Understanding the factors that govern the distribution of species is a central goal of evolutionary ecology. It is commonly assumed that geographic range limits reflect ecological niche limits and that species experience increasingly marginal conditions towards the edge of their ranges. Using spatial data and ecological niche models we tested these hypotheses in *Arabidopsis lyrata*. Specifically, we asked whether range limits coincide with predicted niche limits in this system and whether the suitability of sites declines towards the edge of the species’ range in North America. We further explored patterns of environmental change towards the edge of the range and asked whether genome-wide patterns of genetic diversity decline with increasing peripherality and environmental marginality. Our results suggest that latitudinal range limits coincide with niche limits. Populations experienced increasingly marginal environments towards these limits – though patterns of environmental change were more complex than most theoretical models for range limits assume. Genomic diversity declined towards the edge of the species’ range and with increasing distance from the estimated centre of the species’ niche in environmental space, but not with the suitability of sites based on niche model predictions. Thus while latitudinal range limits in this system are broadly associated with niche limits, the link between environmental conditions and genetic diversity (and thus the adaptive potential of populations) is less clear.

Introduction

Understanding species’ geographic ranges is a major goal in ecology. There are two main ways in which the distribution of environmental conditions is thought to shape species’ ranges. The first is based on Hutchinson’s ‘n-dimensional hypervolume’ niche concept (Hutchinson 1957), which describes the region of multivariate environmental space in which conditions permit population persistence (i.e. non-negative growth rates, Hutchinson 1978, Holt 2009). The configuration of these conditions across geographic space in turn defines the potential geographic range of a species (Jackson and Overpeck 2000). Not mutually exclusive to this concept is the model championed by
Brown (1984) in which environmental conditions deteriorate towards the edge of a species’ range. This scenario sets up a gradient within species’ ranges (i.e. within the geographical realization of Hutchinson’s niche: Brown 1984), whereby populations become smaller and increasingly isolated towards the edge of the range (commonly referred to as the ‘central-peripheral’ or ‘abundant-center’ hypothesis: Sagarin and Gaines 2002, Gaston 2003). These premises have greatly influenced ecological and evolutionary perspectives on range limits (e.g. Table 1 from Sagarin and Gaines 2002, Sexton et al. 2009) and empirical evaluation of these models is thus of fundamental interest.

Evaluating both Hutchison’s and Brown’s models is particularly relevant for narrowing down factors influencing adaptation at the edge of the range. Hutchison’s model compels us to ask whether species’ fill their potential niche on the landscape. Where range limits and niche limits are closely aligned, it follows that range expansion would require ‘niche expansion’. In such cases, subsequent investigation into the traits involved in adaptation to abiotic conditions and the factors influencing the evolution of these traits is warranted. Alternatively, species’ may fail to occupy all parts of the landscape that are suitable for them because of biotic interactions or constraints on dispersal (Soberrón 2007, Peterson et al. 2011). At such range limits, attention is likely to be better focused on understanding antagonistic interactions or dispersal limitation and the evolution of the traits mediating these processes, rather than physiological constraints. In both cases, Brown’s model has implications for thinking about the mechanisms limiting adaptation. Notably, small population sizes at the edge of the range resulting from marginal conditions may decrease the efficacy of selection relative to drift (Kimura 1983, Charlesworth 2009), lead to inbreeding (reviewed by Keller and Waller 2002), and/or establish density gradients that promote swamping gene flow from elsewhere in the range (García-Ramos and Kirk-patrick 1997). All of these factors may diminish the ability of peripheral populations to respond to selection and limit adaptation to novel conditions (Soule 1973, Kawecki 2008). Thus both Hutchison’s and Brown’s models have important implications for understanding adaptation at range limits.

Although both Hutchison’s and Brown’s models are supported by a number of studies, not all species conform to these models (see reviews by Sagarin and Gaines 2002, Eckert et al. 2008, Hargreaves et al. 2014, Lira-Noriega and Manthey 2014, Lee-Yaw et al. 2016, Pironon et al. 2016). Species may fail to meet the expectations of one or both models for a number of reasons. For instance, in the northern hemisphere, a limited amount of time for colonization since the last glacial maximum may mean that many species have yet to colonize the full spatial extent of their Hutchison niche (i.e. a form of dispersal limitation). Likewise, range expansion from southern refugia may mean that the spatial centers of many contemporary ranges in northern areas are displaced from historical areas of high abundance and diversity, thus generating patterns that deviate from expectations under Brown’s model (see also Micheletti and Storfer 2015). Finally, both colonization from multiple historical refugia and the availability of optimal conditions in multiple areas currently may mean that the abundance structure of some species’ ranges are better described as multimodal (Brown et al. 1995, Guo et al. 2005, Poulin and Dick 2007, Soley-Guardia et al. 2014). Thus traditional ways of modeling niche dynamics and the structure of species’ ranges may be overly simplistic in many cases.

