

Table 1 Summary information for the individual climate envelope models, which were performed separately on each single-locus genotype (rows)

Locus	Allele	Training AUC	Test AUC	Test gain	Test gain with only precipitation seasonality	Test gain with only annual precipitation	Test gain with only annual mean cloud cover	Test gain with only annual mean temperature
FLC^{-3312}	G	0.78	0.74	0.28	0.10	− 0.015	0.11	0.10
FLC^{-3312}	T	0.83	0.77	0.57	0.28	0.12	0.29	0.21
FLC^{7207}	A	0.84	0.78	0.62	0.29	0.041	0.33	0.19
FLC^{7207}	C	0.80	0.75	0.34	0.078	0.038	0.14	0.030
$FRL1^{1127}$	G	0.83	0.78	0.55	0.27	0.066	0.27	0.19
$FRL1^{1127}$	T	0.84	0.76	0.36	0.0064	− 0.0056	0.30	0.40
$GA1^{7762}$	A	0.80	0.71	0.29	− 0.011	0.0062	0.04	− 0.0082
$GA1^{7762}$	T	0.82	0.77	0.55	0.29	0.051	0.28	0.18
$HUA2^{5106}$	A	0.80	0.72	0.31	0.16	< 0.001	0.17	0.12
$HUA2^{5106}$	T	0.92	0.89	1.28	0.58	0.21	0.31	0.47
$PFT1^{1593}$	G	0.83	0.78	0.65	0.33	0.12	0.27	0.22
$PFT1^{1593}$	T	0.75	0.70	0.19	− 0.017	0.016	0.22	− 0.0067
$PHYD^{2446}$	A	0.86	0.79	0.52	0.16	0.10	0.067	0.034
$PHYD^{2446}$	G	0.81	0.76	0.50	0.26	0.021	0.26	0.18
$PIE1^{906}$	A	0.85	0.81	0.64	0.17	0.13	0.42	0.21
$PIE1^{906}$	G	0.81	0.75	0.45	0.22	0.034	0.19	0.13
RGA^{1023}	C	0.82	0.75	0.44	0.25	0.055	0.21	0.16
RGA^{1023}	T	0.80	0.75	0.39	0.11	− 0.027	0.23	0.064
$SOC1^{651}$	G	0.83	0.77	0.56	0.33	0.11	0.16	0.16
$SOC1^{651}$	C	0.82	0.78	0.47	0.020	0.025	0.38	0.19
$SOC1^{2742}$	A	0.89	0.85	1.00	0.51	0.26	0.32	0.27
$SOC1^{2742}$	T	0.80	0.73	0.40	0.16	− 0.021	0.24	0.16
$TFL1^{687}$	A	0.86	0.80	0.64	0.32	0.092	0.28	0.24
$TFL1^{687}$	C	0.75	0.68	0.17	0.088	− 0.012	0.10	− 0.021
$TFL2^{890}$	G	0.82	0.79	0.59	0.28	0.10	0.14	0.12
$TFL2^{890}$	T	0.84	0.78	0.52	0.27	0.021	0.30	0.14
$TFL2^{2993}$	A	0.81	0.74	0.38	0.15	0.023	0.19	0.10
$TFL2^{2993}$	G	0.86	0.82	0.81	0.40	0.12	0.41	0.40
TSF^{2661}	A	0.85	0.81	0.65	0.34	0.10	0.11	0.19
TSF^{2661}	G	0.83	0.76	0.50	0.20	0.047	0.27	0.21

cloud cover and average annual temperature contributed roughly equally to the climate envelope model of the single-locus genotype $FLC^{-3312(G)}$ (see the supporting information for an explanation of how genotypes were named). In contrast, average annual cloud cover contributed the most information to the climate envelope model for the single-locus genotype $PIE1^{906(A)}$.

