Climate biogeography of *Arabidopsis thaliana:* linking distribution models, individual performance, and life history

Christina Yim[1], Emily S. Bellis[1,2], Victoria L. DeLeo[1], Diana Gamba[1], Robert Muscarella[3], Jesse R. Lasky[1,*]

[1]Department of Biology
Pennsylvania State University
University Park
USA

[2]Department of Computer Science
Arkansas State University
Jonesboro
USA

[3]Plant Ecology and Evolution
Evolutionary Biology Centre
Uppsala University
Uppsala
Sweden

*Corresponding Author
lasky@psu.edu

Biosketch
The research team is interested in understanding how individual-level processes influence evolutionary and ecological dynamics at continental scales.

Author contributions
All authors contributed to study design. CY, ESB, VLD, DG, and JRL conducted analyses. CY and JRL led the writing with contributions from all authors.

Abstract

AIM

The role of environmental conditions in limiting species distributions is often hypothesized, but it is challenging to gather large-scale data to demonstrate environmental impacts on individual performance. The past and present biogeography of model organisms is key context to understanding how environment shapes species' genetic and phenotypic diversity.

LOCATION

Global

TAXON

Arabidopsis thaliana ("Arabidopsis")

METHODS

We fit occurrence records to climate data, and then projected the distribution of Arabidopsis under last glacial maximum, current, and future climates. We confronted model predictions with individual performance measured on 2,194 herbarium specimens, and we asked whether predicted suitability was associated with life-history variation measured on 898 natural accessions.

RESULTS

The most important climate variables constraining the Arabidopsis distribution were winter cold in northern and high elevation regions and summer heat in southern regions. Herbarium specimens from regions with lower habitat suitability in both northern and southern regions were smaller, supporting the hypothesis that the distribution of Arabidopsis is constrained by climate-associated factors. Climate anomalies partly explained interannual variation in herbarium specimen size, but these did not closely correspond to local limiting factors identified in the distribution model. Late-flowering genotypes were absent from the lowest suitability regions, suggesting slower life histories are only viable closer to the center of the realized niche. We identified glacial refugia farther north than previously recognized, as well as refugia concordant with previous population genetic findings. Lower latitude populations, known to be genetically distinct, are most threatened by future climate change. The recently colonized range of Arabidopsis was well-predicted by our native-range model applied to certain regions but not others, suggesting it has colonized novel climates.

MAIN CONCLUSIONS

Integration of distribution models with performance data from vast natural history collections is a route forward for testing hypotheses about species distributions and their relationship with evolutionary fitness across large scales.

Introduction

Understanding how environmental conditions limit the performance of individuals and species distributions has important applications. First, the environment-distribution relationship can help project species distributions under past climates to understand their ecological and evolutionary history (Forester et al., 2013). Second, the environment-distribution relationship can identify regions with currently suitable habitat but that are unoccupied for a variety of reasons (e.g. dispersal limitation) (Elith et al., 2010). Third, the environment-distribution relationship can predict how distributions will shift under future environments (Thomas et al., 2004). These approaches are valuable to study model organisms like *Arabidopsis thaliana,* where researchers can link detailed information on genetics and organismal biology with population and community processes (Rudman et al., 2019; Takou et al., 2019) to contribute ecological context and advance integrative biology. We focus here on *Arabidopsis thaliana,* a small annual plant (hereafter referred to as "Arabidopsis") (Koornneef & Meinke, 2010). Arabidopsis has been key to understanding how molecular biology and physiology are linked to ecology, but modern distribution modeling tools have not been applied to understand Arabidopsis distributions (Hoffmann, 2002).

Environment-distribution relationships fundamentally arise from processes acting on individuals (Clark, 2010). In general, transplant experiments show that individual performance tends to decline outside species' natural geographic range (Hargreaves et al., 2014) and efforts to integrate information at the individual level into distribution models are emerging (Buckley et al., 2011; Elith et al., 2010; Lasky et al., 2020; Merow et al., 2014; Samis & Eckert, 2007). For example, in the perennial herb *Lathyrus,* Greiser et al. (2020) compared demographic models of performance in dozens of common gardens to distribution model predictions across a landscape. Where populations inhabit harsh environments (e.g. at range margins), local adaptations can emerge, such as life history changes to tolerate or escape harsh periods (Bontrager et al., 2021). In Arabidopsis, there is evidence that local adaptation to environment involves life history (e.g. flowering time) changes (Lovell et al., 2013; Martínez-Berdeja et al., 2020), though the environmental mechanisms of selection are less clear. However, it is challenging to determine how environment shapes individual performance and life history variation across the ranges of broadly distributed species, as well as the consequences for distributions (Lasky et al., 2020).

A vast and underexploited resource of individual-level information can be found in natural history collections (Heberling, 2021; Lopez et al., 2019). For example, Bontrager & Angert (2015) showed with herbarium specimens of *Clarkia* that fecundity decreased with drier summers, and that moving toward the western edge of the species range brought declining summer precipitation and individual fecundity, suggesting a mechanism limiting its distribution. In Arabidopsis, DeLeo et al. (2020) found region-specific decadal shifts in phenology and physiology in herbarium specimens. Public seed banks house great diversity from across species ranges that can also be used to study the geography of natural variation (Estarague et al., 2021; Scholl et al., 2000). Here we use individual-level performance estimates from museum specimens and life history variation in natural seedbank accessions to interrogate predictions from fitted distribution models.

Arabidopsis is native across Eurasia and Africa and with human assistance has more recently colonized North America, South America, Japan, Australia, and New Zealand. Other species in the genus *Arabidopsis* are restricted to narrower ranges of cool temperate climates, with Arabidopsis having expanded to a broader range, *e.g.* including Mediterranean habitats (Hoffmann, 2005). Arabidopsis can have a rapid life cycle (~6 weeks) though many individuals are longer-lived winter annuals (Wilczek et al., 2009). Studies of large-scale environmental response in Arabidopsis have largely focused on local adaptation and environmental patterns explaining genetic differences in the species (e.g. Hancock et al., 2011; Lasky et al., 2018; Martínez-Berdeja et al., 2020; Toledo et al., 2020), but less is known about the determinants of the species' distribution. The last major overview of the climate biogeography of Arabidopsis was Hoffmann (2002), who considered Arabidopsis only native to western and central Eurasia, and non-native in China and sub-Saharan Africa. However, recent studies show the Chinese and sub-Saharan African populations are likely native, being genetically diverse, with unique genetic variants, and genetically distinct from Eurasian populations (Durvasula et al., 2017; Zou et al., 2017). Advances in distribution modeling (Elith et al., 2010; Muscarella et al., 2014) and updates to climate and occurrence datasets, including previously poorly known populations, highlight the need to update to our understanding of Arabidopsis biogeography.

Previous studies suggest that multiple climatic factors limit performance and distribution of Arabidopsis. Experiments have shown that winter cold temperatures are a major factor limiting performance (Ågren & Schemske, 2012; Korves et al., 2007). Additionally, Arabidopsis lacks physiological traits for dealing with severe water deficit so it seems likely individual performance in nature is limited by even moderate drought (Clauw et al., 2015). Based on occurrence data, Hoffmann (2002) concluded that spring and fall cold limited Arabidopsis in northern Europe, as well as winter temperatures in Russia, heat and drought toward the southern range margins, and a lack of snow in central Asia. However, these conclusions were based on visual inspection of congruence between climate isotherms and range margins, not statistical inference. Distribution models fit to occurrence data allow an explicit quantitative statement of environment-distribution relationships and the potential to infer the most important environmental factors limiting distributions (Elith et al., 2010; Li et al., 2015).

Distributions are dynamic through time due to environmental change, dispersal, and demographic stochasticity. Studies have used genetic data to infer Arabidopsis refugia during the last glacial maximum (LGM), where subpopulations (sometimes referred to as "relicts") persisted locally before subsequent admixture with an expanding, now widely distributed "non-relict" lineage (Lee et al., 2017). Whether these refugia corresponded to suitable climates is less clear. Future climate change projected impacts on Arabidopsis have largely focused on *relative* climate impacts on different genotypes or alleles (Exposito-Alonso et al., 2018; Fournier-Level et al., 2016), rather than abundance and distribution dynamics. Additionally, many species (including Arabidopsis), have colonized new regions due to human introduction, sometimes exhibiting traits that appear distinct from native range populations, potentially due to plastic or genetic responses to new environments (Turner et al., 2015). Arabidopsis has colonized many regions, but it is unclear to what degree these places represent novel environments.

To revise our understanding of the biogeography of Arabidopsis, we combined occurrences from its native range (including many outside Europe poorly represented in previous work) with climate data to build distribution models, and then test how model

predictions correspond to performance estimated from herbarium specimens and life history variation in natural accessions. We ask the following questions:

1. What climate factors constrain the distribution of Arabidopsis? We hypothesize that winter cold and summer drought are most important constraints on the species distribution, depending on region.
2. Are individual Arabidopsis performance and life history associated with model-estimated habitat suitability? We hypothesize that individuals reach larger sizes in regions with greater suitability and that populations adapt along gradients in suitability through changes in flowering time, a key component of life history.
3. Are occurrences outside the native range predicted by a native range model, suggesting stable realized niches following colonization? Or is there evidence Arabidopsis has colonized novel environments?
4. Where did Arabidopsis persist during the Last Glacial Maximum (LGM)? And where will Arabidopsis move in future climates?

Methods

Occurrence data

We developed a set of high-quality occurrence data (*i.e*. species ID verified and location checked, N=4,024) from published research (Durvasula et al., 2017; Hsu et al., 2019; Mandáková et al., 2017; Zeng et al., 2017; Zou et al., 2017), publicly available herbarium and germplasm accessions with known collection locations (Alonso-Blanco et al., 2016; DeLeo et al., 2020), and some of our recent field collections in East Africa (Gamba et al. 2022), altogether spanning the years from 1794 to 2018. The herbarium specimens and new collections include samples from little-studied populations in Saudi Arabia, Somalia, Djibouti, Eritrea, Rwanda, Ethiopia, Uganda, Sudan, and Nepal. Duplicate occurrence points were eliminated (samples are often split and sent to different herbariums). For model fitting, we excluded occurrences from putative non-native regions (the Americas, New Zealand, Japan).

We also used occurrence data (with coordinates and without flagged problems N=115,226) from the Global Biodiversity Information Facility (GBIF) to test model predictions in regions outside of the native range of Arabidopsis (downloaded 30 Dec 2020, Gbif.Org, 2020). We deem these occurrences as lower quality given that many have not had the species identity and location closely checked (DeLeo et al., 2020).

Environmental data

Climate data were extracted from CHELSA (Climatologies at High resolution for the Earth's Land Surface Areas) at 30 arc second spatial resolution (Karger et al., 2017). Current conditions are the average of 1979-2013 estimates. We selected the following climate variables based on hypothesized ecological importance (Gienapp et al., 2017; Hancock et al., 2011; Lasky et al., 2014, 2018) and relatively low inter-correlation (Pearson correlation coefficients among variables at occurrences all < 0.75): isothermality (Bio3), minimum temperature of coldest month (Bio6), temperature annual range (Bio7), mean temperature of wettest quarter (Bio8), mean temperature of the warmest quarter (Bio10), precipitation seasonality (Bio15), precipitation of wettest quarter (Bio16), and precipitation of driest quarter (Bio17). We also included altitude from Hijmans et al. (2005).

For projecting past distributions, we obtained climate estimates from the last glacial maximum (LGM) at 21k yrs before present (Karger et al., 2017). We used the global altitude and bathymetry map with 15 arc second resolution from Tozer et al. (2019) with a sea level 125 m lower than present (Poore et al., 2011) to project potential suitable habitat at the LGM on land in areas currently submerged.

For projecting future distributions, we used climate projections for 2050 from CHELSA v1.2 (Karger et al., 2017). We separately used five divergent global climate models from the fifth phase of the Coupled Model Integration Project (CMIP5) under Relative Concentration Pathway (RCP) 4.5 (an intermediate emissions scenario): ACCESS-1-0, MIROC5, CESM1-BCG, CMCC-CM, and MPI-ESM-MR.

For our analysis of climate anomaly effects on individual performance of herbarium specimens, we used the Climate Research Unit (CRU) TS 4.01 dataset, providing a global time series of monthly temperature and precipitation for the period 1900-2010 at a 0.5º resolution (Harris et al., 2014). From the CRU data we calculated the same bioclimatic parameters that we used in the Maxent distribution model, but in the CRU data these bioclimatic variables were specific to each individual herbarium specimen in the time period it was collected (Supplemental Methods) (DeLeo et al., 2020). We then calculated local anomalies for each of these variables by taking the observed value, subtracting mean across the entire time series, and dividing by the standard deviation (DeLeo et al., 2020).

Individual performance estimates from herbarium specimens

We investigated whether predictions from our models corresponded to performance variation among Arabidopsis individuals in the wild. We estimated fecundity on herbarium specimens using two traits. First, we measured the length of the longest inflorescence, reasoning that longer inflorescences would have more fruits and seeds. Second, we measured maximum rosette leaf length, reasoning that larger rosettes would support later reproductive investment if these collected plants were allowed to continue growth *in situ* (see Supplemental Methods for a validation). We used ImageJ on specimen images to estimate the tallest point of each inflorescence and the longest rosette leaf length (see Supplemental Methods for more detail). This yielded 1,264 measurements of maximum rosette leaf length and 2,188 measurements of inflorescence height on 2,194 herbarium specimens.

Range-wide genetic variation in life history

Many late-flowering Arabidopsis genotypes require cold cues (vernalization) to flower and also show slower growth and more stress tolerance, delineating a major life history axis (Lovell et al., 2013; Vasseur et al., 2018). To test associations between life history variation and predicted habitat suitability, we used published experimental data on flowering time for 898 whole-genome resequenced accessions from the native range with reliable geographic coordinates grown at 10 and 16ºC (The 1001 Genomes Consortium 2016). To estimate vernalization sensitivity, we calculated the difference between days to flower at 10 and 16ºC.

Species Distribution Modeling

We thinned the high quality occurrence points to one sample per 1 km grid cell to reduce sampling bias (N=672) using the ‘gridsample’ and ‘aggregate’ functions in the ‘sp’ package

(Bivand et al., 2008). To characterize potentially inhabited environments across the studied region, we generated pseudoabsence background points using the 'dismo' package (Hijmans & Elith, 2013) by sampling 10,000 random points within a 500 km buffer around occurrence points.

We used Maxent version 3.4.0 to generate a species distribution model (Phillips et al., 2006). MaxEnt was implemented with the R package 'ENMeval' v1, and parameters were optimized using the 'checkerboard2' method for cross validation (Muscarella et al., 2014). Among the tested settings (ENMeval defaults), we chose the model with lowest AICc value and used this to project habitat suitability under recent (1978-2013) conditions across the globe, including the regions where Arabidopsis is not native. For all models we used the logistic output of MaxEnt that scales suitability from zero to unity.

We used the permutation importance of each climate factor to determine which climatic factors drove predictions in the Arabidopsis distribution model. We also used the 'limiting' function in the R package 'rmaxent' to determine the most limiting climatic factors in each location (Baumgartner et al., 2017). The limiting factor is the environmental covariate that has the largest decrease in suitability at a given location relative to the suitability value achieved if a given environmental covariate had the value equal to the global mean (Elith et al., 2010).

In non-native regions, we evaluated whether Arabidopsis is limited from further expansion at range edges by climate, *i.e.* whether there were no more unoccupied climatically-suitable environments near existing populations. To do so, we calculated habitat suitability in a zone from 50 to 100 km distant from existing GBIF occurrences and compared occupancy in these buffers in the native range to invaded regions.

We also projected the optimized MaxEnt model using the past (LGM) and future climate conditions described above. For future conditions, we calculated the mean and standard deviation of habitat suitability projected for the five global climate models. To assess whether model predictions were extrapolating into poorly characterized or novel climates, we compared climates for prediction to the present-day model training data. We implemented this comparison by calculating multivariate environmental similarity surfaces (MESS) following Elith et al. (2010).