The difficulty of empirically quantifying environmental suitability and population size (i.e. with the amount of replication and across the scales necessary for understanding range limits) has limited previous investigations of the

Table 1. Environmental variables used to generate niche models for Arabidopsis lyrata sub. lyrata and their effects on model performance.

<table>
<thead>
<tr>
<th>Type</th>
<th>Variable</th>
<th>Abbreviation in text</th>
<th>Source*</th>
<th>Basis of inclusion†</th>
<th>Percent contribution†</th>
<th>Permutation importance†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Climate</td>
<td>Mean diurnal range (temperature)</td>
<td>Bio2</td>
<td>BioClim</td>
<td>RF</td>
<td>8.65</td>
<td>2.87</td>
</tr>
<tr>
<td></td>
<td>Maximum temperature of warmest month</td>
<td>Bio5</td>
<td>BioClim</td>
<td>BI</td>
<td>9.32</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>Precipitation seasonality</td>
<td>Bio15</td>
<td>BioClim</td>
<td>RF</td>
<td>4.04</td>
<td>6.74</td>
</tr>
<tr>
<td></td>
<td>Precipitation of wettest quarter</td>
<td>Bio16</td>
<td>BioClim</td>
<td>BI/RF*</td>
<td>9.48</td>
<td>22.3</td>
</tr>
<tr>
<td></td>
<td>Priestley–Taylor coefficient</td>
<td>Alpha</td>
<td>CGAIR</td>
<td>BI</td>
<td>4.51</td>
<td>18.47</td>
</tr>
<tr>
<td></td>
<td>Average minimum temperature of early spring (March, April)</td>
<td>Tmin_sp</td>
<td>Derived from WorldClim</td>
<td>BI</td>
<td>61.63</td>
<td>46.71</td>
</tr>
<tr>
<td>Topography</td>
<td>Compound topographic index</td>
<td>CTI</td>
<td>USGA (Hydro1k)</td>
<td>BI</td>
<td>0.21</td>
<td>0.43</td>
</tr>
<tr>
<td>Vegetation</td>
<td>Variability in maximum green vegetation fraction</td>
<td>mgvf_sd</td>
<td>Derived from USGS-LCI</td>
<td>BI</td>
<td>2.15</td>
<td>2.03</td>
</tr>
</tbody>
</table>

†Variables were either chosen based on information about the biology of the species (BI) or because important for classifying presence and background sites in a random forest analysis (RF); Supplementary material Appendix 1.
‡Based on Maxent output and averaged across those k-fold models passing modeling evaluation.
§Bio16 was also highly correlated with the derived mean spring and mean summer precipitation variables that we had a priori singled out as being important for the species but that were subsequently dropped.
relationship between suitable habitat and range limits, and thus evaluation of niche-based explanations for range limits. In this study we use ecological niche modeling and a large genomic dataset to overcome some of these limitations and to address these questions in Arabidopsis lyrata (lyrate rock-cress) across its North American range. We specifically asked: 1) are range limits niche limits in this system? 2) Do suitable conditions decline towards the edge of the range? 3) How do important niche variables change at the edge of the range? 4) In line with an effect of environmental conditions on population size and genetic diversity, is there a relationship between environmental suitability and genome-wide patterns of genetic diversity, or is genetic diversity explained by range position alone, particularly in light of range dynamics associated with the Pleistocene glaciations?

Material and methods

Study system

Arabidopsis lyrata ssp. lyrata (hereafter A. l. lyrata for simplicity) occupies disturbed habitats in the eastern USA and southern Canada (Fig. 1a) and is one of the southern-most members of the circumpolar Arabidopsis lyrata species complex (Schmidl et al. 2010). In addition to geographic location, A. l. lyrata differs from its closest relatives in phenotype (Arabidopsis lyrata ssp. petraea: Remington et al. 2013) and mating system (Arabidopsis arenicola: Hohmann et al. 2014, Willi unpubl.), and European and North American populations are at least partially reproductively isolated (Lep-pälä and Savolainen 2011). Within A. l. lyrata, previous work has documented two genetic groups (Willi and Määttänen 2010, Griffin and Willi 2014), likely reflecting repeated contraction into and expansion from distinct eastern and western refugia during the Pleistocene glaciations (Griffin and Willi 2014). Genomic data suggest divergence between these lineages is low and that some amount of gene flow has occurred in the recent past (Fracassetti and Willi unpubl.). Crosses between the two lineages also result in viable offspring with no evidence of reduced fitness (Willi unpubl.). For these reasons, we treat all North American populations of A. l. lyrata as a single taxonomic unit of interest and ask about range limits and the niche of this particular subspecies. At the same time, evidence for post-glacial range expansion from at least two refugia prompts us to consider this recent history in our analysis of the distribution of genome-wide diversity below.