In most cases, the niche breadths and potential range sizes of allelic single-locus genotypes were different from each other, as indicated by permutation tests (Figs. 3 and S4; Table S6). Late flowering alleles were associated with smaller potential range sizes regardless of how flowering time was measured (days to bolting or rosette leaf number) and whether potential range size was estimated using the median habitat suitability score (days to bolting: $r^2 = 0.16$, $P = 0.0031$; number of rosette leaves: $r^2 = 0.18$, $P = 0.0019$; Fig. 4a,b) or the number of habitable grid cells (days to bolting: $r^2 = 0.10$, $P = 0.031$; number of rosette leaves: $r^2 = 0.12$, $P = 0.019$; Fig. 4d,e). The two different estimates of potential range size were also highly correlated with one another ($r = 0.67$; $P < 0.001$). Furthermore, the relationship between potential range size and niche breadth was quite strong, and was also robust to the potential range size estimator (median habitat suitability score vs. niche breadth: $r^2 = 0.86$, $P < 0.001$; number of habitable cells vs. niche breadth: $r^2 = 0.73$, $P < 0.001$; c.f. Fig. 4). Finally, flowering time (however measured) was inversely correlated with niche breadth (bolting date and niche breadth: $r = -0.33$, $P = 0.027$; number of rosette leaves and niche breadth: $r = -0.34$, $P = 0.021$).

DISCUSSION

Differential adaptation to environmental conditions is a major factor shaping the spatial structure of species, but this process also plays an important role in how individuals are spatially distributed within a species. We used climate envelope models to understand the intraspecific relationships between two macroecological phenomena (niche breadth and potential range size). Previous work has shown that niche breadth and range size are correlated interspecifically within the *Arabidopsis* genus (Hoffmann 2005). Here, we report on that pattern (and others) on an intraspecific level. The key to our approach was the use of genotypes as the unit of analysis. We used single-locus genotypes that have contrasting effects on flowering time – an important plant phenotype associated with ecological differentiation – and in the aggregate these genotypes formed the intraspecific variation upon which we based our analyses.

The climate envelope models presented here suggest that climate plays a strong role in determining the various geographical distributions of the single-locus genotypes. Biotic factors (pathogens, herbivores, and competitors), however, clearly also play an important role in species distributions (Meier *et al.* 2010). To the extent that biotic factors are correlated with climate, modelling tolerances to climate indirectly takes into account tolerances to biotic factors; but biotic factors can also be orthogonal to climate (Meier *et al.* 2010). Global climate maps are readily available at a variety of scales (New *et al.* 2002; Hijmans *et al.* 2005), but future work is needed to