Plant performance, life history, and predicted habitat suitability

We asked whether the model-predicted habitat suitability corresponded to plant size, suggesting its influence on individual fitness. We first tested these relationships with Pearson's and Spearman's correlations. We also fit Generalized Additive Models (GAMs) where herbarium specimen sizes were the response variable. The model included covariates for the model-estimated habitat suitability at the collection location (square-root transformed to reduce the lower-tail suitability influence) and (as a nuisance variable) the year of collection (scaled to mean zero and unit variance) to account for potential changes in average size in recent decades. We used GAMs to allow smooth spatial variation in coefficients, so that we could capture any regional variation in covariate effects. The model can be represented as:

$$\text{Phenotype}_{i,year} = \beta_0 + \beta_{i,year}\text{Year} + \beta_{i,HS}\text{Habitat Suitability}_i + \varepsilon_{i,year}$$

where *i* represents the location of the collected record. Spatially-varying parameters were constrained to have smooth spatial variation using the 'mgcv' package for R, which we used with REML to fit all GAMs (Wood, 2011). We considered covariates to be significant in a given

location if their 95% confidence interval (CI) did not overlap with 0, and the Moran's I (Cooper, 2021) for model residuals did not indicate spatial autocorrelation (all $p > 0.38$).

We also asked whether local limiting factors corresponded to climate variables with stronger effects on plant fecundity in a given locale compared to elsewhere. To do so, we compared herbarium specimen sizes to anomalies in climate conditions corresponding to the climate variables used in the MaxEnt model. We used these yearly climate anomalies in GAMs with spatially varying coefficients, where for a phenotype measured at location *i*:

$$\text{Phenotype}_{i,year} = \beta_{i,0} + \beta_{i,year}\text{Year} + \beta_{i,climate}\text{BioClim Variable Anomaly} + \varepsilon_{i,year}$$

We used three strategies to test how life history varies with model-predicted suitability. First, we estimated simple linear correlations between suitability and flowering time. Second, to account for population structure and neutral processes that can affect spatial variation in flowering time, we tested whether suitability was associated with flowering time after using random effects to account for genome-wide similarity among accessions. Such an association would suggest selection linked to suitability acts on flowering time. This test is akin to Q_{ST}-F_{ST} contrasts (Whitlock & Guillaume, 2009), except that an explicit environmental gradient is tested (here, suitability) as opposed to studying trait variance among discrete populations. We implement this mixed model in R using the function 'lmekin' from the package coxme (Therneau & Therneau, 2015), along with the package kinship2 (Sinnwell et al., 2014). The kinship matrix was obtained from a published whole genome resequencing SNP data (Alonso-Blanco et al., 2016; Durvasula et al., 2017). Third, to test for geographic variation in suitability-flowering time relationships, we implemented GAMs of flowering time with spatially varying coefficients:

$$\text{FloweringTime}_{i,year} = \beta_0 + \beta_{i,HS}\text{Habitat Suitability} + \varepsilon_i \ .$$

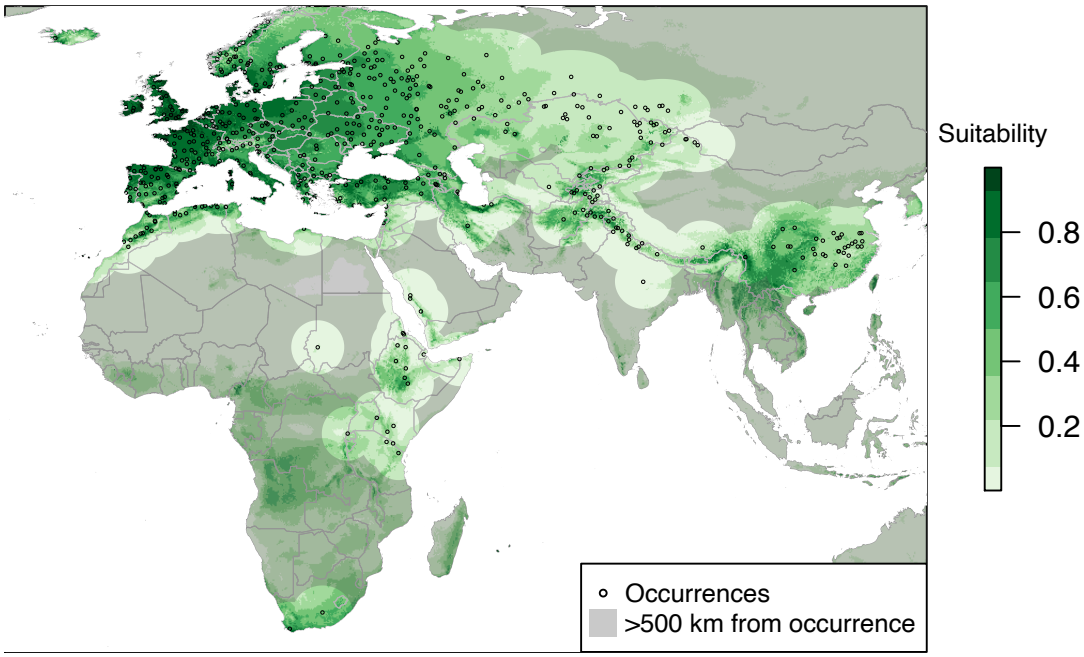
Results

Q1. Climate constraints on the distribution of Arabidopsis

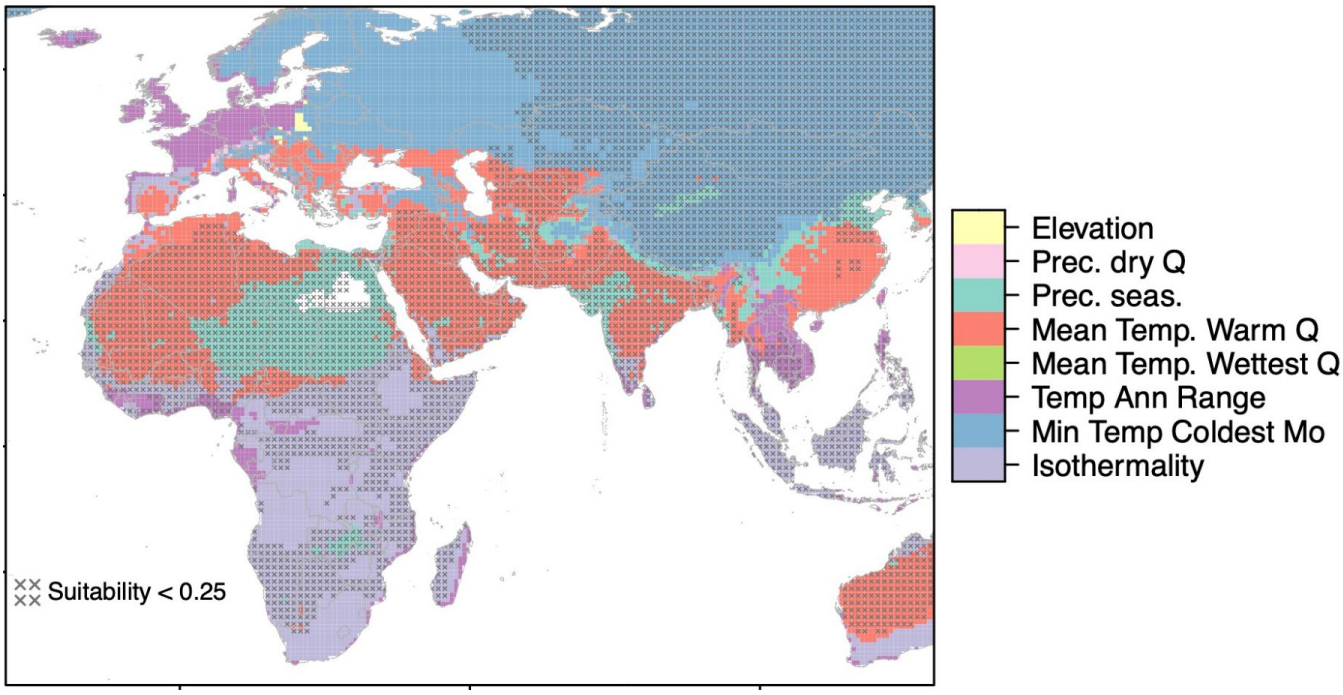
Our optimization of the MaxEnt models resulted in a selection of a model with linear and quadratic effects, with the regularization multiplier of 0.5. The training AUC value was 0.77 and the average test AUC with checkerboard2 cross validation was 0.77 (Figure 1A). The omission rate of the 10th percentile of suitability for training points was 0.10, which suggests our models were not overfit as they were able to predict low probability occurrences as well as expected (Fielding & Bell, 1997; Peterson et al., 2011). The areas of high suitability overall correspond well to the documented Arabidopsis distribution, with one notable exception being tropical lowland sites (Congo basin) in sub-Saharan Africa. The MESS map showed this region was near zero multivariate environmental similarity, indicating the model may have been poorly constrained in this region, while most of the regions of high predicted suitability where Arabidopsis is documented have positive similarity (Figure S2).

Figure 1. (A**)** Habitat suitability (green) under current climate conditions with thinned occurrences (n = 672) used in fitting shown in black circles and (B) limiting factors from the MaxEnt model fit to current climates. Regions far (>500 km) from known occurrence have a gray mask in (A). Regions of low suitability (less than 0.25) in (B) are makred with gray 'x' symbols.

A.



B.



The permutation importance of the distribution model covariates across the entire native range revealed that the minimum temperature of the coldest month (permutation importance = 37%)

and the mean temperature of the warmest quarter (permutation importance = 34%) were the two most important variables overall, suggesting winter cold stress and summer heat stress (or correlated factors such as drought) are most important in constraining Arabidopsis's distribution (Table S1). The next most important variable was isothermality (permutation importance =13%), as Arabidopsis tends to be found in regions with low isothermality (e.g. most temperate zones).

To identify spatial variation in the variables constraining model predictions, we also identified local limiting factors (Figure 1B). Across northern Eurasia, minimum temperature of the coldest month limited habitat suitability. Across the Mediterranean and tropical/subtropical regions, the temperature of the warmest quarter was limiting. Notably, across Eastern Europe and Central Asia, winter cold was limiting in regions adjacent to other regions where summer heat was limiting, highlighting the multiple potential stressors in this region. In Atlantic Europe, where conditions were highly favorable, a low annual range of temperature was limiting. A low annual range of temperature was also limiting in tropical lowland moist sites in Africa and SE Asia, and high isothermality (low seasonality) of temperature was limiting in other tropical sites, both where Arabidopsis is largely absent (and showing near zero on the MESS map, Figure S2).

We hypothesized that for a given region, temporal fluctuations in local limiting factors would be associated with temporal variation in individual plant size in herbarium specimens. We found that years with warmer minimum winter temperatures were associated with significantly taller inflorescences from Turkey to central Asia (Figure S3). Concordantly, in much of this region the Maxent model indicated that minimum temperature of the coldest of the month was the most limiting factor (namely in the Caucuses and the region from Kazakhstan to northern India, Figure 1B). We also found that years with greater seasonality of precipitation were associated with shorter inflorescences in central Asia, but taller inflorescences in central Europe (Figure S3). Partly consistent, in the southern part of this region in central Asia the MaxEnt model indicated precipitation seasonality was limiting (Figure 1B). Years with higher isothermality were positively associated with inflorescence height in Eastern Europe but this was not a limiting factor in this region (Figure S3). Other climate anomalies were not significantly associated with temporal variation in plant size.

Q2. Evidence of declining fitness and stress escape life history in regions of low predicted habitat suitability

As model-estimated habitat suitability increased, so did the size of individual plants in herbarium specimens, measured as inflorescence height (Pearson's $r = 0.11$, $p < 10^{-6}$; Spearman's rho = 0.08, $p < 10^{-4}$; $n = 2053$). The relationship with suitability was stronger for maximum rosette leaf length ($r = 0.24$, $p < 10^{-16}$; rho = 0.24, $p < 10^{-16}$; $n = 1179$, Figure 2), which was also more correlated with total fruit + flower number (see Supplement), suggesting fecundity was greater in regions of high predicted habitat suitability. We also tested these relationships using GAMs with spatially varying suitability coefficients (but not including climate anomalies – distinct from the above section Q1 on limiting factors). We found a consistently significant positive relationship between suitability and maximum rosette leaf length across Europe and inflorescence height across Eurasia (Figs S5 and S6). Unexpectedly, in these models where year of collection was considered primarily as a nuisance variable, plant size was significantly related to year of collection in northwest Europe (Figs S5 and S6), with declining size over time, perhaps because

of changing climate or evolving life histories (e.g. within the significant region in Scandinavia, maximum rosette leaf length vs year Spearman's rho = -0.17) (DeLeo et al., 2020).

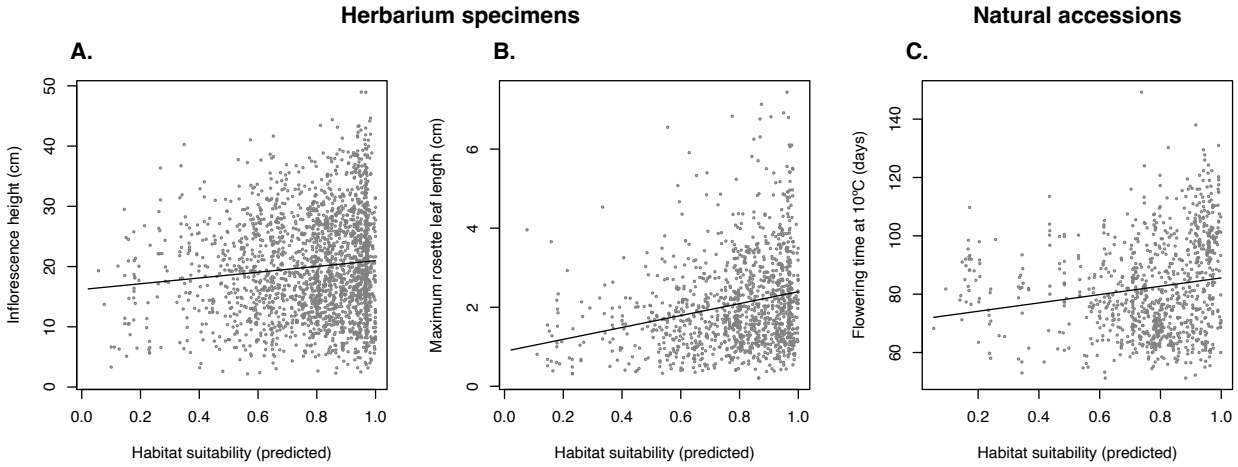


Figure 2. Predicted current habitat suitability compared with individual plant size (A & B) and a measure of life history (C). Size measures include inflorescence height (A, relationship with suitability: Pearson's r = 0.11, Spearman's rho = 0.08) and maximum rosette leaf length (B, r = 0.24, rho = 0.24) from herbarium specimens. Flowering time of natural accessions (C, r = 0.17) was taken from published data on a growth chamber experiment at 10ºC (Alonso-Blanco et al., 2016). Linear model fits are shown. N = 2053 for inflorescence height, N = 1179 for maximum rosette leaf length, and N=898 for flowering time.