Quantifying environmental suitability

We used ecological niche models to assess the distribution of suitable habitat for A. l. lyrata within and beyond its range in North America. Using published studies, herbaria and government databases, we collected close to 600 localities from all States and Provinces where the species occurs and georeferenced these records to within 1 km (average estimated error: 0.25 km). Records along lakeshores or on small islands that resulted in missing data at the resolution of our raster dataset (i.e. coordinates fell into a water-dominated grid cell in the rasters described below) were moved a maximum distance of 5 km to their nearest ‘onshore’ grid cell or were discarded. In order to reduce the potential effects of sampling bias on our niche models (Boria et al. 2014, Varela et al. 2014), we thinned records using an approach that considered both the density of points (geographic filtering) and environmental novelty (environmental filtering; full details in Supplementary material Appendix 1). After thinning, our final locality dataset included 279 records (including 55 that had been moved a median distance of 1.28 km ‘onshore’; Fig. 1).

The environmental data used to calibrate our models included eight climatic, soil, hydrological and vegetation variables at a resolution of 30 arc-seconds (Table 1; selection process from a larger set of variables is explained in Supplementary material Appendix 1). Niche models were built using Maxent (Phillips et al. 2006, Phillips and Dudík 2008) and the ‘raster’ (Hijmans et al. 2016a) and ‘dismo’ packages (Hijmans et al. 2016b) in R (R Core Team). The background area for model calibration included all ecotones (Commission for Environmental Cooperation 1997) occupied by the species (i.e. presumably accessible to the species based on the general distribution of the communities within which it has assembled: Soberón 2010, Barve et al. 2011) with the maximum study extent set as 100°-70°W and 32°-53°N (Supplementary material Appendix 1 Fig. A1-1). Prior to model calibration, we used the approach described by Warren and Seifert (2011, see also Warren et al. 2014) to tune both the features and regularization multiplier value (RM; Supplementary material Appendix 1). All other parameters in Maxent were set at default values. Following feature and RM tuning, we used ten-fold cross-validation (Fielding and Bell 1997) to calibrate a final set of models, assessing the ability of each model to discriminate between withheld presence data and background data based on the area under the curve statistic (AUC; of the receiver operating characteristic plot). For each model, we also compared observed AUC values to a null distribution of values generated from 99 niche models built using random points (Raes and ter Steege 2007). Models that had an AUC score that was both ≥0.70 and that fell outside the 95th quantile of this null distribution (i.e. performed significantly better than random) were used to predict environmental suitability across the study region. The mean of these predictions was taken as an estimate of the suitability of each cell for the species (Peterson et al. 2011), which in turn served as the basis for subsequent analyses. We note that AUC values from k-fold cross-validation are often inflated by spatial autocorrelation in the locality dataset (Araújo and Guisan 2006, Veloz 2009, Hijmans 2012, Bahn and McGill 2013). To test the effects of this known issue on model evaluation, we also generated models using the masked geographically structured cross-validation approach of Radosavljevic and Anderson (2014) using the ‘ENMeval’ package (Muscarella et al. 2014) in R (R Core Team).
As a second metric of environmental suitability, we used the method of Lira-Noriega and Manthey (2014) to calculate the distance of sites to an estimate of the centre of the species’ niche in environmental space. Specifically, we converted continuous suitability scores (based on average niche model predictions) to a map of suitable habitat (i.e. representing the species’ niche in geographic space) using two thresholds for defining suitable areas: the lowest suitability score for any A. l. lyrata locality (‘lowest presence threshold’: Pearson et al. 2007; Fig. 1c) and the 5th percentile of locality suitability scores (‘5th percentile threshold’: Fig. 1d). We then took a random sample of 5000 points from within the area of suitable habitat from each of these maps and used these points, as well as the locality data, to conduct a principal
component analysis (PCA) of the eight environmental variables considered in the niche models. The species’ niche centroid in each case was defined as the vector of median scores of all points used in the PCA for PC axes one to six (which captured >99% of the variance in environment). The distance between any location and this niche centroid in environmental space served as a measure of the niche centrality (or conversely marginality) of that point. As a sensitivity test, we reran the niche centroid calculation for both suitability thresholds, and with different numbers of random sites in the PCA (500, 1000, 2500).

Are range limits niche limits?