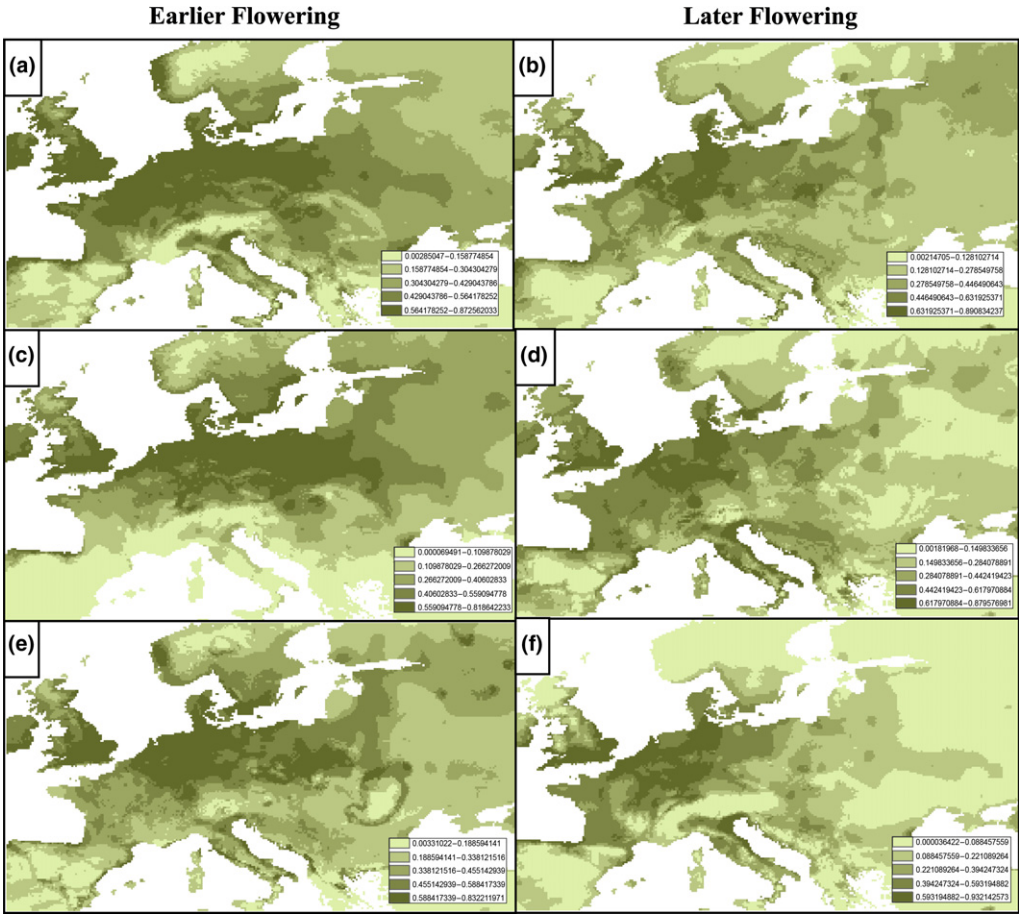

Figure 3 Climate envelope models for the genotypes $FLC^{3312(G)}$ and $FLC^{3312(T)}$ (a and b), $SOC1^{651(G)}$ and $SOC1^{651(C)}$ (c and d) and $HUA2^{5106(A)}$ and $HUA2^{5106(T)}$ (e and f). The genotypes on the left are earlier flowering, and the corresponding genotypes on the right are later flowering. The colorisation scheme represents the predicted habitat suitability for the genotypes across the landscape. The darkest shade represents the most suitable habitat (logistic scores near 1) and the lightest shade indicates the least suitable habitat (logistic scores near 0). The rest of the climate envelope models can be seen in Fig. S3. See the supplemental information for details on the nomenclature of the genotypes.

develop continental-scale maps of biotic factors, with resolution similar to that of the climate maps, so that these factors can also be incorporated into large-scale ecological niche models such as ours.

Intraspecific variation in niche breadths and potential range sizes

Our results suggest that individuals with small genetic differences can have different niche breadths and potential range sizes, provided those genetic differences are associated with adaptive differentiation in an ecologically important phenotype such as flowering time. For example, genotypes $SOC1^{651(G)}$ and $SOC1^{651(C)}$, while differing only in their nucleotide sequences in the region around position 651 within the *SOC1* gene (relative to our alignment; see Table S2 and the supplemental information), have significantly different niche breadths and potential range sizes from one another (Fig. 3c and Table S6). Thus, we concur with other studies that suggest species' responses to environmental changes (such as geographical clines or climate change) are best understood on an intraspecific level (Fournier-Level *et al.* 2011; Rehfeldt *et al.* 1999; Angert *et al.* 2011; Hancock *et al.* 2011). Models of species' responses to environmental change that assume equivalency of individuals will incorrectly assume that certain areas of the landscape are inhabitable/habitable. For instance, some areas of the landscape could be deemed inhabitable because countervailing associations between genotypes and environmental factors are averaged out, and some areas of the landscape could be deemed habitable even if the appropriate genotypes cannot reasonably establish there due to geographical barriers or slow dispersal rates. For example, in *A. thaliana* there are areas of Spain that are projected contain suitable habitat for some genotypes and not for others (compare $SOC1^{651(G)}$ and $SOC1^{651(C)}$ in Fig. 3c, d, for instance). This is information that would have been washed out at the whole-species level, although whether or not the appropriate genotypes actually exist at those locations or could potentially distribute there would need to be investigated.