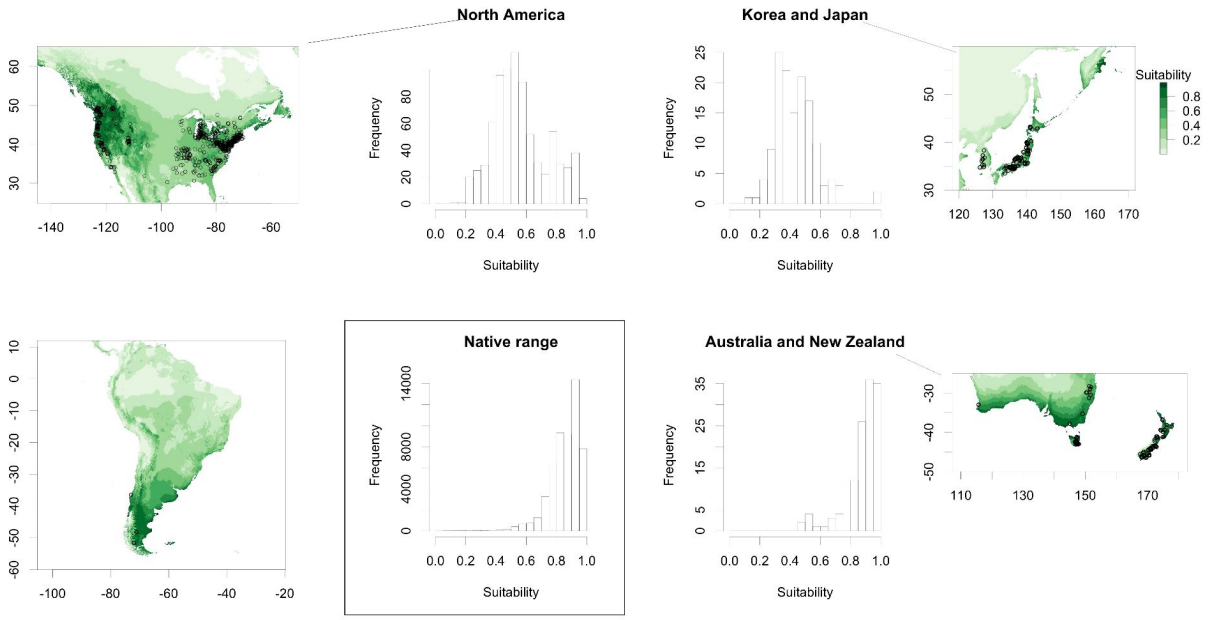
We compared habitat suitability with published data on flowering time, a key axis of Arabidopsis life history. We found that flowering time at 10ºC and flowering time plasticity were significantly and positively associated with predicted suitability (r=0.17, p<0.0001 and r=0.11, p=0.001, respectively), but not for days to flower at 16 ºC (r=0.05, p=0.15, Figures 2 & S7). In GAMs with spatially-varying effects of suitability on flowering time, these patterns were largely consistent across Eurasia (Figure S8). The flowering time-suitability association was significant even when accounting for genomic similarity among accessions, suggesting selection associated with suitability could maintain geographic clines in flowering time. Specifically, linear mixed-effects models found a positive association for flowering time at 10ºC (p=0.0001, n=898), and also at 16ºC, the latter of which may have been obscured by population structure that was unaccounted for in the simple linear regression model (p=0.002, n=898). The suitability association was not significant for plasticity in flowering time when accounting for genomic similarity (p=0.66, n=898) (Tables S1-S3). Consistent with this potential obscuring of population structure, we found that 10 ADMIXTURE genetic clusters in Arabidopsis (Alexander & Lange, 2011; Alonso-Blanco et al., 2016) were significantly different in their predicted habitat suitability (F(9,888)=191.1, p<0.0001, Table S5, Figure S9).

Q3. Using the native range model to predict occurrences outside the native range

We used the occurrences from GBIF to explore whether our native-range model predicted occurrences across the non-native range. Australia and New Zealand occurrences were largely in areas predicted to be highly suitable (Figure 3). By contrast, North America includes a group

of occurrences in highly suitable areas (Pacific Northwest) and another in low (but non-zero) suitability areas (interior east, where summer heat was predicted limiting, Figure S10). The eastern US also has near zero similarity on the MESS map, suggesting a relatively novel set of conditions compared to the native range (Figure S2). Similarly, the regions where Arabidopsis occurred in Korea and Japan were of lower suitability than in the native range, also with summer heat as the predicted limiting factor (Figure S10), suggesting Arabidopsis in eastern North America and East Asia inhabits climates with distinctly warm summers. There were too few occurrences in South America to make strong conclusions.

Figure 3. Distribution of predicted habitat suitability (based on our native-range MaxEnt model, underlying map surfaces in green) for GBIF occurrences in various regions (black circles). We do not include a histogram for South America because there are too few occurrences.



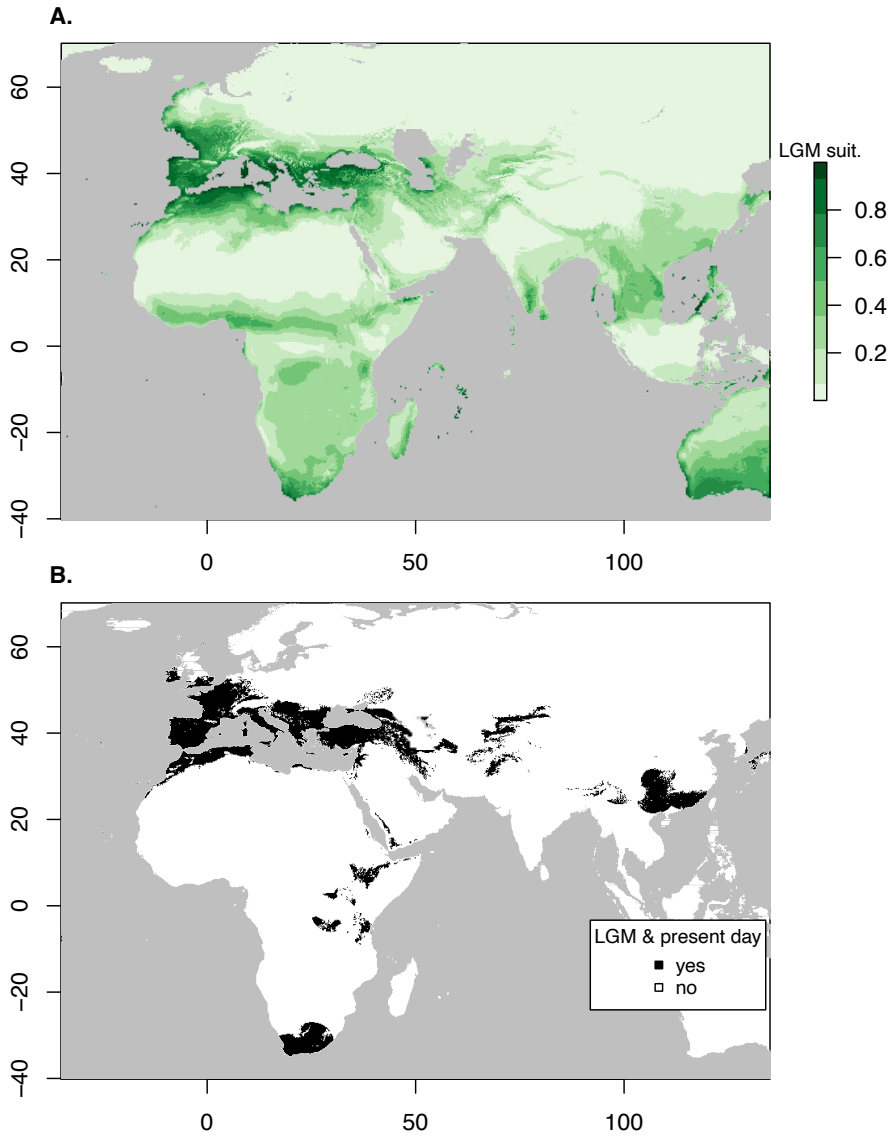
We investigated whether non-native Arabidopsis occupy all available habitat in their regions or whether apparently suitable habitat remains unoccupied. We found that locales >50 km but <100 km from occurrences included many areas of high suitability in North America, suggesting suitable sites remain unoccupied (12.7% of grid cells in these 50-100 km regions had suitability > 0.8). In comparison there were fewer such sites in the native range (4.4% of these 50-100 km regions had suitability > 0.8). There were extensive areas of high suitability along the Pacific coast of North America to Alaska around 60ºN, but no occurrences north of 50º latitude, which we confirmed with Alaskan expert botanists (pers comm. Matthew Carlson). The timing of the invasion is likely not a factor given that occurrences from the Vancouver area date at least to 1939. This region has near zero similarity on the MESS map, potentially indicating truly unsuitable habitat poorly modeled by the native range model (Figure S2). Similarly, the southern coast of Australia is highly suitable but Arabidopsis is apparently absent (34.6% of these 50-100 km regions with suitability > 0.8). Records are restricted to the southeast and southwest (confirmed by botanists, Shelley James, Tim Entwistle, Neville Walsh pers. comm.),

even though records date at least to 1959 in SE Australia. In South America there are a few Arabidopsis records from the Southern Cone, and it appears to be rare in the region (pers. comm. Diego Salariato), while apparently suitable environments occur throughout Patagonia and high elevation Andean sites that are apparently unoccupied (pers comm. Gwendolyn Peyre, Santiago Madriñán). Additionally, large areas of southern Australia and South America show positive similarity on the MESS map suggesting the model is well constrained in those regions. By contrast in Korea and Japan there are very few sites expected to be highly suitable that are not already occupied by Arabidopsis (2.3% of these 50-100 km regions with suitability > 0.8).

Q4. The distribution of Arabidopsis during the Last Glacial Maximum and in the future

We projected suitability onto the climate at the LGM 21 kya and found several putative refugia, where past environmental conditions could have supported Arabidopsis persistence. In particular, the Mediterranean/Caucuses/south Caspian Sea, much of sub-Sharan Africa > 1000 m (current) asl, and China and SE Asia appear as refugia (Figure 4). The islands of Sicily, Corsica, and Sardinia, and North Africa appear as highly suitable potential refugia as well. We also found that the Atlantic European coast was highly suitable during the LGM.

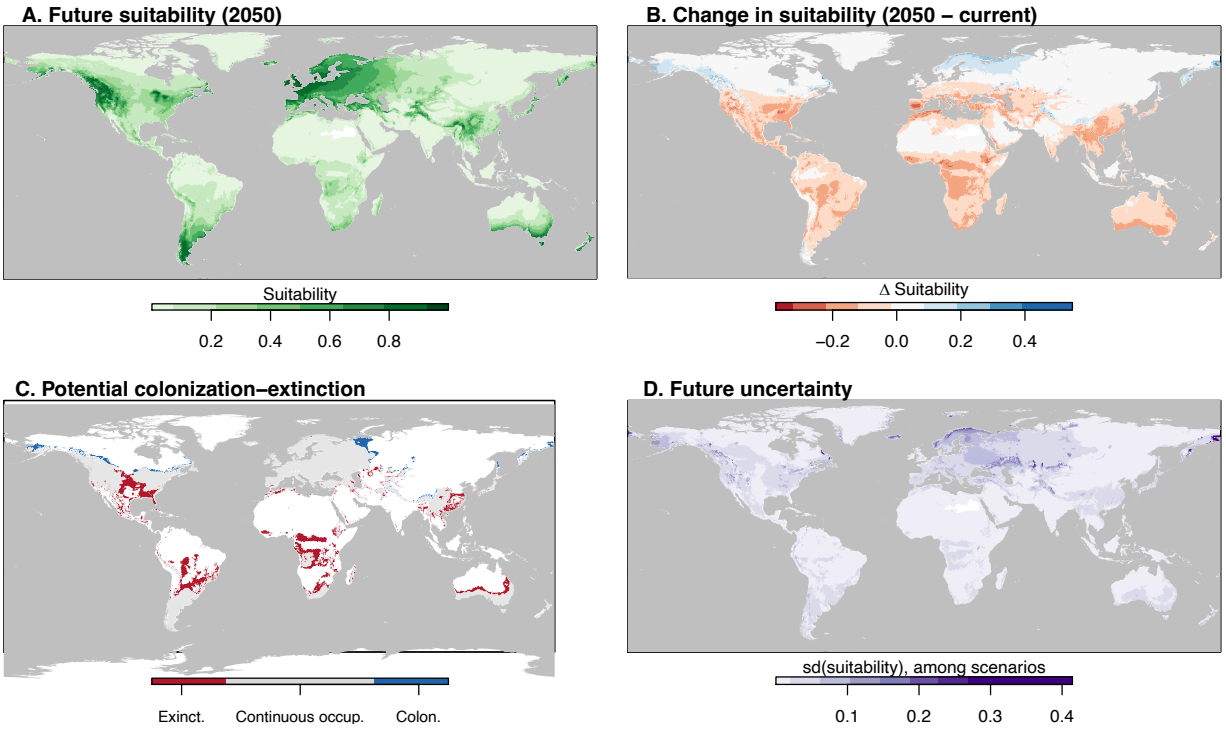
Figure 4. (A) Predicted distribution during the Last Glacial Maximum and (B) areas in black with suitability > 0.25 during both the LGM and current conditions as well as within 500 km of known current occurrences. Because of their greater level at the LGM, in (A) we masked the LGM Caspian and Aral Seas (including regions of high putative suitability) from the map (Prentice et al., 1993). Lake Victoria was left unmasked as it was likely very low during the LGM (Johnson et al., 1996).



We projected climate suitability for Arabidopsis in the year 2050. A coarse global view shows some of the same regions still having high suitability, such as in Northern Europe in the native range. At the same time, we generally found poleward and up-elevation shifts in regions of high suitability and retreats at lower latitudes and elevations. In the native range, all African sites show declining suitability, as did most Mediterranean sites, highlighting potential vulnerability of these populations. By contrast, northern Scandinavia, NW Russia, and the mountains of central

Asia/Tibetan plateau show increased suitability, highlighting regions for potential range expansion (Figure 5).

Figure 5. Future predicted climate suitability for Arabidopsis (A), change in suitability from future compared with present (blue indicates improving suitability and red decreasing, B), regions of potential colonization (blue) continued occupancy (gray), and extinction (red) based on a threshold suitability of 0.25 for occupancy (C), and the standard deviation in suitability among the 5 tested climate models giving uncertainty (D).



Discussion

We used the model plant Arabidopsis in a case study of climate biogeography of a species' past, present, and future distributions. This species is key to a large body of plant biology research, and an in-depth study of its biogeography provides important context for the biology of Arabidopsis. We found that the most important climate variables constraining the Arabidopsis distribution were winter cold and summer heat, depending on the region. The size of individual plants in herbarium specimens, as well as genetic variation in flowering time, suggested that regions with lower predicted suitability harbored populations with reduced fitness and altered life history. We identified new glacial refugia, and looking to the future, we found that genetically distinct lower latitude populations are most threatened by climate change.

The current climatic distribution of Arabidopsis and its limiting factors

The fitted native-range model largely corresponded to the occurrences and indicates Arabidopsis broadly distributed across Europe, moister regions of central and eastern Asia, and mountains across Africa. Our results advance beyond the most recent climate biogeographical study of Arabidopsis by Hoffmann (2002), who used monthly climate data (but notably no synthetic bioclimatic variables) from Leemans & Cramer (1991) mostly to qualitatively describe the regions occupied by Arabidopsis in Eurasia, north Africa, and North America. Our study included populations in sub-Saharan Africa and Asia that were overlooked or considered non-native by Hoffmann (2002). We also included many occurrence records apparently missed by Hoffmann (2002). Zou et al. (2017) more recently built models of Arabidopsis's distribution using Eurasian populations and default MaxEnt settings, but did not include occurrences from Arabia, sub-Saharan Africa, and much of the Himalayas, Russia, and central Asia, and did not subsample occurrence to reduce bias. Likely as a result of these issues, the predictions of Zou et al. (2017) show a pronounced peak in suitability in Germany, which was densely sampled in the genomics studies used for occurrences by Zou et al. (2017), but predicted low suitability in much of the core European range and near zero suitability in sub-Saharan Africa, Arabia, and most of the Russian part of the species range.

The two dominant limiting factors in our MaxEnt model across the range were winter cold (at lower latitudes and elevations) and summer heat (at higher latitudes and elevations). Winter cold is recognized to impose limitations on Arabidopsis, in particular winter cold appears to be a dominant force in local adaptation of populations of Arabidopsis (Ågren & Schemske, 2012; Gienapp et al., 2017; Monroe et al., 2016). Because the phenology of Arabidopsis in southern Europe (where summer heat was inferred to be limiting) is such that plants often flower in early spring (DeLeo et al., 2020), summer heat is not directly experienced by many vegetative plants. Nevertheless, the worsening conditions in late spring may constrain plants in these regions. At these times, warm temperatures do not reach consistently stressful levels (e.g. to induce fruit abortion) (Warner & Erwin, 2005) but it may be that moisture limitation driven by evaporative demand is directly limiting. In the regions where MaxEnt suitability was estimated to be highest for Arabidopsis, a region including the British Isles and a belt along the coast from France to Poland, temperature annual range was identified as limiting. However, interpreting limiting factors in an area of extremely high suitability (near one) is not meaningful given that suitability can scarcely be increased.