Apart from qualitative comparisons between model prediction surfaces and the species’ range, we used a novel range-fitting test to evaluate the statistical significance of the fit between observed range limits and predicted range limits based on the species’ niche. The steps of this test were as follows: 1) we represented the species’ observed geographic range as the minimum convex polygon (MCP) around all known localities; 2) continuous niche model predictions were converted to a map of suitable habitat representing the species’ niche limits; 3) depending on the range limit in question (e.g. northern limit, etc.) this raster of discrete suitability was then subset to include only cells to the north (or south etc.) of the centroid of the range MCP; 4) treating all of the cells inside the MCP in this subsetted raster as ‘in-range’ and all cells outside the MCP as ‘out-of-range’, we used a modified version of the true skill statistic (TSS; Allouche et al. 2006), to assess the fit of suitable habitat to the range MCP. Specifically, TSS\textsubscript{range} = proportion of in-range cells predicted as suitable + proportion of out-of-range cells predicted as non-suitable – 1; 5) the range MCP was then shifted randomly in the direction of the range limit being evaluated (e.g. to the north etc.); 6) based on the centroid of the shifted MCP, the raster of suitable habitat was subset anew and TSS\textsubscript{range} was recalculated; 7) steps 5 and 6 were repeated 99 times to generate a null distribution of TSS\textsubscript{range} scores for each range limit. The fit between a given range limit and the niche limit of the species in the same direction was considered significant if the observed value of TSS\textsubscript{range} was more extreme (indicating a better fit) than 95% of the values in the null distribution. We conducted this test to evaluate the fit of the species’ northern, southern and western range limits to predicted niche limits (the species’ eastern limit coincides with the Atlantic ocean precluding the test this direction). The MCP was repeatedly shifted in each direction by a random amount between 50 and 500 km (i.e. reasonable for our study region). We ran this test twice, using maps of suitable habitat based on the lowest presence and 5th percentile thresholds.

Environmental changes towards the edge of the range

We used the average predicted suitability of sites based on our niche models, as well as our index of niche centrality, to test the hypothesis that environmental conditions become more marginal (i.e. less suitable) towards the edge of the species’ range. Brown’s hypothesis specifically posits that conditions are most suitable at the geographic centre of species’ ranges. As an index of peripherality, we took the minimum great circle distance from each locality to the centroid of the MCP encompassing all localities. These values were standardized by dividing by the sum of this distance and the minimum distance between the locality and the MCP hull (following Griffin and Willi 2014). Localities were also assigned to non-overlapping groups according to their primary direction (north, south, east, west) away from the MCP centroid. We then used linear models to test whether environmental suitability declines and distance to the niche centroid increases with increasing peripherality. Direction was included as a second, potentially interacting, explanatory variable. Peripherality scores were arcsin transformed and distance to the niche centroid was log transformed to better meet the assumptions of a linear model. We repeated the test for the estimates of distance to niche centroid based on the two thresholds used above for defining suitable habitat (which influenced the area of sampling for the PCA) and with the PCAs involving different numbers of points to estimate the niche centroid (above). One of the sites towards the eastern edge of the species’ range (a coastal population from Sandy Hook, NJ, USA) was an environmental outlier in both datasets. We ran the models with and without this point included.

To understand environmental patterns underlying changes in suitability at the species’ range limits on a smaller scale, we used a model selection approach to evaluate the nature of the relationship between range position and the two variables that were most important for predicting presence in the niche models (see Results). Focusing on 200 km transects spanning the range edge in 12 different places and centered on different peripheral populations, we specifically tested whether minimum temperature of early spring and precipitation of the wettest quarter change across the range edge and whether changes in these variables tend to be gradual or abrupt (Fig. 1a; see Supplementary material Appendix 1 for full details of transect designation). We extracted environmental values at 5 km intervals along each transect and used model selection based on AIC\textsubscript{c} to determine whether an intercept-only (no change), linear (environmental gradient) or four-parameter logistic (abrupt change) model better described the relationship between each environmental variable and geographic position within each transect. Finally, because these simple model types may fail to account for more complex changes in the environment at range limits, we estimated the number of significant structural changes in the relationship between each environmental variable and transect position using the ‘strucchange’ (Zeileis et al. 2002, Achim et al. 2003) package in R (R Core Team). The minimum number of observations per segment was set to five (permitting a maximum of seven ‘breakpoints’ in the relationship between position and environment per transect). No structural changes are expected where an intercept or linear model best describe environmental change and two changes are expected where a four-parameter logistic model applies.
Environmental marginality and patterns of genomic diversity

Small population size associated with increasingly marginal habitat is expected to result in a decrease in genetic diversity towards the edge of species' ranges (Eckert et al. 2008, Lira-Noriega and Manthey 2014). To test this hypothesis, we looked at the relationship between environmental suitability and genome-wide estimates of genetic diversity. Following the protocol and pipeline outlined by Fracassetti et al. (2015), we sequenced pools of 25 individuals from 42 outcrossing populations (also included in the niche models) from across the species' range (Fig. 1a). After aligning reads to the published nuclear genome of *A. l. lyrata* (ver. 1.0; Hu et al. 2011) and filtering (Fracassetti et al. 2015), single nucleotide polymorphisms (SNPs) were called for each population using Varscan (Koboldt et al. 2012). We retained biallelic SNPs with a minimum mapping quality score of 20, a range of coverage of 50–500×, a minimum allele count of 3, \( p < 0.15 \), and a minimum variant allele frequency of 0.03. Our final genomic dataset consisted of a mean of > 1.6 million SNPs per population with an average depth of coverage of 125×. Nucleotide diversity (\( \pi \)) was estimated in 5000 bp windows across the genome of each population using NPStat ver. 0.99c (Ferretti et al. 2013). The weighted median value across windows was taken as an estimate of genomic diversity for each population.