Our results also suggest that intraspecific variation in niche breadths and potential range sizes are associated with phenological variation in flowering: earlier flowering genotypes had larger niche breadths and potential range sizes than later flowering genotypes (Fig. 4a,b,d,e). In other words, earlier flowering genotypes are projected to have a more even habitat-affinity profile than later flowering genotypes, and they can also potentially inhabit a larger area of the landscape. There are several possible explanations for this pattern, which warrant further investigation. One possibility is that later-flowering (and slower-growing) genotypes are adapted to

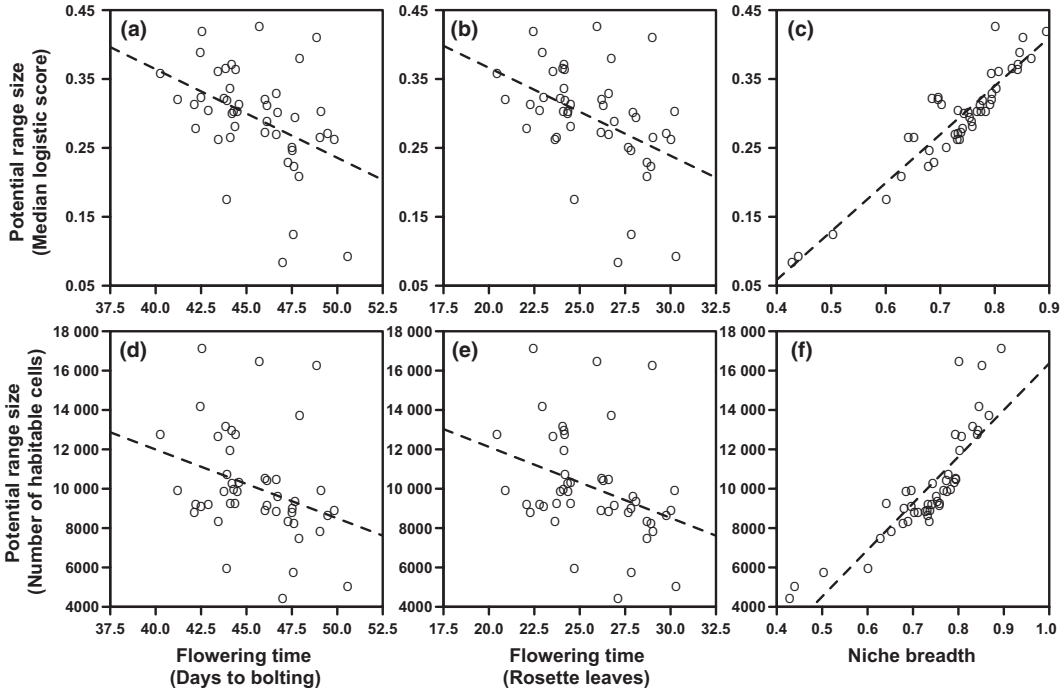

Figure 4 Potential range size vs. flowering time (a, b, d and e) and potential range size vs. niche breadth (c and f). Potential range size was measured as either the median habitat suitability (a–c) or the number of habitable cells (d–f). Flowering time was measured as either the number of days to bolting (a and d) or the number of rosette leaves at bolting (b and e). The dotted lines indicate the slopes of the best-fit linear regressions.