Despite MaxEnt inference of winter cold and summer heat as primarily limiting, these were only partly reflected by temporal fluctuations in individual plant performance from herbarium specimens. Winter cold and precipitation variability anomalies were significantly associated with specimen size in much of Asia, where these were also the limiting factors in the Maxent model from Iran and Kazakhstan to Afghanistan and the Himalayas, suggesting these climate factors truly limit Arabidopsis populations and that perhaps occurrence models can be used to infer limiting factors. However, there were some discrepancies between individual performance and model-estimated limiting factors, likely for several reasons. First, the size of specimens is an imperfect fitness proxy. For annuals with indistinct vegetative traits like Arabidopsis, usually only reproductive individuals are collected, and individuals that die before reproduction and collection are missed, and our size measurements are an imperfect measure of fecundity. Second, individuals are not randomly sampled from populations (Daru et al., 2018). Third, MaxEnt models face limitations due to misspecification, problems with occurrence data,

or a mismatch between covariates and the true ecological factors limiting populations. Nonetheless, the limiting factors we identified here fit well with our knowledge of Arabidopsis ecophysiology and natural history.

Habitat suitability, individual performance, and life history

We found a decrease in plant size with decreasing habitat suitability across the range of Arabidopsis, suggesting that our model suitability captured a substantial part of environmental effects on individual fitness. Given the great variability in individual plant performance within populations obvious to casual observers, it is unsurprising that suitability explains a minority of total variation in individual size leaving most variation unexplained. Furthermore, fecundity and fecundity response to treatments often have low heritability (Price & Schluter, 1991), even in controlled experiments (Lasky et al., 2015).

There have still been few studies of individual performance with the geographic scope that allows inference across a species range (Angert & Schemske, 2005; Csergő et al., 2017; Greiser et al., 2020; Samis & Eckert, 2007). In a synthesis of published experimental studies of 40 species, Lee-Yaw et al. (2016) found that on average individual performance measures and distribution-model inferred suitability decline beyond range margins. From 42 studies Lee-Yaw et al. (2021) found that 38% identified some predictive ability of distribution models for individual performance. However, many previous studies relied on intensive observations of a small number of populations, while our estimates of performance from thousands of herbarium specimens allowed us to cover most of the species range and observe where specific performance measures did (and did not) correspond to suitability. The increased availability of digitized museum specimens trait data indicates an opportunity to estimate performance across distributions for many species (Bontrager & Angert, 2015).

We found that low suitability regions tended to have earlier flowering genotypes. This relationship was noisy, with early flowering genotypes frequent in all levels of suitability but later flowering ecotypes lacking from the least suitable regions. The suitability-flowering time associations were significant when accounting for genome-wide similarity between accessions, suggesting they are not merely due to drift but instead reflect selection associated with suitability. We interpret the direction of the relationship as indicating that when suitability is low, Arabidopsis employs stress escape strategies, *i.e*. a rapid life cycle during favorable conditions (Ludlow, 1989). The restriction of late flowering genotypes to more suitable regions is counterintuitive given physiological work showing these are more stress tolerant (Lovell et al., 2013). However, it may be true that more favorable conditions make possible a slow growing, slow flowering, freezing-tolerant winter annual strategy. Note that our life history data did not include sub-Saharan populations, which have not been evaluated for flowering time.

The distribution of Arabidopsis outside its native range

Arabidopsis has spread across the globe, largely to climates well-predicted by native range climate models, but to some regions predicted to have less suitable climates. Populations in western North America, Australia, and New Zealand, occur in climates well-predicted by the native range, suggesting stable realized niches following colonization. But in eastern North America, Korea, and Japan, these climates tend to be of low predicted suitability. Whether these are truly of lower suitability is unknown without performance data. It may be that Arabidopsis

has colonized novel environments in these regions facilitated by plasticity and adaptation, or that the low suitability is only a model artefact. Published studies on other species often show limited transferability among the native and introduced range (Early & Sax, 2014; Liu et al., 2020).

In southern Australia, the Southern Cone, and Alaska there are widespread seemingly suitable habitats where Arabidopsis is absent. The lack of more populations may partly owe to the strict quarantine on imports to Australia to avoid species invasions (Pheloung, 1999). Additionally, a lack of human disturbance in some of these regions (e.g. Alaska) may limit the potential for Arabidopsis colonization given it often is found in anthropogenically disturbed sites.

The distribution of Arabidopsis at the last glacial maximum

Lee et al. (2017) hypothesized five glacial refugia for Eurasia using population genetic approaches to be in Iberia, Sicily, Balkans, the Levant, and Turkmenistan. These are consistent with our range reconstruction, although we do not find clear barriers of unsuitable climates that would have isolated Italian, Balkan, and Levant LGM populations. Furthermore, we note that Sardinia/Corsica appears as a refugium, and we note that narrow, highly suitable areas along the southern and eastern Caspian Sea could be the location of the hypothesized Turkmenistan refugium. Based on patterns of genome-wide diversity, Lee et al. (2017) hypothesized that the bulk of current European genotypes derive from an expansion originating on the northwest coast of the Black Sea, where we found a strip of highly suitable conditions for Arabidopsis during the LGM. Additional genetic studies hypothesized a long history of existence in sub-Saharan Africa (Durvasula et al., 2017) and conditions during the LGM suggest the species could have been much more widespread than currently. Similar to our findings, Chala et al., (2017) used models of the distribution of Afroalpine habitat generally during the LGM to find expanded areas in East Africa compared to present day, though some locations such as Jebel Marra, Sudan were still surrounded by landscapes of unsuitable habitat even at the LGM.

By contrast, we found that the North Atlantic European coast was highly suitable (including areas currently inhabited on the British Isles) during the LGM, but current British Isle populations do not show genetic signatures of refugia (Lee et al., 2017). Additionally much of France was highly suitable during the LGM, but current French populations do not show genetic signatures of refugia (Lee et al., 2017). In both cases, local, distinct genotypes that survived the LGM may have gone extinct following expansion by the now dominant genetic cluster ("non-relicts") (Fulgione & Hancock, 2018; Lee et al., 2017). The previous Arabidopsis distribution modeling of the LGM (Zou et al., 2017) did not identify France and NE Iberia as highly suitable during the LGM and did not mask the Caspian Sea (thus overinflating that refugium). Zou et al. (2017) also did not identify the Africa refugia (except for a small region in SE Africa) and potential refugia in mountains of SE Asia.

The future distribution of Arabidopsis

Over the next decades climate conditions are predicted to worsen for Arabidopsis across large areas of its range. Central Spain and mountains in north and east Africa and Arabia may see the most worsening. These climate change effects may already be emerging: our intensive field exploration in the Dai Forest of Djibouti in 2018 failed to yield any Arabidopsis, despite the presence of abundant (but likely more drought tolerant) annual mustard (Brassicaceae family)

species *Erucastrum arabicum* and *Sisymbrium erysimoides* in otherwise appropriate habitat for Arabidopsis. Arabidopsis was collected in this juniper woodland habitat in 1954 but recent decades have seen increased drought and tree mortality (Witsen, 2012), and the potential extinction of this isolated marginal Arabidopsis population. Given the unique genetic diversity in the model plant Arabidopsis housed in lower latitude Arabidopsis populations (Durvasula et al., 2017; Hsu et al., 2019; Lee et al., 2017; Zou et al., 2017) the conservation of these populations could benefit biology and agricultural research.

By contrast, conditions are expected to improve for the northernmost populations in Europe, suggesting a potential current colonization front of Arabidopsis, in addition to higher elevation locations in Tibet, the Caucuses, Ural, Alps, and Hengduan mountains adjacent to currently inhabited regions. Similar currently unoccupied higher elevations are available in many African mountains though these are usually small mountaintop areas and may be poorly characterized by CHELSA climate data 3/6/22 10:15:00 PM. Future distribution models would benefit from improved environmental data for these high elevation tropical sites.

Conclusion

Species distribution models provide potentially powerful windows into past, present, and future macroecology, but they are rarely confronted with individual performance data. Here we showed that lower suitability habitats inferred from distribution models had smaller plants with distinct life history, suggesting a stress escape strategy. While the relationships were noisy it may still be remarkable that they emerge against the many microsite contributors to individual level variation in a habitat generalist annual plant. Arabidopsis populations are distributed across diverse climates, but genetically distinct populations in lower latitudes that are potentially valuable for research are also highly threatened by anthropogenic climate change in the next few decades.

Acknowledgments

Victoria Meagher and Jaden Hill measured morphology on herbarium specimens. Patrick Herné assisted in digitizing many herbarium specimens and Eugene Shakirov translated many Russian specimen labels. Data from the MNHN in Paris were obtained from the participatory science program ‘Les Herbonautes’ (MNHN/Tela Botanica) which is part of Infrastructure Nationale e-RECOLNAT: ANR-11-INBS-0004. This work was supported by NIH award 1R35GM138300-01 to JRL. RM was supported by grant 2019-03758 from Vetenskapsrådet. We also thank volunteers for digitization and georeferencing through DigiVol, a website developed by the Australian Museum in collaboration with the Atlas of Living Australia.

References

Ågren, J., & Schemske, D. W. (2012). Reciprocal transplants demonstrate strong adaptive differentiation of the model organism Arabidopsis thaliana in its native range. *New Phytologist*, *194*(4), 1112–1122. https://doi.org/10.1111/j.1469-8137.2012.04112.x

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 246. https://doi.org/10.1186/1471-2105-12-246

Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K. M., Cao, J., Chae, E., Dezwaan, T. M., Ding, W., Ecker, J. R., Exposito-Alonso, M., Farlow, A., Fitz, J., Gan, X., Grimm, D. G., Hancock, A. M., Henz, S. R., Holm, S., … Zhou, X. (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell*, *166*, 481–491. https://doi.org/10.1016/j.cell.2016.05.063

Angert, A. L., & Schemske, D. W. (2005). The Evolution of Species' Distributions: Reciprocal Transplants Across the Elevation Ranges of Mimulus Cardinalis and M. Lewisii. *Evolution*, *59*(8), 1671–1684. https://doi.org/10.1111/j.0014-3820.2005.tb01817.x

Baumgartner, J., Wilson, P., & Esperón-Rodríguez, M. (2017). rmaxent: Tools for working with Maxent in R. *R Package Version 0.8*, *3*.

Bivand, R. S., Pebesma, E. J., Gómez-Rubio, V., & Pebesma, E. J. (2008). *Applied spatial data analysis with R* (Vol. 747248717). Springer.

Bontrager, M., & Angert, A. L. (2015). Effects of range-wide variation in climate and isolation on floral traits and reproductive output of Clarkia pulchella. *American Journal of Botany*, *103*, 10–21. https://doi.org/10.3732/ajb.1500091

Bontrager, M., Usui, T., Lee-Yaw, J. A., Anstett, D. N., Branch, H. A., Hargreaves, A. L., Muir, C. D., & Angert, A. L. (2021). Adaptation across geographic ranges is consistent with strong

selection in marginal climates and legacies of range expansion. *Evolution*, *75*(6), 1316–1333. https://doi.org/10.1111/evo.14231

Buckley, L. B., Waaser, S. A., MacLean, H. J., & Fox, R. (2011). Does including physiology improve species distribution model predictions of responses to recent climate change? *Ecology*, *92*(12), 2214–2221. https://doi.org/10.1890/11-0066.1

Chala, D., Zimmermann, N. E., Brochmann, C., & Bakkestuen, V. (2017). Migration corridors for alpine plants among the 'sky islands' of eastern Africa: Do they, or did they exist? *Alpine Botany*, *127*(2), 133–144. https://doi.org/10.1007/s00035-017-0184-z

Clark, J. S. (2010). Individuals and the Variation Needed for High Species Diversity in Forest Trees. *Science*, *327*(5969), 1129–1132.

Clauw, P., Coppens, F., Beuf, K. D., Dhondt, S., Daele, T. V., & Maleux, K. (2015). Leaf responses to mild drought stress in natural variants of Arabidopsis. *Plant Physiol.*, *167*. https://doi.org/10.1104/pp.114.254284

Cooper, M. (2021). *moranfast: Calculate Moran's I quickly with low memory footprint for Big Data.* [C++]. https://github.com/mcooper/moranfast

Csergő, A. M., Salguero-Gómez, R., Broennimann, O., Coutts, S. R., Guisan, A., Angert, A. L., Welk, E., Stott, I., Enquist, B. J., McGill, B., Svenning, J.-C., Violle, C., & Buckley, Y. M. (2017). Less favourable climates constrain demographic strategies in plants. *Ecology Letters*, *20*(8), 969–980. https://doi.org/10.1111/ele.12794

Daru, B. H., Park, D. S., Primack, R. B., Willis, C. G., Barrington, D. S., Whitfeld, T. J. S., Seidler, T. G., Sweeney, P. W., Foster, D. R., Ellison, A. M., & Davis, C. C. (2018). Widespread sampling biases in herbaria revealed from large-scale digitization. *New Phytologist*, *217*(2), 939–955. https://doi.org/10.1111/nph.14855

DeLeo, V. L., Menge, D. N. L., Hanks, E. M., Juenger, T. E., & Lasky, J. R. (2020). Effects of two centuries of global environmental variation on phenology and physiology of

Arabidopsis thaliana. *Global Change Biology*, *26*(2), 523–538. https://doi.org/10.1111/gcb.14880

Durvasula, A., Fulgione, A., Gutaker, R. M., Alacakaptan, S. I., Flood, P. J., Neto, C., Tsuchimatsu, T., Burbano, H. A., Picó, F. X., Alonso-Blanco, C., & Hancock, A. M. (2017). African genomes illuminate the early history and transition to selfing in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, 201616736. https://doi.org/10.1073/pnas.1616736114

Early, R., & Sax, D. F. (2014). Climatic niche shifts between species' native and naturalized ranges raise concern for ecological forecasts during invasions and climate change. *Global Ecology and Biogeography*, *23*(12), 1356–1365. https://doi.org/10.1111/geb.12208

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, *1*(4), 330–342. https://doi.org/10.1111/j.2041-210X.2010.00036.x

Estarague, A., Vasseur, F., Sartori, K., Bastias, C., Cornet, D., Rouan, L., Beurier, G., Exposito-Alonso, M., Herbette, S., Bresson, J., Vile, D., & Violle, C. (2021). *Into the range: A latitudinal gradient or a center-margins differentiation of ecological strategies in Arabidopsis thaliana?* (p. 2021.10.15.461205). https://doi.org/10.1101/2021.10.15.461205

Exposito-Alonso, M., Vasseur, F., Ding, W., Wang, G., Burbano, H. A., & Weigel, D. (2018). Genomic basis and evolutionary potential for extreme drought adaptation in Arabidopsis thaliana. *Nature Ecology & Evolution*, *2*(2), 352. https://doi.org/10.1038/s41559-017-0423-0

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, *24*(1), 38–49. https://doi.org/10.1017/S0376892997000088

Forester, B. R., DeChaine, E. G., & Bunn, A. G. (2013). Integrating ensemble species distribution modelling and statistical phylogeography to inform projections of climate change impacts on species distributions. *Diversity and Distributions*, *19*(12), 1480–1495. https://doi.org/10.1111/ddi.12098

Fournier-Level, A., Perry, E. O., Wang, J. A., Braun, P. T., Migneault, A., Cooper, M. D., Metcalf, C. J. E., & Schmitt, J. (2016). Predicting the evolutionary dynamics of seasonal adaptation to novel climates in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, *113*(20), E2812–E2821. https://doi.org/10.1073/pnas.1517456113

Fulgione, A., & Hancock, A. M. (2018). Archaic lineages broaden our view on the history of Arabidopsis thaliana. *New Phytologist*, *219*(4), 1194–1198. https://doi.org/10.1111/nph.15244

Gbif.Org. (2020). *Occurrence Download*. The Global Biodiversity Information Facility. https://doi.org/10.15468/DL.H95GHR

Gienapp, P., Fior, S., Guillaume, F., Lasky, J. R., Sork, V. L., & Csilléry, K. (2017). Genomic Quantitative Genetics to Study Evolution in the Wild. *Trends in Ecology & Evolution*, *32*(12), 897–908. https://doi.org/10.1016/j.tree.2017.09.004

Greiser, C., Hylander, K., Meineri, E., Luoto, M., & Ehrlén, J. (2020). Climate limitation at the cold edge: Contrasting perspectives from species distribution modelling and a transplant experiment. *Ecography*, *43*(5), 637–647. https://doi.org/10.1111/ecog.04490

Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G., Toomajian, C., Roux, F., & Bergelson, J. (2011). Adaptation to Climate Across the Arabidopsis thaliana Genome. *Science*, *334*(6052), 83–86.