We used Pearson correlation tests to examine the relationship between genomic diversity and environmental marginality. However, genetic diversity is known to decline with increasing peripherality as the result of founder effects during colonization (Griffin and Willi 2014). We thus used linear models to test for an effect of environmental marginality on genetic diversity, above and beyond any effects associated with range position. In this case, range position was defined in terms of distance away from putative refugial areas based on the results of Griffin and Willi (2014). Specifically, Griffin and Willi (2014) described a pattern of declining microsatellite diversity away from the centroids of the western and eastern lineages that they described. These centroids roughly coincide with the Driftless Area of Wisconsin and the central Appalachian Mountains respectively (Fig. 1a), both of which have previously been proposed as refugial areas for a number of other taxa (Jackson et al. 2000, Soltis et al. 2006, Lee-Yaw et al. 2008, Beatty and Provan 2011, Li et al. 2013). We assigned sites to genetic group (based on a modified version of the boundaries from Griffin and Willi 2014: Supplementary material Appendix 1) and used the formula above to calculate the peripherality of each site with respect to the centroid of its genetic group. We regressed nucleotide diversity against peripherality and then asked whether environmental marginality explained any of the residual variation from this relationship. Models were run separately for the two measures of environmental marginality (suitability and distance from niche centre). We note that the suitability outlier was removed for this analysis, as it was an extreme outlier that polarized the values of all other sites.

Data deposition


Results

Niche models

The niche models generated by standard cross-validation in Maxent resulted in AUC values ranging from 0.84–0.91 (mean: 0.88). All ten models had higher AUC values than models built using random locations from within the species' range (\( p < 0.05 \); on average, AUC scores were 0.16 higher than the 95th percentile cutoff of the null distribution). A masked geographically structured approach to model calibration and evaluation resulted in a lower average but wider range of AUC scores (mean: 0.68; range: 0.49–0.84). Relying only on the models with high support in this exercise (AUC > 0.70) produced suitability scores that were highly correlated with the average suitability from the k-fold cross-validation procedure (\( r = 0.89, p << 0.01 \)) and a suitability map that was qualitatively similar to the average of the k-fold cross-validation models (not shown). In light of this similarity, and given that the k-fold cross-validation models were trained with localities spanning the full geographic range of the species (rather than limiting the training data to a portion of the region occupied by the species), the remainder of the analyses were based on the k-fold cross-validation models.

Of the eight variables included in our models, average minimum temperature during early spring (March and April) made the largest percent contribution across all models and resulted in the largest drop in model performance when values were randomly permuted across the training dataset (Table 1). The suitability of sites tended to be highest at intermediate to higher values of this variable and decline towards very high or very low values within the study region (Supplementary material Appendix 2 Fig. A2-1a). Precipitation of the wettest quarter and the Priestly–Taylor coefficient (representing moisture availability) were of moderate importance to most models (based on permutation importance scores but not necessarily percent contribution; Table 1). Suitability tended to decrease with very high values of spring precipitation (Supplementary material Appendix 2 Fig. A2-1b) and towards very high or low values of the Priestly–Taylor coefficient. Non-climate variables, associated with topography and vegetation, were the least important for predicting the occurrence of the species across the landscape (Table 1).
Are range limits niche limits?

Examination of average model predictions across the study region revealed that the species’ range limits were closely aligned with the predicted distribution of suitable conditions for the species in many places – although areas of moderate to high suitability extended beyond the species’ range to the west and to the northeast along the St Lawrence River (Fig. 1). Our range fitting tests generally supported these conclusions (although they do not speak to non-cardinal directions i.e. the northeast). Observed TSS$_{range}$ was significantly higher than values calculated after randomly shifting range limits 50 to 500 km to the north, regardless of whether the lowest presence or 5th percentile threshold was used to define suitable habitat ($p = 0.01$ in both cases). In the southward direction, TSS$_{range}$ was significantly higher than values calculated for shifted range limits when the 5th percentile threshold was used for defining suitable habitat ($p = 0.01$) but not when the lowest presence threshold was used ($p = 0.33$). In the westward direction, observed values were non-significant regardless of the threshold used, supporting the qualitative observation that range limits are not well-predicted by niche limits in this direction.