harsher environments, where it is necessary to flower later, and are competitively excluded from more favourable environments by earlier-flowering (and faster-growing) genotypes; this appears to be the case for *Pinus contorata* in western North America (Rehfeldt *et al.* 1999). Another possibility is that early flowering individuals track the spread of human agriculture, environments that favour early-flowering ecotypes (Toomajian *et al.* 2006). However, another possibility is that early flowering is a drought-avoidance adaptation, allowing survival in more marginal warm environments (Sexton *et al.* 2011). Finally, it is possible that that early-flowering genotypes are simply better dispersers (because of shorter generation times) and that propagule pressure is more important than adaptation in determining range limits (Vellend *et al.* 2007). Regardless of the precise reasons for these patterns, they show how the evolution of flowering time can be constrained or enhanced because of landscape-level properties: local selection for later flowering has the correlated consequence of restricting the potentially colonisable space for the population; conversely, local selection for earlier flowering opens up the opportunity for those populations to spread geographically. Thus, local selection for earlier flowering can influence the range limits of the entire species by favouring genotypes that are then able to colonize areas outside of their area of origination.

The intraspecific relationship between niche breadth and potential range size

We found niche breadth and potential range size to be highly intraspecifically correlated, regardless of how potential range size is measured (Fig. 4c,f). In other words, genotypes that utilise different climate spaces more evenly also have larger potential distributions. This matches theoretical expectations at the species level, laid out by Brown (1984), that species with a broad niche breadth should have a large potential range, because they can tolerate a wide variety of environments, whereas species with a narrow niche breadth should have a small potential range, because they can only tolerate a small range of environments. Thus, not only does the *Arabidopsis* genus follow this pattern interspecifically, but one member of this genus, *A. thaliana*, follows this pattern intraspecifically as well.

We found that our two different potential range size metrics (the number of habitable cells meeting a certain threshold or the median habitat suitability of the landscape) were highly correlated with one another ($r = 0.67$; $P < 0.001$). In fact, the two sets of results based on the respective metrics were practically interchangeable. On the one hand, the congruence between the two metrics suggests that they are measuring the same thing, which lends authority to our new approach. On the other hand, we caution that this may not always be the case, and so it should not be taken for granted. A potential range size estimate based on the number of habitable cells may change in nonlinear ways, depending on the distribution of climatic and other relevant factors on the landscape and on the suitability threshold deemed habitable. This brings up some related problems, namely that it remains unclear what the appropriate habitat suitability threshold should be, at least in the absence of empirical calibration (Elith *et al.* 2011), and that a strict threshold is not biologically realistic anyway (Gaston 1994). For these reasons, we recommend that median habitat suitability be used to estimate potential range size whenever comparing several potential range sizes to one another. Of course, this approach will not yield straightforward answers if the size of potential range is needed in absolute terms (e.g. if a specific area of space must delimited for conservation purposes); in that case, a threshold-based measurement of potential range size must be used, and the threshold should be well justified.

The relationship between potential range size and the realised range size

The patterns we have identified here refer to the potential geographical range size as opposed to the realised range size. Although the latter is clearly of great interest, understanding, let alone predicting, realised range limits based on potential range limits would require natural history information about dispersal, contingent events and geographical barriers (Hoffmann 2005; Paul *et al.* 2009). But fortunately, some information about the biogeographical history of *A. thaliana* across its native range is available (Hoffmann 2002, 2005; Toomajian *et al.* 2006; François *et al.* 2008), and so the relationship between the potential range and the realised range could be explored in the future. Furthermore, *A. thaliana*'s is widespread in North America, where it is was introduced within the last few centuries (Hoffmann 2002, 2005), and this affords the opportunity to study the interplay of colonisation, selection and other factors on the degree of potential range filling over shorter time frames.

CONCLUSION

Our study is one of the first to explicitly study the relationship between genetic diversity and macroecological patterns in a species. We found that niche breadth and potential range size are positively correlated in *A. thaliana* at an intraspecific level, implying that genotypes that use the available climate spaces more evenly are also more widespread. This relationship is driven in part by an important plant phenotype, flowering time, that exhibits substantial heritable trait variation within the species. Our findings suggest that local selection on flowering time can either constrain the distribution of certain genotypes or 'pre-adapt' them to colonise other areas. Importantly, we were only able to detect these patterns by focusing on the effects of genetic loci that cause flowering time differences within *A. thaliana*.