Hargreaves, A. L., Samis, K. E., Eckert, C. G., Schmitz, A. E. O. J., & Bronstein, E. J. L. (2014). Are Species' Range Limits Simply Niche Limits Writ Large? A Review of Transplant

Experiments beyond the Range. *The American Naturalist*, *183*(2), 157–173. https://doi.org/10.1086/674525

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *International Journal of Climatology*, *34*(3), 623–642. https://doi.org/10.1002/joc.3711

Heberling, J. M. (2021). Herbaria as Big Data Sources of Plant Traits. *International Journal of Plant Sciences*. https://doi.org/10.1086/717623

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. *International Journal of Climatology*, *25*(15), 1965–1978.

Hijmans, R. J., & Elith, J. (2013). Species distribution modeling with R. *R CRAN Project*.

Hoffmann, M. H. (2002). Biogeography of Arabidopsis thaliana (L.) Heynh. (Brassicaceae). *Journal of Biogeography*, *29*(1), 125–134.

Hoffmann, M. H. (2005). EVOLUTION OF THE REALIZED CLIMATIC NICHE IN THE GENUS: ARABIDOPSIS (BRASSICACEAE). *Evolution*, *59*(7), 1425-1436.

Hsu, C.-W., Lo, C.-Y., & Lee, C.-R. (2019). On the postglacial spread of human commensal Arabidopsis thaliana: Journey to the East. *New Phytologist*, *222*(3), 1447–1457. https://doi.org/10.1111/nph.15682

Johnson, T. C., Scholz, C. A., Talbot, M. R., Kelts, K., Ricketts, R. D., Ngobi, G., Beuning, K., Ssemmanda, I., & McGill, J. W. (1996). Late Pleistocene Desiccation of Lake Victoria and Rapid Evolution of Cichlid Fishes. *Science*, *273*(5278), 1091–1093. https://doi.org/10.1126/science.273.5278.1091

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*(1), 1–20. https://doi.org/10.1038/sdata.2017.122

Koornneef, M., & Meinke, D. (2010). The development of Arabidopsis as a model plant. *The Plant Journal*, *61*(6), 909–921. https://doi.org/10.1111/j.1365-313X.2009.04086.x

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. *The American Naturalist*, *169*(5), E141–E157.

Lasky, J. R., Des Marais, D. L., Lowry, D. B., Povolotskaya, I., McKay, J. K., Richards, J. H., Keitt, T. H., & Juenger, T. E. (2014). Natural variation in abiotic stress responsive gene expression and local adaptation to climate in Arabidopsis thaliana. *Molecular Biology and Evolution*, *31*(9), 2283–2296. https://doi.org/10.1093/molbev/msu170

Lasky, J. R., Forester, B. R., & Reimherr, M. (2018). Coherent synthesis of genomic associations with phenotypes and home environments. *Molecular Ecology Resources*, *18*(1), 91–106. https://doi.org/10.1101/051110

Lasky, J. R., Hooten, M. B., & Adler, P. B. (2020). What processes must we understand to forecast regional-scale population dynamics? *Proceedings of the Royal Society B: Biological Sciences*, *287*(1940), 20202219. https://doi.org/10.1098/rspb.2020.2219

Lasky, J. R., Upadhyaya, H. D., Ramu, P., Deshpande, S., Hash, C. T., Bonnette, J., Juenger, T. E., Hyma, K., Acharya, C., Mitchell, S. E., Buckler, E. S., Brenton, Z., Kresovich, S., & Morris, G. P. (2015). Genome-environment associations in sorghum landraces predict adaptive traits. *Science Advances*, *1*(6), e1400218. https://doi.org/10.1126/sciadv.1400218

Lee, C.-R., Svardal, H., Farlow, A., Exposito-Alonso, M., Ding, W., Novikova, P., Alonso-Blanco, C., Weigel, D., & Nordborg, M. (2017). On the post-glacial spread of human commensal Arabidopsis thaliana. *Nature Communications*, *8*(1), 14458. https://doi.org/10.1038/ncomms14458

Leemans, R., & Cramer, W. P. (1991). The CLIMATE data base of monthly normals for globally gridded climatic variables (Version 2.1) [Data set]. In *Potsdam Institute for Climate Impact Research*. PANGAEA. https://doi.org/10.1594/PANGAEA.759491

Lee-Yaw, J. A., Kharouba, H. M., Bontrager, M., Mahony, C., Csergő, A. M., Noreen, A. M. E., Li, Q., Schuster, R., & Angert, A. L. (2016). A synthesis of transplant experiments and ecological niche models suggests that range limits are often niche limits. *Ecology Letters*, *19*(6), 710–722. https://doi.org/10.1111/ele.12604

Lee-Yaw, J. A., McCune, J., Pironon, S., & Sheth, S. N. (2021). Species distribution models rarely predict the biology of real populations. *Ecography*, *n/a*(n/a). https://doi.org/10.1111/ecog.05877

Li, G., Du, S., & Guo, K. (2015). Evaluation of Limiting Climatic Factors and Simulation of a Climatically Suitable Habitat for Chinese Sea Buckthorn. *PLoS ONE*, *10*(7), e0131659. https://doi.org/10.1371/journal.pone.0131659

Liu, C., Wolter, C., Xian, W., & Jeschke, J. M. (2020). Species distribution models have limited spatial transferability for invasive species. *Ecology Letters*, *23*(11), 1682–1692. https://doi.org/10.1111/ele.13577

Lopez, L., Bellis, E. S., Wafula, E., Hearne, S. J., Honaas, L., Ralph, P. E., Timko, M. P., Unachukwu, N., dePamphilis, C. W., & Lasky, J. R. (2019). Transcriptomics of host-specific interactions in natural populations of the parasitic plant purple witchweed (Striga hermonthica). *Weed Science*, *67*(4), 397–411. https://doi.org/10.1017/wsc.2019.20

Lovell, J. T., Juenger, T. E., Michaels, S. D., Lasky, J. R., Platt, A., Richards, J. H., Yu, X., Easlon, H. M., Sen, S., & McKay, J. K. (2013). Pleiotropy of FRIGIDA enhances the potential for multivariate adaptation. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1763), 20131043. https://doi.org/10.1098/rspb.2013.1043

Ludlow, M. M. (1989). Strategies of response to water stress. In K. H. Kreeb, H. Richter, & T. M. Hinckley (Eds.), *Structural and Functional Responses to Environmental Stresses* (pp. 269–281). SPB Academic.

Mandáková, T., Thorbjörnsson, H., Pisupati, R., Reichardt, I., Lysak, M. A., & Anamthawat-Jónsson, K. (2017). Icelandic accession of Arabidopsis thaliana confirmed with cytogenetic markers and its orign inferred from whole-genome sequencing. *Icelandic Agricultural Sciences*, *30*, 29–38. https://doi.org/10.16886/IAS.2017.03

Martínez-Berdeja, A., Stitzer, M. C., Taylor, M. A., Okada, M., Ezcurra, E., Runcie, D. E., & Schmitt, J. (2020). Functional variants of DOG1 control seed chilling responses and variation in seasonal life-history strategies in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, *117*(5), 2526–2534. https://doi.org/10.1073/pnas.1912451117

Merow, C., Latimer, A. M., Wilson, A. M., McMahon, S. M., Rebelo, A. G., & Silander Jr, J. A. (2014). On using integral projection models to generate demographically driven predictions of species' distributions: Development and validation using sparse data. *Ecography*, *37*(12), 1167–1183. https://doi.org/10.1111/ecog.00839

Monroe, J. G., McGovern, C., Lasky, J. R., Grogan, K., Beck, J., & McKay, J. K. (2016). Adaptation to warmer climates by parallel functional evolution of CBF genes in Arabidopsis thaliana. *Molecular Ecology*, *25*(15), 3632–3644. https://doi.org/10.1111/mec.13711

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, *5*(11), 1198–1205. https://doi.org/10.1111/2041-210X.12261

Peterson, A., Soberón, J., Pearson, R., Anderson, R., Martínez-Meyer, E., Nakamura, M., & Araújo, M. (2011). Ecological Niches and Geographic Distributions. In *Monographs in Population Biology* (Vol. 49). https://doi.org/10.1515/9781400840670

Pheloung, P. (1999). Preventing the introduction of potential new weeds to Australia. *Practical Weed Management: Protecting Agriculture and the Environment. 10th Biennial Noxious Weeds Conference, Ballina, Australia, 20th-22nd July 1999.*, 18–24.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Poore, R., Williams, R., & Tracey, C. (2011). *Fact Sheet fs002-00: Sea Level and Climate*. https://pubs.usgs.gov/fs/fs2-00/

Prentice, I. C., Sykes, M. T., Lautenschlager, M., Harrison, S. P., Denissenko, O., & Bartlein, P. J. (1993). Modelling Global Vegetation Patterns and Terrestrial Carbon Storage at the Last Glacial Maximum. *Global Ecology and Biogeography Letters*, *3*(3), 67–76. https://doi.org/10.2307/2997548

Price, T., & Schluter, D. (1991). On the Low Heritability of Life-History Traits. *Evolution*, *45*(4), 853–861. https://doi.org/10.2307/2409693

Rudman, S. M., Greenblum, S., Hughes, R. C., Rajpurohit, S., Kiratli, O., Lowder, D. B., Lemmon, S. G., Petrov, D. A., Chaston, J. M., & Schmidt, P. (2019). Microbiome composition shapes rapid genomic adaptation of Drosophila melanogaster. *Proceedings of the National Academy of Sciences*, *116*(40), 20025–20032. https://doi.org/10.1073/pnas.1907787116

Samis, K. E., & Eckert, C. G. (2007). Testing the abundant center model using range-wide demographic surveys of two coastal dune plants. *Ecology*, *88*(7), 1747–1758. https://doi.org/10.1890/06-1153.1

Scholl, R. L., May, S. T., & Ware, D. H. (2000). Seed and Molecular Resources for Arabidopsis. *Plant Physiology*, *124*(4), 1477–1480. https://doi.org/10.1104/pp.124.4.1477

Sinnwell, J. P., Therneau, T. M., & Schaid, D. J. (2014). The kinship2 R Package for Pedigree Data. *Human Heredity*, *78*(2), 91–93. https://doi.org/10.1159/000363105

Takou, M., Wieters, B., Kopriva, S., Coupland, G., Linstädter, A., & De Meaux, J. (2019). Linking genes with ecological strategies in Arabidopsis thaliana. *Journal of Experimental Botany*, *70*(4), 1141–1151. https://doi.org/10.1093/jxb/ery447

Therneau, T. M., & Therneau, M. T. M. (2015). Package 'coxme.' *R Package Version*, *2*(5).

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., de Siqueira, M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Townsend Peterson, A., Phillips, O. L., & Williams, S. E. (2004). Extinction risk from climate change. *Nature*, *427*(6970), 145–148. https://doi.org/10.1038/nature02121

Toledo, B., Marcer, A., Méndez-Vigo, B., Alonso-Blanco, C., & Picó, F. X. (2020). An ecological history of the relict genetic lineage of Arabidopsis thaliana. *Environmental and Experimental Botany*, *170*, 103800. https://doi.org/10.1016/j.envexpbot.2019.103800

Tozer, B., Sandwell, D. T., Smith, W. H. F., Olson, C., Beale, J. R., & Wessel, P. (2019). Global Bathymetry and Topography at 15 Arc Sec: SRTM15+. *Earth and Space Science*, *6*(10), 1847–1864. https://doi.org/10.1029/2019EA000658

Turner, K. G., Fréville, H., & Rieseberg, L. H. (2015). Adaptive plasticity and niche expansion in an invasive thistle. *Ecology and Evolution*, *5*(15), 3183–3197. https://doi.org/10.1002/ece3.1599

Vasseur, F., Sartori, K., Baron, E., Fort, F., Kazakou, E., Segrestin, J., Garnier, E., Vile, D., & Violle, C. (2018). Climate as a driver of adaptive variations in ecological strategies in Arabidopsis thaliana. *Annals of Botany*, *122*(6), 935–945. https://doi.org/10.1093/aob/mcy165

Warner, R. M., & Erwin, J. E. (2005). Naturally occurring variation in high temperature induced floral bud abortion across Arabidopsis thaliana accessions. *Plant, Cell & Environment*, *28*(10), 1255–1266. https://doi.org/10.1111/j.1365-3040.2005.01361.x

Whitlock, M. C., & Guillaume, F. (2009). Testing for Spatially Divergent Selection: Comparing QST to FST. *Genetics*, *183*(3), 1055–1063. https://doi.org/10.1534/genetics.108.099812

Wilczek, A. M., Roe, J. L., Knapp, M. C., Cooper, M. D., Lopez-Gallego, C., Martin, L. J., Muir, C. D., Sim, S., Walker, A., Anderson, J., Egan, J. F., Moyers, B. T., Petipas, R., Giakountis, A., Charbit, E., Coupland, G., Welch, S. M., & Schmitt, J. (2009). Effects of Genetic Perturbation on Seasonal Life History Plasticity. *Science*, *323*(5916), 930–934.

Witsen, D. (2012). *Djibouti Biodiversity and Tropical Forests: 118/119 Assessment* (p. 56). United States Agency for International Development. https://usaidgems.org/Documents/FAA&Regs/FAA118119/Djibouti2012.pdf

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Zeng, L., Gu, Z., Xu, M., Zhao, N., Zhu, W., Yonezawa, T., Liu, T., Qiong, L., Tersing, T., Xu, L., Zhang, Y., Xu, R., Sun, N., Huang, Y., Lei, J., Zhang, L., Xie, F., Zhang, F., Gu, H., … Zhong, Y. (2017). Discovery of A high-altitude ecotype and ancient lineage of Arabidopsis thaliana from Tibet. *Science Bulletin*, *62*(24), 1628–1630. https://doi.org/10.1016/j.scib.2017.10.007

Zou, Y.-P., Hou, X.-H., Wu, Q., Chen, J.-F., Li, Z.-W., Han, T.-S., Niu, X.-M., Yang, L., Xu, Y.-C., Zhang, J., Zhang, F.-M., Tan, D., Tian, Z., Gu, H., & Guo, Y.-L. (2017). Adaptation of Arabidopsis thaliana to the Yangtze River basin. *Genome Biology*, *18*(1), 239. https://doi.org/10.1186/s13059-017-1378-9

Supplemental material

Methods

When calculating anomalies for bioclimatic variables, we used months we deemed most relevant to the herbarium specimen based on its date of collection. Specifically, for the minimum temperature of the coldest month, precipitation of the wettest and driest quarters, and precipitation variability, we used the 12 months preceding collection because we reasoned it would best capture these potential stressors. For the mean temperature of the warmest and wettest quarter we used the same calendar year of the collection, because we reasoned this would allow us to capture growing season conditions for the most relevant period for an individual. For temperature annual range and isothermality we used the fall (October) preceding collecting to the summer in the year of collection (September) to attempt to capture both the growing season of collection and its prior cool season.

To test if inflorescence height and maximum rosette length were good proxies for total number of fruits, we counted flowers and fruits using ImageJ on 311 broadly distributed specimens (Figure S1) from the *Muséum national d'histoire naturelle* (herbarium code P), which provided high-resolution scans, allowing us to resolve individual flowers and fruits. We found that inflorescence height and maximum rosette leaf length were positively correlated with total number of flowers and fruits (inflorescence height $r = 0.46$, $n = 304$; maximum rosette leaf length $r = 0.50$, $n = 183$).

When there were multiple plants on an herbarium sheet, we selected one based on tissue sampling for a previous study (DeLeo et al., 2020) or we selected one at random using a random number generator when there were fewer than 6 plants present. We did not measure inflorescence length on specimens where the main inflorescence was obviously bent or broken. Likewise we only measured maximum rosette leaf length for specimens where rosettes were preserved flat (as opposed to crumpled, folded, or broken).

For all GAMs, relationships were predicted over a 140 by 200 grid from -28.5 to 140.5° longitude and 23.9 to 71.0° latitude.

Figure S1. 311 specimens from the P herbarium (MNHN, Paris) for which we counted individual fruits and flowers to compare with inflorescence height and maximum rosette leaf length.