Environmental changes towards the edge of the range

Patterns of predicted suitability suggested that conditions generally become more marginal towards the edge of the species’ range. Average predicted suitability based on the niche models declined with distance from the centre of the MCP (linear model: estimate $\pm$ SE $= -0.309 \pm 0.032$, $F = 93.16$, df $= 1$, 274, $p < 0.001$; Fig. 2). Although eastern populations tended to have higher suitability scores (Fig. 2a), the main effect of the direction of sites relative to the geographic centre (i.e. north, south, east or west) was not significant ($F = 0.33$, df $= 3$, 274, $p = 0.80$; type III sums of squares). However, the interaction between peripherality and direction was marginally significant ($F = 2.50$, df $= 3$, 274, $p = 0.059$), with suitability declining slightly more steeply towards the northern edge of the range; Fig. 2a). These results were robust to the inclusion or exclusion of a single high suitability outlier near the eastern edge of the species’ range (outlier shown in Fig. 2a).

The relationship between niche centrality and geographic peripherality depended on direction away from the range centre (Fig. 2b). As expected if conditions become more marginal towards the edge of the range, distance to the niche centroid increased moving away from the geographic centre of the species’ range to the north and south. However, distance to the niche centroid increased only gradually to the west and tended to decline (i.e. conditions became less marginal) towards the eastern edge of the species’ range (Fig. 2b). These results were consistent regardless of the threshold of suitability (Supplementary material Appendix 2 Table A2-1) and number of points in the PCA used to calculate the niche centroid (the significance of the main effect of peripherality varied across the latter set of iterations; however we note that it is difficult to interpret main effects in the presence of a significant interaction, and in all cases, results were qualitatively the same). Results were once again robust to the inclusion or exclusion of the outlier.

The variables that best predicted presence in the niche models exhibited a diversity of patterns at the edges of the species’ range (Fig. 3). Average minimum spring temperature...
tended to decline across northern range limits and increase across southern range limits. Model selection suggested that these changes were more abrupt than simple linear models would predict (i.e. the four-parameter logistic model was chosen in seven out of twelve transects; the linear model was chosen for the remaining five transects). Examination of the data revealed that change was most abrupt or step-like for transect 6 in the northeastern part of the species’ range. The direction of change in precipitation of the wettest quarter was more variable across transects, although a consistent pattern of increasing precipitation towards the most southern range limits was observed (transects 8, 10, 11, 12; Fig. 3b). Our analysis of structural changes in the relationship between position and environment along each transect suggested that the three types of relationships tested (intercept-only, linear, four-parameter logistic) may have been overly simplistic. For both variables, the estimated number of breakpoints was usually higher than the number expected for the selected model (0 for intercept and linear models, 2 for four-parameter logistic model; Supplementary material Appendix 2 Table A2-2).

Visual examination of the data for each transect also suggested that patterns of environmental change were complex in many cases (Fig. 3a, b).

Environmental marginality and patterns of genomic diversity

In line with the results of Griffin and Willi (2014), genomic diversity was negatively correlated with distance from the centroids of the two genetic groups that comprise the species ($r = -2.51$, $df = 40$, $p = 0.016$; Fig. 4a). Genomic diversity was not significantly associated with the predicted suitability of sites based on our niche models ($r = 1.47$, $df = 40$, $p = 0.15$; Fig. 4b). However, genomic diversity did decline with distance to the centre of the species’ niche in environmental space ($r = -3.72$, $df = 39$, $p = 0.00063$; Fig. 4c). This latter measure of marginality also explained a significant proportion of the residual variation in genomic diversity after accounting for range position (linear model: estimate ± SE = −0.00030 ± 0.00011, $F = 6.89$, $df = 1, 39$, $p = 0.012$).

Figure 3. Changes in (a) average temperature during the early spring and (b) precipitation of the wettest quarter towards and across range limits in *Arabidopsis lyrata* based on transects centered on different peripheral populations. Plots in each panel correspond to the transects depicted in Fig. 1 and are ordered from west to east for northern (top row) and southern (bottom row) sections of the species’ range. Transects spanned the range edge (100 km in either direction) with the focal peripheral population located at position 0. Negative values along the x-axis represent locations within the species range, positive values represent locations over the edge of the range. For clarity, the y-axis varies across plots and colour is used instead to represent the environmental value at a given location.
Discussion

We evaluated niche-based explanations for range limits in *Arabidopsis lyrata* ssp. *lyrata*. Northern and southern range limits in this system were well predicted by our niche models, suggesting that the species is niche limited in these directions. In contrast, suitable habitat to the west and northeast of the species’ range indicates that other factors shape range limits in those directions. Environmental suitability declined away from the geographic centre of the species’ range, with the underlying environment changing in a variety of ways at the edge of the species’ range. However, the relationship between the environmental suitability of sites and genome-wide patterns of genetic diversity depended on the measure of suitability used. Thus although our results generally support niche-based explanations for northern and southern range limits, variation in conditions within the range likely nuance the actual distribution and abundance of populations (Brown et al. 1995, Guo et al. 2005, Poulin and Dick 2007, Soley-Guardia et al. 2014).