Our work represents just a starting point for determining how heritable trait variation causes macroecological differentiation within a species. Future research can follow-up on these results using an allele distribution modelling approach (e.g. Manel *et al.* 2010; Hancock *et al.* 2011) to identify some of the axes of environmental variation that are separating these ecologically differentiated genotypes geographically. Another way to expand this work is by using more complex genotypes, with specific combinations of alleles at multiple rather than single loci. This could more precisely determine how niche breadths and range sizes vary within species. Furthermore, defining genotypes by their differentiation in other important phenotypes, in addition to flowering, would likely identify some other significant sources of intraspecific macroecological variation as well as some other interesting correlations. Since phenotypes are integrated and therefore correlated with one another, and since some phenotypes are more substantial contributors to a species' total phenotypic variation than others (Klingenberg 2008), it should be possible to get a relatively clear picture of instraspecific macroecological variation by studying a limited and thoughtfully selected set of important phenotypes that represent orthogonal axes of phenotypic variation. Ultimately, ecological niche models that are nuanced, incorporating intraspecific phenotypic and macroecological variation, are needed to accurately forecast the spread/contraction of species' ranges under future climates and to make informed decisions about conservation priorities and threats from biological invasions.

ACKNOWLEDGEMENTS

We would like to thank Christopher Tracey and Ashley Dunithan for their help with ArcMap and Dan Warren for helpful discussions and advice about niche breadth. We would also like to thank Naeha Bhambra, Melissa Pentony, Gabriel Lebec, Kelly Clemenza, Jeanmarie Molina, Yoshie Honzawa and Si Li for their help with plant care and data collection. This work was supported by NSF Frontiers in Integrative Biological Research program grant EF-0425759. PXK was supported by a Biotechnology and Biological Sciences Research Council (BBSRC) grant.

AUTHORSHIP

JB, IE and MP designed the study. LC, SG, IE, JB and PK collected genetic data. LC, SG and JB performed the experimental work. JB, MP, IE, AW and JS analysed the results, with AW and JS providing some novel climate envelope modelling methods. JB and MP wrote the first draft of the manuscript, and all authors contributed substantially to the revisions.

REFERENCES

Angert, A.L, Sheth, S.N. & Paul, J. (2011). Incorporating population-level variation in thermal performance into predictions of geographic range shifts. *Integr. Comp. Biol.*, 51, 733–750.

Balasubramanian, S., Sureshkumar, S., Lempe, J. & Weigel, D. (2006). Potent induction of *Arabidopsis thaliana* flowering by elevated growth temperature. *PLoS Genet.*, 2, 980–989.

Bolnick, D.I., Svanback, R., Fordyce, J.A., Yang, J.H., Davis, J.M., Hulsey, C.D. *et al.* (2003). The ecology of individuals: Incidence and implications of individual specialization. *American Naturalist*, 161, 1–28.

Brown, J.H. (1984). On the relationship between abundance and distribution of species. *Am. Nat.*, 124, 255–279.

Callahan, H.S. & Pigliucci, M. (2002). Shade-induced plasticity and its ecological significance in wild populations of *Arabidopsis thaliana*. *Ecology*, 83, 1965–1980.

Clausen, J., Keck, D.D. & Hiesey, H.M. (1940). *Experimental Studies on the Nature of Species. I. Effect of Varied Environments on Western North American Plants*. Carnegie Institute of Washington, Washington, DC.

Dabney, A. & Storey, J.D. (2010). qvalue: Q-value Estimation for False Discovery Rate Control. R package version 1.22.0. Available at: http://CRAN.R-project.org/package=qvalue. Last accessed 1 May 2012.