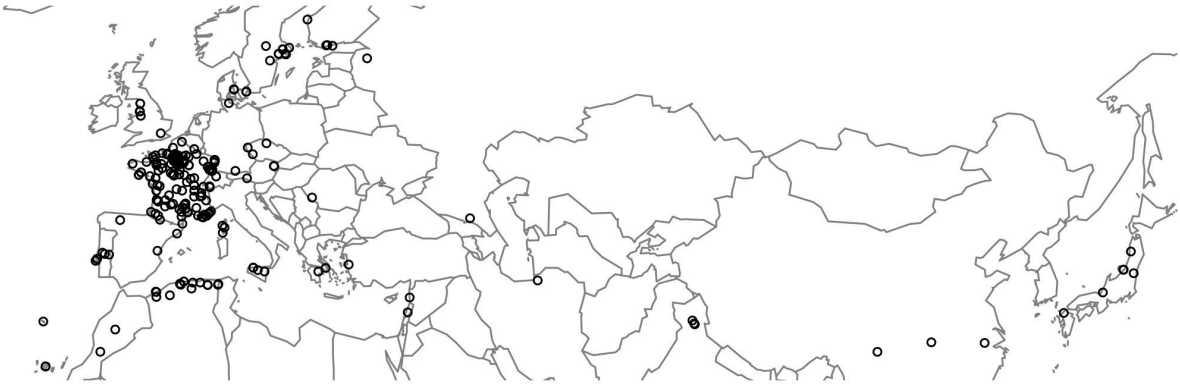


Figure S2. MESS maps of current, LGM, and future climate conditions. Blue indicates positive similarity and red indicates negative similarity with conditions used for fitting the model on present occurrences.

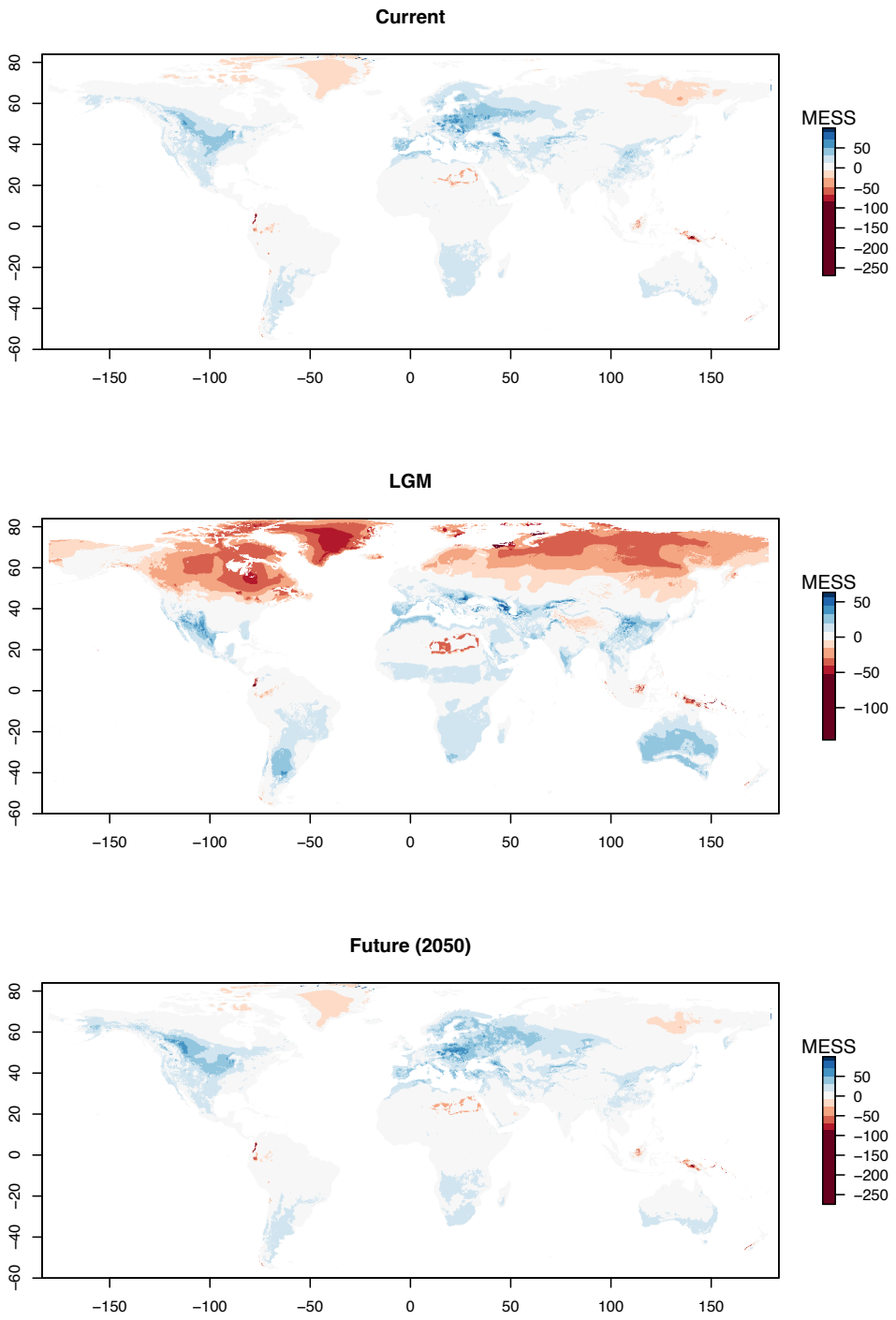


Figure S3. Spatial varying coefficients and intercepts from GAMs showing the relationship between inflorescence height and yearly anomalies in climate variables of isothermality (A, BIO3), mean daily minimum air temperature of the coldest month (B, BIO6), annual range of air temperature (C, BIO7), and mean daily mean air temperature of the wettest quarter (D, BIO8). Gray shading indicates areas where parameter values are not statistically significantly different from zero (e.g. western Europe in for the BIO3 effect in panel A), specifically where 95% confidence intervals overlap with 0.

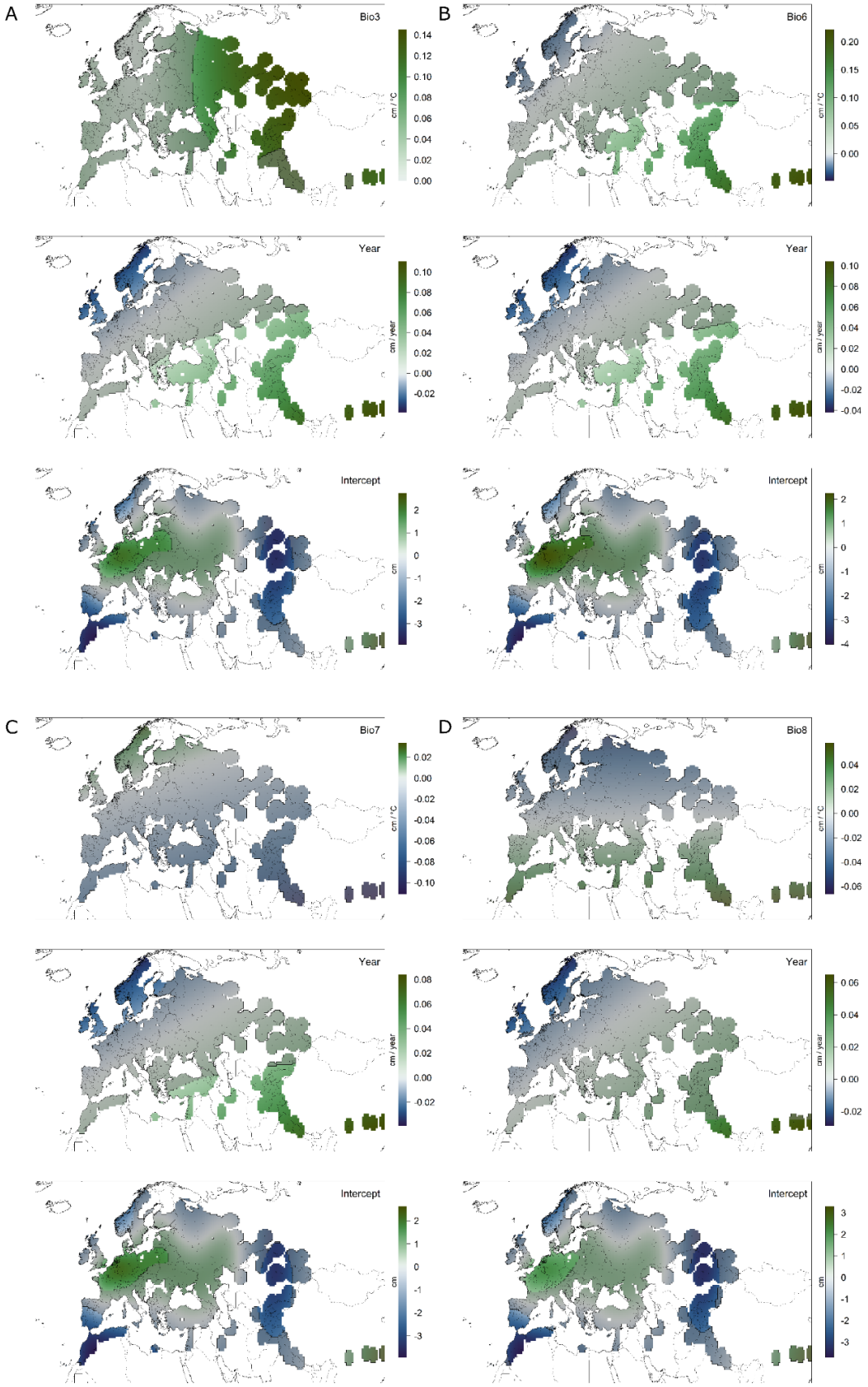

A
Bio3
cm / °C
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
B
Bio6
cm / °C
0.20
0.15
0.10
0.05
0.00
Year
cm / year
0.10
0.08
0.06
0.04
0.02
0.00
-0.02
Year
cm / year
0.10
0.08
0.06
0.04
0.02
0.00
-0.02
-0.04
Intercept
cm
2
1
0
-1
-2
-3
Intercept
cm
2
1
0
-1
-2
-3
-4
C
Bio7
cm / °C
0.02
0.00
-0.02
-0.04
-0.06
-0.08
-0.10
D
Bio8
cm / °C
0.04
0.02
0.00
-0.02
-0.04
-0.06
Year
cm / year
0.08
0.06
0.04
0.02
0.00
-0.02
Year
cm / year
0.06
0.04
0.02
0.00
-0.02
Intercept
cm
2
1
0
-1
-2
-3
Intercept
cm
3
2
1
0
-1
-2
-3

Figure S3 (cont.). Spatial varying coefficients and intercepts from GAMs showing the relationship between inflorescence height and yearly anomalies in climate variables mean daily mean air temperatures of the warmest quarter (E, BIO10), precipitation seasonality (F, BIO15), mean monthly precipitation amount of the wettest quarter (G, BIO16), and mean monthly precipitation amount of the driest quarter (H, BIO17). Gray shading indicates areas where 95% confidence intervals overlap with 0.

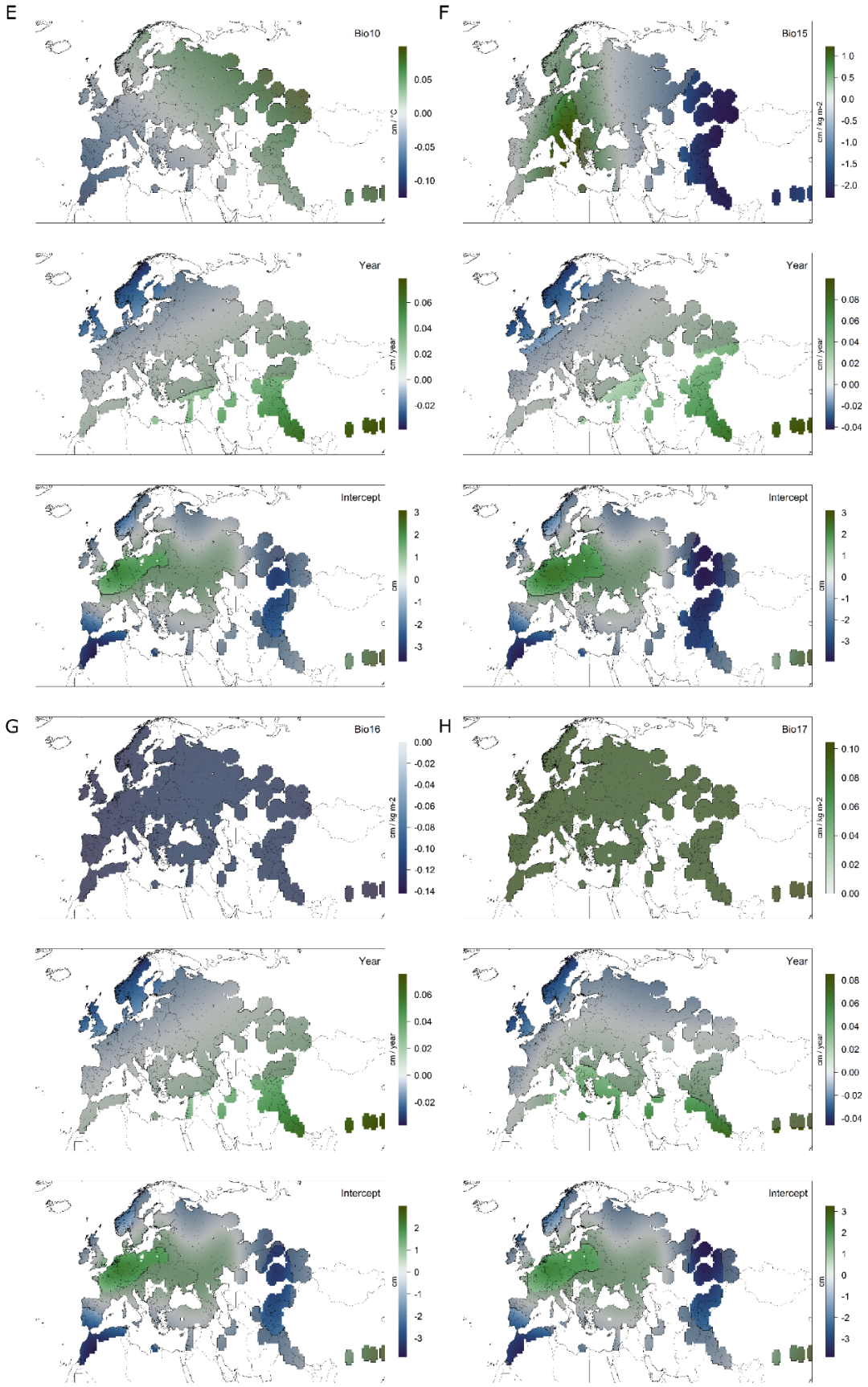

E
Bio10
cm / °C
0.05
0.00
-0.05
-0.10
Year
cm / year
0.06
0.04
0.02
0.00
-0.02
Intercept
cm
3
2
1
0
-1
-2
-3
F
Bio15
cm / kg m-2
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
Year
cm / year
0.08
0.06
0.04
0.02
0.00
-0.02
-0.04
Intercept
cm
3
2
1
0
-1
-2
-3
G
Bio16
cm / kg m-2
0.00
-0.02
-0.04
-0.06
-0.08
-0.10
-0.12
-0.14
Year
cm / year
0.06
0.04
0.02
0.00
-0.02
Intercept
cm
2
1
0
-1
-2
-3
H
Bio17
cm / kg m-2
0.10
0.08
0.06
0.04
0.02
0.00
Year
cm / year
0.08
0.06
0.04
0.02
0.00
-0.02
-0.04
Intercept
cm
3
2
1
0
-1
-2
-3

Figure S4. Spatial varying coefficients and intercepts from GAMs showing the relationship between maximum rosette leaf length and yearly anomalies in climate variables of isothermality (A, BIO3), mean daily minimum air temperature of the coldest month (B, BIO6), annual range of air temperature (C, BIO7), and mean daily mean air temperature of the wettest quarter (D, BIO8). Gray shading indicates areas where parameter values are not statistically significantly different from zero (e.g. central Europe in for the intercept parameter in panel A), specifically where 95% confidence intervals overlap with 0.