In addition, not all range limits appear to be niche limits in this system. Suitable habitat extended continuously to the west and to the northeast of the species’ range. Although the predicted suitability of sites did decline towards range limits in these directions, the distance of sites to the centre of the species’ niche in environmental space (i.e. the marginality of sites) showed only a moderate increase towards the species’ western limit, and actually decreased towards the eastern limit. Thus it is less clear that populations at the eastern and western periphery of the range experience marginal conditions. Whereas the Atlantic Ocean (i.e. a hard physical barrier) coincides with the eastern range limit, factors limiting the species to the west remain unknown. The availability of moderately suitable habitat beyond this limit may mean that the species is dispersal limited in this direction. However, it is also possible that other abiotic or biotic factors not incorporated in our models set the western bounds of the species’ range. All together, these findings underscore the potential for peripheral populations to be influenced by dramatically different processes in different parts of the range.

Niche limits, environmental marginality and range limits in *Arabidopsis lyrata*

Consistent with the idea that range limits are manifestations of species’ ecological niches (i.e. sensu Hutchinson 1957, e.g. Jackson and Overpeck 2000, Soberón 2007), we found that the northern and southern range limits of *A. l. lyrata* were generally predicted by the niche models. Furthermore, in line with the central-peripheral hypothesis, there was an overall decline in the predicted suitability of sites towards these range limits. At the same time, qualitative examination of the distribution of suitable habitat (e.g. Fig. 1b) revealed patches of suitable habitat near both range limits. Likewise, not all central areas were highly suitable. Thus, although our results generally support niche-based explanations for northern and southern range limits, variation in conditions within the range likely nuance the actual distribution and abundance of populations (Brown et al. 1995, Guo et al. 2005, Poulin and Dick 2007, Soley-Guardia et al. 2014).

Implications for adaptation at the edge of the range

Where niche models predict range limits, the variables with the greatest influence on model predictions may represent dimensions along which adaptation at the range edge is constrained. The variables that were most important for predicting the presence of *A. l. lyrata* were average minimum temperature during the early spring and precipitation during the wettest quarter. Other variables were less important to the models; although general moisture availability (e.g. the Priestley–Taylor coefficient) was of moderate influence. That extreme temperatures and levels of precipitation during the spring exert a strong influence of the species’ distribution is not surprising given that *A. l. lyrata* breaks seed dormancy during the spring and that growth during this time...
is compromised by frost and drought stress (Paccard et al. 2014, Wos and Willi 2015).

Apart from identifying variables associated with range limits in *A. l. lyrata*, our analysis of environmental change at the edge of the range also has implications for thinking about adaptation. Theoretical models for adaptation generally assume that range limits fall along linear environmental gradients (Kirkpatrick and Barton 1997, Bridle et al. 2010), or occur in coupled discrete patches (e.g. source-sink scenarios: Holt and Barfield 2011). These assumptions in turn have implications for the number and effect size of mutations necessary for adaptation, as well as for the effects of gene flow on adaptation (Kawecki 2008, Gomulkiewicz et al. 2010, Holt and Barfield 2011). Range limits in *A. l. lyrata* were associated with roughly linear changes in the environment in 46% of the tests we conducted. Thus theoretical results based on linear gradients may be relevant to this system in many places. At the same time, environmental change across range limits was often best characterized by a four-parameter logistic model or was more complex than any of the models considered presently. The frequency of such patterns indicates that fully understanding range limits in this system (and possibly others) will require greater theoretical understanding of the impacts of non-linear environmental changes on adaptation at the edge of the range (García-Ramos and Huang 2012, Schieffers et al. 2014).

Finally, the patterns of genomic diversity revealed in our analyses speak to the adaptive potential of peripheral populations. A long-held explanation for the failure of adaptation at range limits is that peripheral populations have limited genetic variation upon which selection can act (Hoffmann and Blows 1994, Blows and Hoffmann 2005). Griffin and Willi (2014) previously documented two genetic groups within *A. l. lyrata* and a pattern of declining genetic diversity away from the geographic centers of these groups based on microsatellite data. Our genomic data confirm a decline in genetic diversity at the range edge. Thus low levels of diversity at the edge of the species’ range may compromise the ability of populations to respond to the environmental changes discussed above.