Dudik, M., Phillips, S.J. & Schapire, R.E. (2010). Maxent, Version 3.3.3Available at: http://www.cs.princeton.edu/~schapire/maxent/. Last accessed 13 December 2010.

Eckert, A.J., Bower, A.D., Wegrzyn, J.L., Pande, B., Jermstad, K.D., Krutovsky, K.V. *et al.* (2009). Asssociation genetics of coastal Douglas Fir (*Pseudotsuga menziesu* var. menziesii, Pinaceae). I. Cold-hardiness related traits. *Genetics*, 182, 1289–1302.

Ehrenreich, I.M., Honzawa, Y., Chou, L., Roe, J.L., Kover, P.X. & Purugganan, M.D. (2009). Candidate gene association mapping of *Arabidopsis* flowering time. *Genetics*, 183, 325–335.

Elith, J. (2002). Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. In: *Quantitative Methods for Conservation Biology* (eds Ferson, S. & Burgman, M.). Springer, New York, pp. 39–58.

Elith, J., Phillips, S.J., Hastie, T., Dudik, M., Chee, Y.E. & Yates, C.J. (2011). A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.*, 17, 43–57.

ESRI Inc. (2008). *ArcMap Version 9.3*. ESRI Inc., Redlands, CA.

Fournier-Level, A., Korte, A., Cooper, M.D., Nordborg, M., Schmitt, J. & Wilczek, A.M. (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science*, 333, 86–89.

François, O., Blum, M.G.B., Jakobsson, M. & Rosenberg, N.A. (2008). Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genet.*, 4, e1000075.

Garbutt, K. & Bazzaz, F.A. (1987). Population niche structure: differential response of *Abutilon theophrasti* progeny to resource gradients. *Oecologia*, 72, 291–296.

Gaston, K.J. (1994). Measuring geographic range sizes. *Ecography*, 17, 198–205.

Hancock, A.M., Brachi, B., Faure, N., Horton, M.W., Jarymowycz, L.B., Sperone, F.G. *et al.* (2011). Adaptation to climate across the *Arabidopsis thaliana* genome. *Science*, 333, 83–86.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.*, 25, 1965–1978.

Hoffmann, M.H. (2002). Biogeography of *Arabidopsis thaliana* (L.) Heynh (Brassicaceae). *J. Biogeogr.*, 29, 125–134.

Hoffmann, M.H. (2005). Evolution of the realized climatic niche in the genus *Arabidopsis* (Brassicaceae). *Evolution*, 59, 1425–1436.

Hutchinson, G.E. (1957). Population studies: animal ecology and demography: concluding remarks. *Cold Spring Harb. Symp. Quant. Biol.*, 22, 415–427.

Inouye, D.W. (2000). The ecological and evolutionary significance of frost in the context of climate change. *Ecol. Lett.*, 3, 457–463.

Klingenberg, C.P. (2008). Morphological integration and developmental modularity. *Annu. Rev. Ecol. Evol. Syst.*, 39, 115–132.

Köckemann, B., Buschmann, H. & Leuschner, C. (2009). The relationships between abundance, range size and niche breadth in Central European tree species. *J. Biogeogr.*, 36, 854–864.

Koorneef, M. & Meinke, D. (2010). The development of *Arabidopsis* as a model plant. *Plant J.*, 61, 909–921.

Korves, T.M., Schmid, K.J., Caicedo, A.L., Mays, C., Stinchcombe, J.R., Purugganan, M.D. & Schmitt, J. (2007). Fitness effects associated with the major flowering time gene *FRIGIDA* in *Arabidopsis thaliana* in the field. *Am. Nat.*, 169, E141–E157.

Kover, P.X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I.M., Purugganan, M.D. *et al.* (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.*, 5, e1000551.

Levins, R. (1968). *Evolution in Changing Environments*. Princeton University Press, Princeton, NJ.