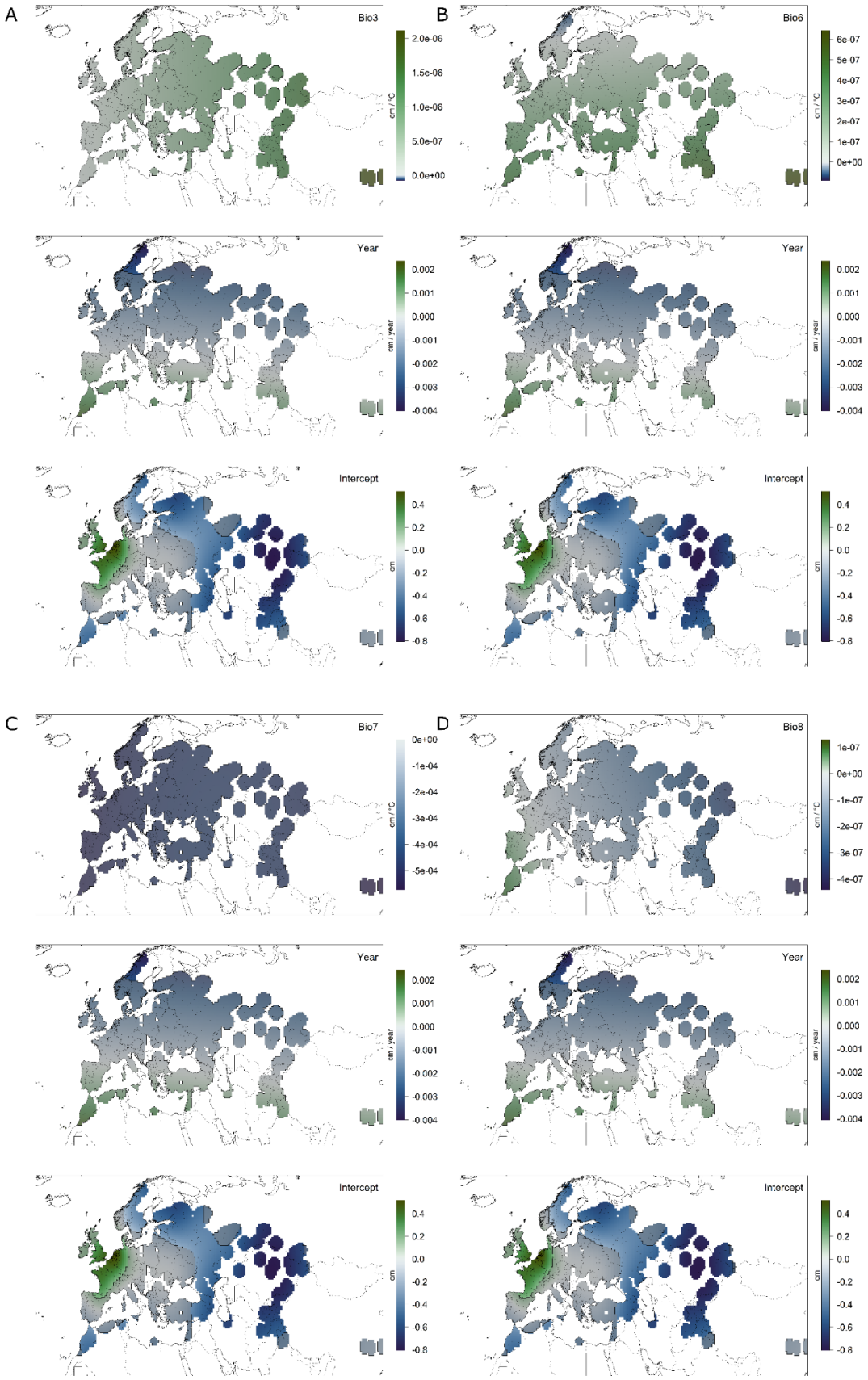

A
Bio3
2.0e-06
1.5e-06
1.0e-06
5.0e-07
0.0e+00
cm / °C
Year
0.002
0.001
0.000
-0.001
-0.002
-0.003
-0.004
cm / year
Intercept
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
cm
B
Bio6
6e-07
5e-07
4e-07
3e-07
2e-07
1e-07
0e+00
cm / °C
Year
0.002
0.001
0.000
-0.001
-0.002
-0.003
-0.004
cm / year
Intercept
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
cm
C
Bio7
0e+00
-1e-04
-2e-04
-3e-04
-4e-04
-5e-04
cm / °C
Year
0.002
0.001
0.000
-0.001
-0.002
-0.003
-0.004
cm / year
Intercept
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
cm
D
Bio8
1e-07
0e+00
-1e-07
-2e-07
-3e-07
-4e-07
cm / °C
Year
0.002
0.001
0.000
-0.001
-0.002
-0.003
-0.004
cm / year
Intercept
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
cm

Figure S4 (cont.). Spatial varying coefficients and intercepts from GAMs showing the relationship between maximum rosette leaf length and yearly anomalies in climate variables mean daily mean air temperatures of the warmest quarter (E, BIO10), precipitation seasonality (F, BIO15), mean monthly precipitation amount of the wettest quarter (G, BIO16), and mean monthly precipitation amount of the driest quarter (H, BIO17). Gray shading indicates areas where 95% confidence intervals overlap with 0.

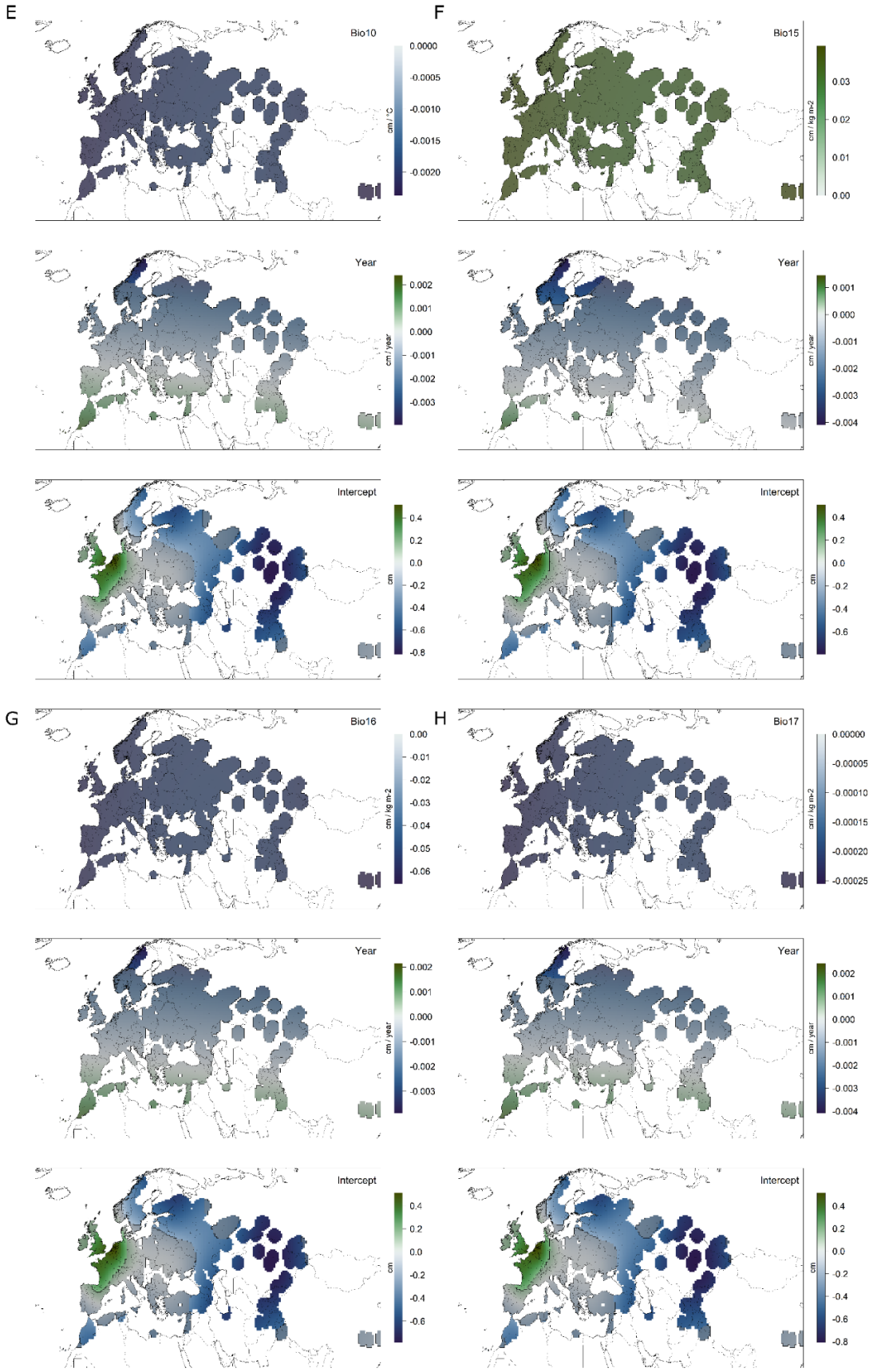

E
Bio10
0.0000
-0.0005
-0.0010
-0.0015
-0.0020
cm / °C
Year
0.002
0.001
0.000
-0.001
-0.002
-0.003
cm / year
Intercept
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
cm
F
Bio15
0.03
0.02
0.01
0.00
cm / kg m-2
Year
0.001
0.000
-0.001
-0.002
-0.003
-0.004
cm / year
Intercept
0.4
0.2
0.0
-0.2
-0.4
-0.6
cm
G
Bio16
0.00
-0.01
-0.02
-0.03
-0.04
-0.05
-0.06
cm / kg m-2
Year
0.002
0.001
0.000
-0.001
-0.002
-0.003
cm / year
Intercept
0.4
0.2
0.0
-0.2
-0.4
-0.6
cm
H
Bio17
0.00000
-0.00005
-0.00010
-0.00015
-0.00020
-0.00025
cm / kg m-2
Year
0.002
0.001
0.000
-0.001
-0.002
-0.003
-0.004
cm / year
Intercept
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
cm

Figure S5. The relationship between maximum rosette leaf length and habitat suitability is significantly positive across western and central Europe in a GAM with spatially varying coefficients. Color indicates fitted parameter values and gray shading covers not statistically significant areas, i.e. where the 95% CI includes zero.

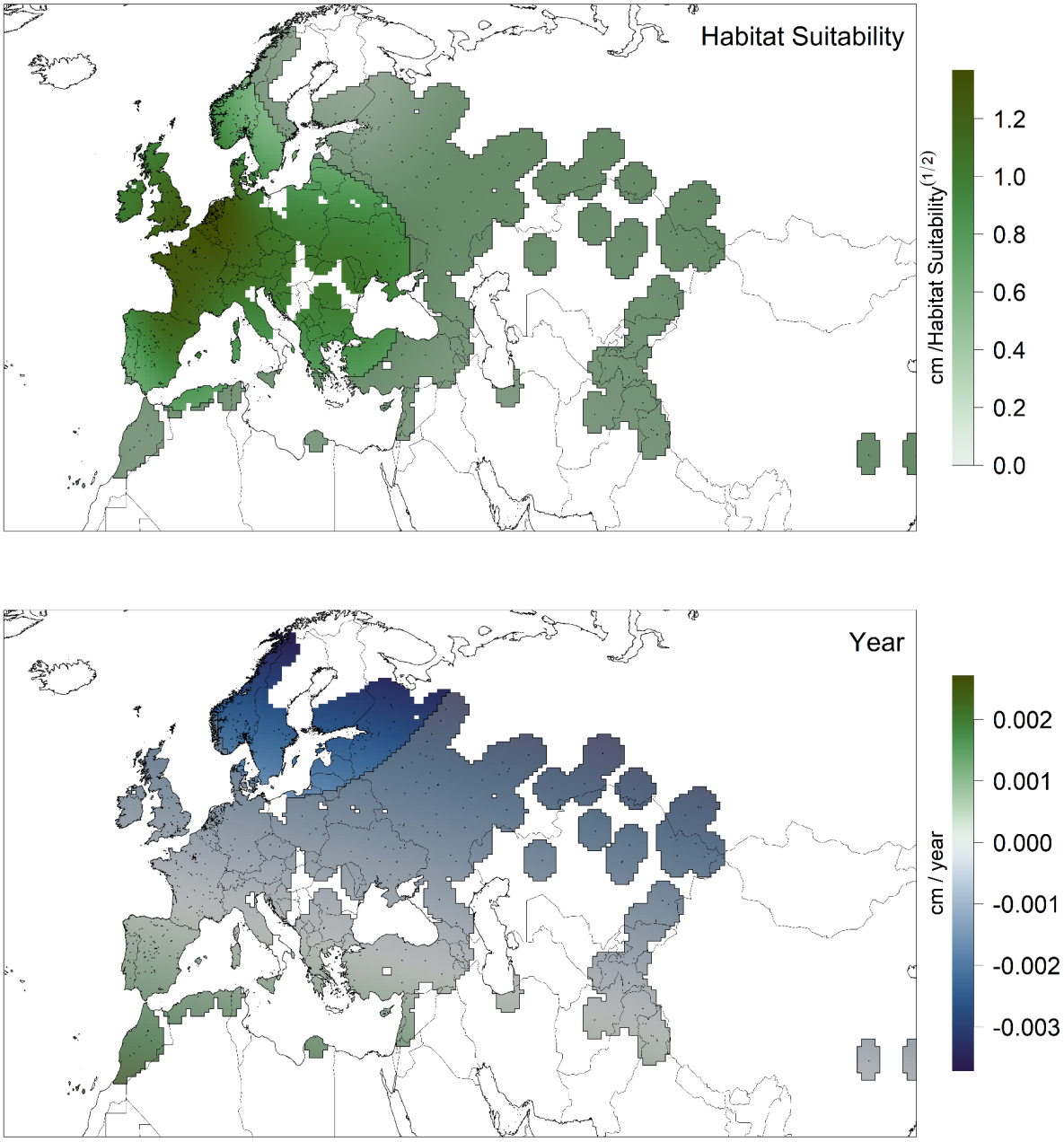


Figure S6. The relationship between inflorescence height and habitat suitability is significantly positive across most of Eurasia in a GAM with spatially varying coefficients. Color indicates fitted parameter values and gray shading covers not statistically significant areas, i.e. where the 95% CI includes zero.