### The impact of environmental marginality on genomic diversity

In addition to range position, we tested the importance of marginal conditions for explaining levels of genomic diversity. Following from the central-peripheral hypothesis, genetic diversity is expected to decline towards the edge of the range as populations become smaller and more isolated in response to increasingly marginal conditions (Brown 1984, Sagarin and Gaines 2002, reviewed by Gaston 2003). We found mixed support for this hypothesis. Genomic diversity declined significantly with increasing distance from the estimated center of the species’ niche in environmental space. Furthermore, this measure of suitability explained a significant amount of the residual variation between genomic diversity and range position. However, niche model estimates of suitability were unrelated to genomic diversity. These results caution against reliance on a single metric of suitability when conducting this type of analysis and suggest that more work is needed to understand the relationship between environmental conditions and genetic diversity in this system.

At the same time, a growing number of studies have questioned the relationship between genetic variation, range limits, and the environmental suitability of sites. Eckert et al. (2008) and, more recently, Pironon et al. (2016), reviewed the evidence for a genetic signature of the central-peripheral hypothesis. Both studies found mixed to low support for this hypothesis. Lira-Noriega and Manthey (2014) analyzed genetic variation in relation to both geography and niche centrality. They found that genetic diversity was consistently negatively associated with niche centrality but not with geography, concluding that the central-peripheral hypothesis may only apply when environmental marginality and geographic peripherality are positively correlated (Lira-Noriega and Manthey 2014). In their review, Pironon et al. (2016) pointed out that such correlations are far from universal and suggested that colonization dynamics may play a more important role than ecological suitability in shaping patterns of genetic diversity. Similar conclusions have been reached by a number of studies for a variety of taxa (Soares et al. 2014, Duncan et al. 2015, Ortego et al. 2015, Pironon et al. 2015, Trumbo et al. 2016). In *A. l. lyrata* clear support for an association between genomic diversity and range position, with only limited evidence of an association between genomic diversity and environmental suitability, suggests that colonization dynamics may have been the main factor shaping genetic diversity in this system as well. Phylogeographic analyses confirming the location of historical refugia and colonization routes in this system are needed to fully test this hypothesis. At the same time, we note that the most likely axes of historical colonization (e.g. away from the genetic group centers identified here) covary with changes in environmental conditions and it can be hard to disentangle the relative impacts of the environment versus founder effects on genetic diversity in such cases.

### Limitations and future directions

Although our niche models highlight specific variables that may be influencing range limits in *A. l. lyrata*, we note that unmodeled variables that systematically exclude the species from regions of environmental space that are otherwise suitable can influence model predictions (Peterson et al. 2011). Direct experimentation is thus ultimately necessary to establish a causal link between the variables examined here and range limits. Experimental validation of results is also pertinent given known issues with model evaluation metrics (Araújo and Guisan 2006, Elith and Leathwick 2009, Veloz 2009, Hijmans 2012, Bahn and McGill 2013). The implementation of geographically structured evaluation with our dataset resulted in a wider range of AUC scores (some of which were quite low), suggesting that caution is needed when interpreting the statistical support for models based on standard cross-validation. Nevertheless, carefully calibrated
models serve as a useful starting point for investigation, pointing to variables that may influence the range and where their effects on range limits are likely most pronounced.

Our results are also interesting in light of previous observations that selfing has evolved repeatedly at the edge of the range of *A. l. lyrata* (Griffin and Willi 2014). Specifically, our results highlight the potential for such shifts in mating system to correspond not only with geography, but with increasing marginal habitat. That a selfing congener (*A. arenicola*) replaces *A. l. lyrata* to the north — the range limit qualitatively associated with the sharpest niche limit (e.g. Fig. 1) — further highlights a link between selfing and environmental conditions. These observations suggest that our understanding of mating system evolution would benefit from consideration of the potential influence of the environment on the processes thought to promote the evolution of selfing (e.g. mate limitation: Busch and Schoen 2008, Busch and Delph 2012).

Finally, our study contributes one of the first tests of genome-wide patterns of genetic diversity in relation to range position and the environmental suitability of sites. The patterns we observed were consistent with results from previous work based on microsatellites (Griffin and Willi 2014). This observation is reassuring given that most studies investigating patterns of genetic diversity to date have relied on a limited number of markers (reviewed by Eckert et al. 2008, Lira-Noriega and Manthey 2014). At the same time, genomic data paves the way for a better understanding of the adaptive potential of peripheral populations. For instance, identifying the specific genes and traits involved in local adaption, and examining variation in these genes at different types of range limits (i.e. ‘leading’ versus ‘rear’ edges: Hampe and Petit 2005) are important avenues of future investigation. Likewise, genomic data can be used to evaluate patterns of gene flow and assess the extent to which diversity in peripheral populations has been influenced by demographic events such as founder effects and population bottlenecks. The coupling of such data with the types of spatial analyses presented here represents a powerful approach for advancing our understanding of species’ geographic range limits.

**Funding** — We received funding from Natural Sciences and Engineering Research Council of Canada (PDF-487035-2016) and Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (PP00P3-123396 and PP00P3-146342).

**References**