Manel, S., Poncet, B.N., Legendre, P., Gugerli, F. & Holderegger, R. (2010). Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Mol. Ecol.*, 19, 3824–3835.

May, R. (1999). Unanswered questions in ecology. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 354, 1951–1959.

Meier, E.S., Kienast, F., Pearman, P.B., Svenning, J.-C., Thuiller, W., Araújo, M. B. *et al.* (2010). Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography*, 33, 1038–1048.

Montesinos-Navarro, A., Wig, J., Pico, F.X. & Tonsor, S.J. (2011). *Arabidopsis thaliana* populations show clinal variation in a climatic gradient associated with altitude. *New Phytol.*, 189, 282–294.

New, M., Lister, D., Hulme, M. & Makin, I. (2002). A high-resolution data set of surface climate over global land areas. *Clim. Res.*, 21, 1–25.

Paul, J.R., Morton, C., Taylor, C.M. & Tonsor, S.J. (2009). Evolutionary time for dispersal limits the extent but not the occupancy of species' potential ranges in the tropical plant genus *Psychotria* (Rubiaceae). *Am. Nat.*, 173, 188–199.

Phillips, S.J. (2005). A Brief Tutorial on Maxent. AT&T Research. Available at: http://www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc Last accessed 14 February 2011.

Phillips, S.J. & Dudik, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31, 161–175.

R Development Core Team. (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rehfeldt, G.D., Ying, C.C., Spittlehouse, D.L. & Hamilton, D.A. (1999). Genetic responses to climate in *Pinus contorta*: niche breadth, climate change, and reforestation. *Ecol. Monogr.*, 69, 375–407.

Reinhardt, K., Smith, W.K. & Carter, G.A. (2010). Clouds and cloud immersion alter photosynthetic light quality in a temperate mountain cloud forest. *Bot-Botanique*, 88, 462–470.

Sexton, J.P., Strauss, S.Y. & Rice, K.J. (2011). Gene flow increases fitness at the warm edge of a species' range. *Proc. Nat. Acad. Sci. U.S.A.*, 108, 11704–11709.

Svenning, J.-C. & Skov, F. (2004). Limited filling of the potential range in European tree species. *Ecol. Lett.*, 7, 565–573.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C. *et al.* (2004). Extinction risk from climate change. *Nature*, 427, 145–148.

Toomajian, C., Hu, T.T., Aranzana, M.J., Lister, C., Tang, C., Zheng, H. *et al.* (2006). A nonparametric test reveals selection for rapid flowering in the *Arabidopsis* genome. *PLoS Biol.*, 4, e137.

Vellend, M., Verheyen, K., Flinn, K.M., Jacquemyn, H., Kolb, A., Van Calster, H. *et al.* (2007). Homogenization of forest plant communities and weakening of species-environment relationships via agricultural land use. *J. Ecol.*, 95, 565–573.

Warren, D.L., Glor, R.E. & Turelli, M. (2008). Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, 62, 2868–2883.

Welk, E. & Bruelheide, H. (2006). There may be bias in R/P ratios (realized vs. potential range) calculated for European tree species – an illustrated comment on Svenning & Skov (2004). *J. Biogeogr.*, 33, 2013–2021.

Wilczek, A.M., Roe, J.L., Knapp, M.C., Cooper, M.D., Lopez-Gallego, C., Martin, L.J. *et al.* (2009). Effects of genetic perturbation on seasonal life history plasticity. *Science*, 323, 930–934.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. *et al.* (2008). Effects of sample size on the performance of species distribution models. *Divers. Distrib.*, 14, 763–773.

SUPPORTING INFORMATION

Additional Supporting Information may be downloaded via the online version of this article at Wiley Online Library (www.ecology-letters.com).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Editor, Janneke Hille Ris Lambers
Manuscript received 23 August 2011
First decision made 2 October 2011
Second decision made 19 January 2012
Third decision made 16 March 2012
Manuscript accepted 10 April 2012