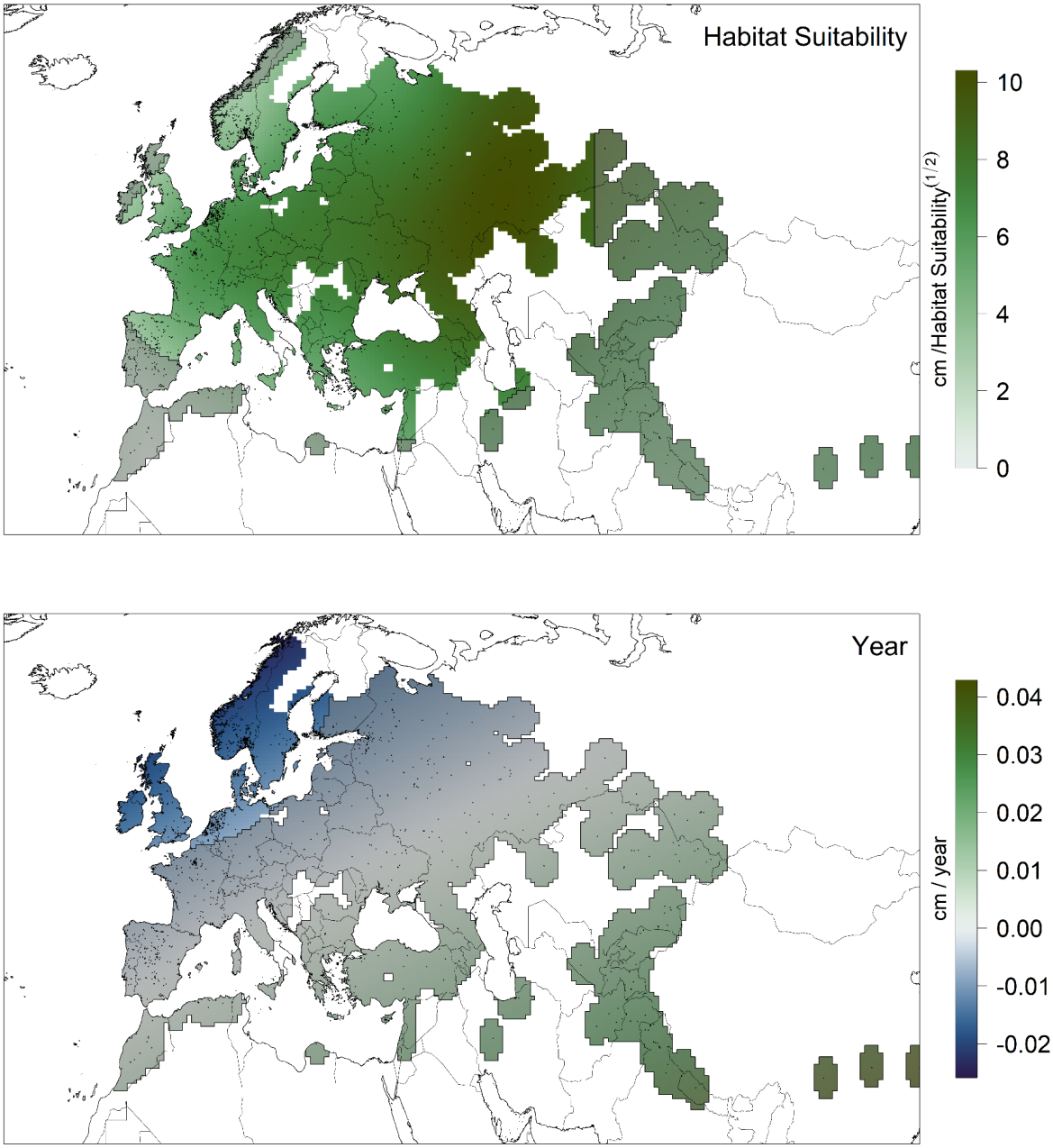


Figure S7. Suitability relationship with genetic variation in flowering time (N=898). Linear regression fits are shown.

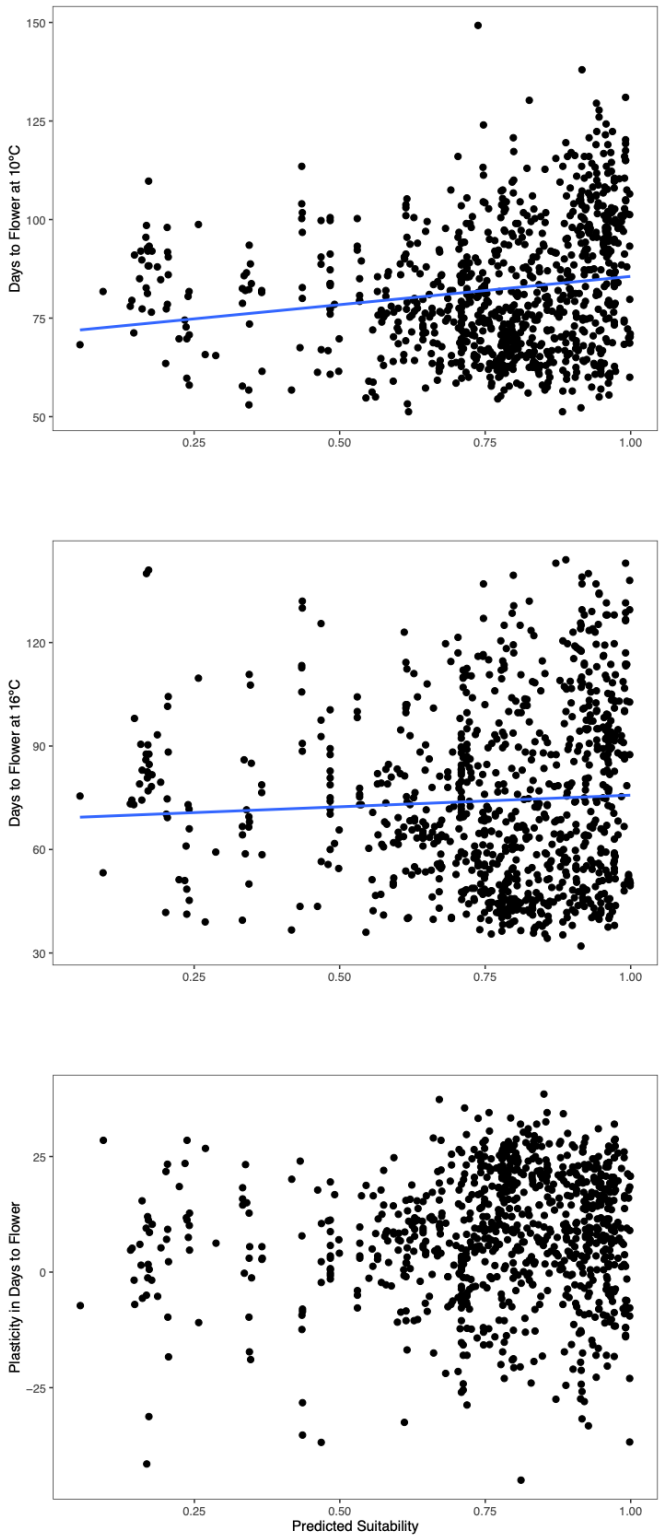


Figure S8. Association between habitat suitability and flowering time at 10°C (top) or plasticity of flowering time (bottom) in a GAM. Flowering time was significantly positively related to habitat suitability for most of the locations where Arabidopsis has been collected, while plasticity in flowering time (difference between flowering time at 10°C vs 16°C) is significantly negatively related to habitat suitability. Gray shading indicates regions where parameters are not significantly different from zero, i.e. where the 95% confidence intervals overlap with 0.

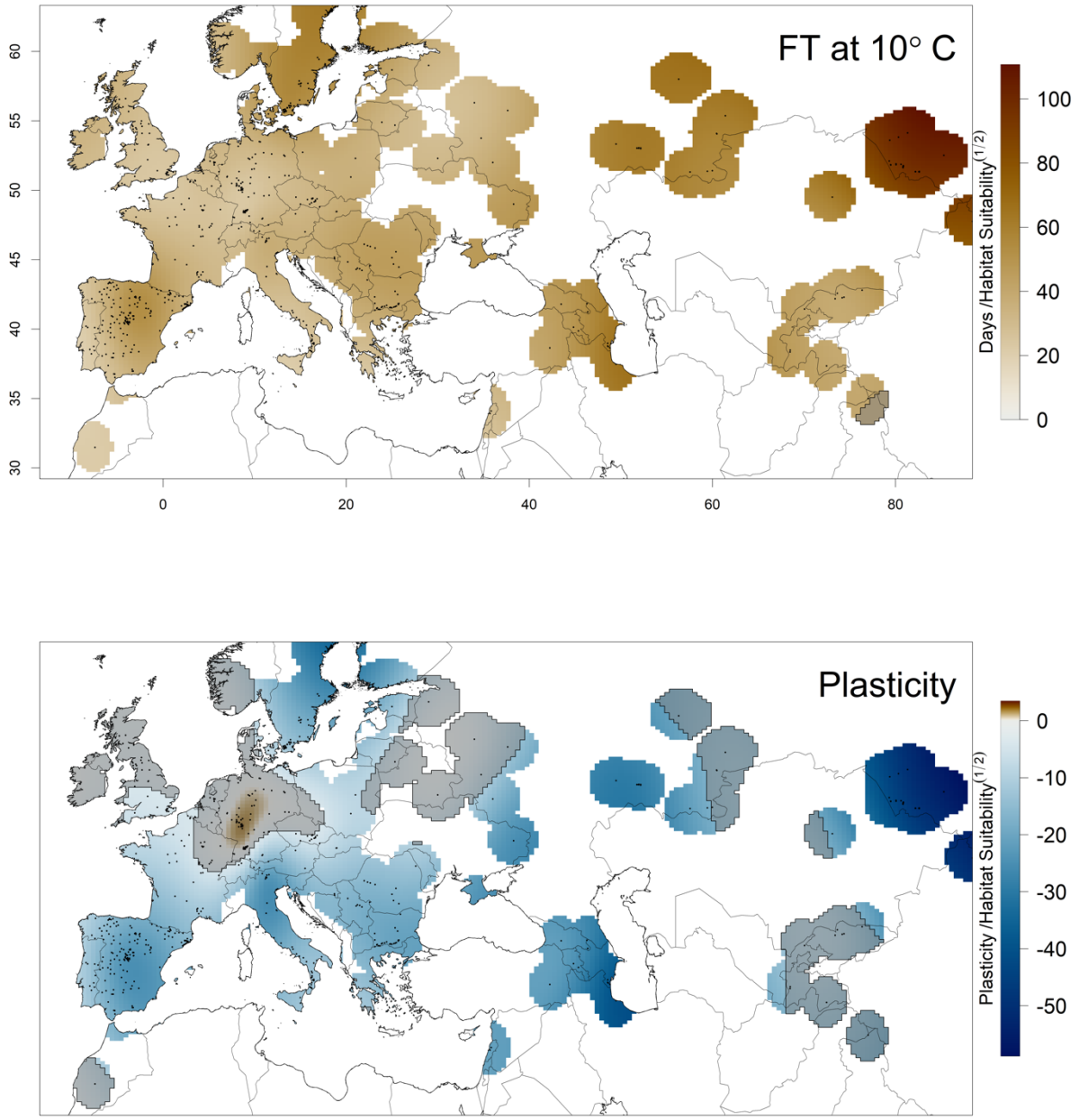


Figure S9. Variation in suitability among genetic clusters. Admixed (N=116 ecotypes), Asia (67), Central Europe (173), Germany (58), Italy-Balkan-Caucasus (92), North Sweden (41), South Sweden (129), Relict (23), Spain (103), and Western Europe (96).

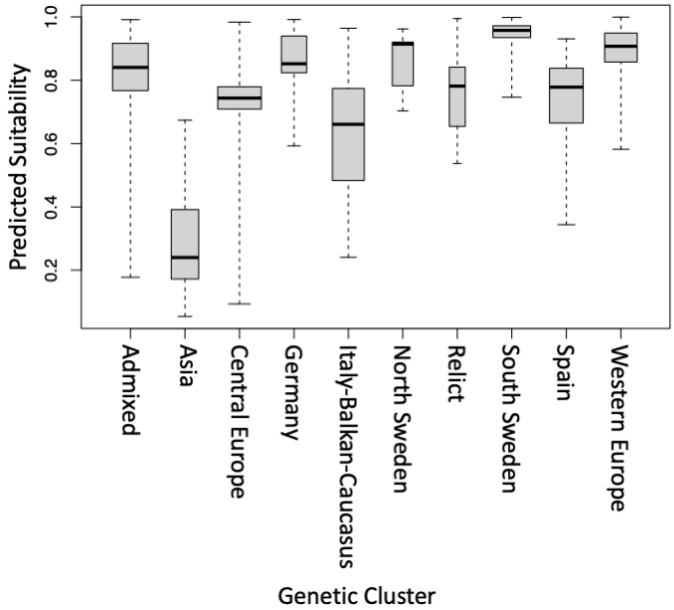


Figure S10. Limiting factors map for globe.

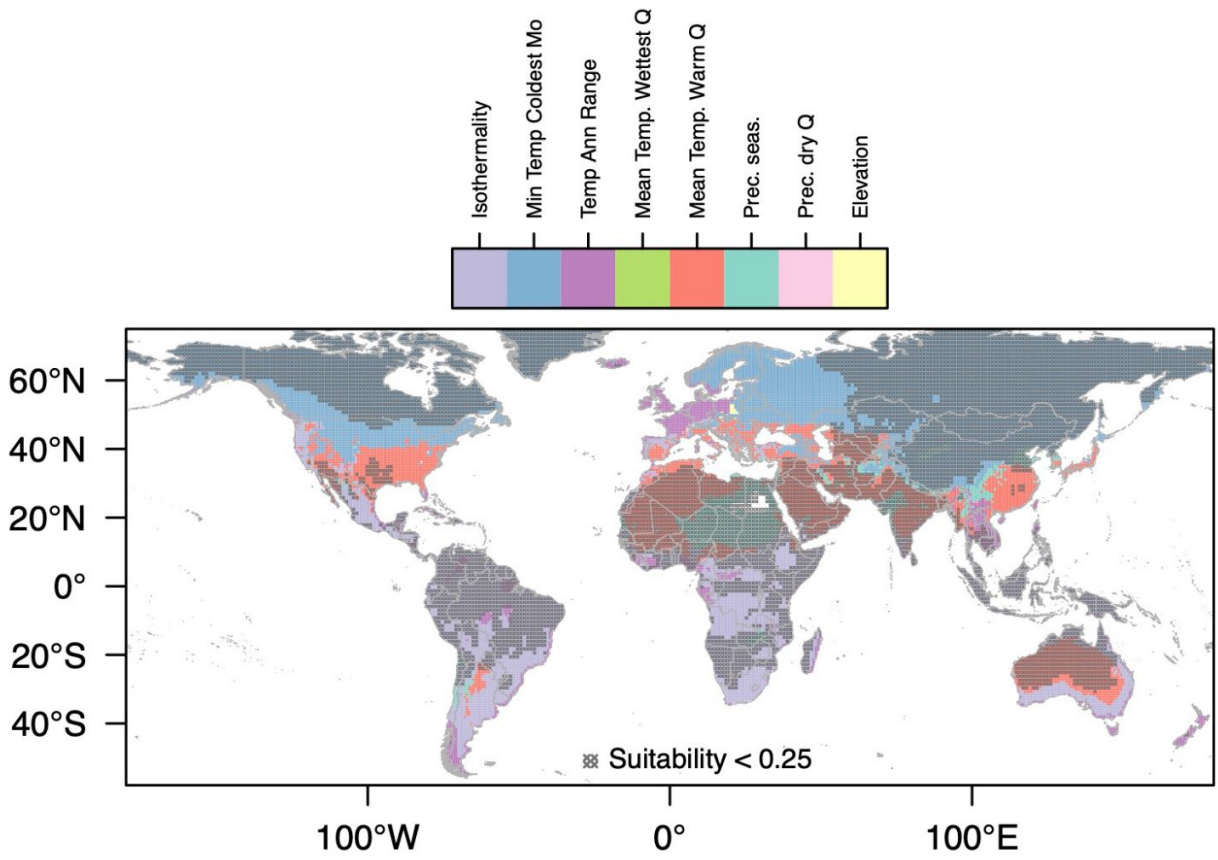


Table S1. Permutation importance of tested variables in the fitted habitat suitability model.

Environmental Variable	Permutation Importance (%)
Min Temperature of Coldest Month	37.4693
Mean Temperature of Warmest Quarter	33.7603
Isothermality	12.7888
Precipitation Seasonality	6.6833
Temperature Annual Range	3.6676
Elevation	2.8841
Mean Temperature of Wettest Quarter	2.0659
Precipitation of Driest Quarter	0.6369
Precipitation of Wettest Quarter	0.0439

Table S2. Results from a linear-mixed effect model testing for an association of days to flower at 10ºC and predicted suitability while accounting for kinship among ecotypes using lmekin in R.

Linear mixed-effects kinship model fit by maximum likelihood

Data: ft.cluster

Log-likelihood =-3402.342

n=898

Model: FT10 ~ preds + (1 | id)

Fixed coefficients

	Value	Std Error	z	p
(Intercept)	73.87	20.17	3.66	0.00025
preds	14.24	3.74	3.81	0.00014

Random effects

Group	Variable	Std Dev	Variance
id	Vmat.1	2.63e+07	6.91E+14

Residual error= 3.61e-07

Table S3. Results from a linear-mixed effect model testing for an association of days to flower at 16ºC and predicted suitability while accounting for kinship among ecotypes using lmekin in R.

Linear mixed-effects kinship model fit by maximum likelihood

Data: ft.cluster

Log-likelihood = -3781.65

n=898

Model: FT16 ~ preds + (1 | id)

Fixed coefficients

	Value	Std Error	z	p
(Intercept)	64.64	30.77	2.10	0.036
preds	17.95	5.71	3.15	0.0017

Random effects

Group	Variable	Std Dev	Variance
id	Vmat.1	5.67E+08	3.22E+17

Residual error= 5.51e-07

Table S4. Results from a linear-mixed effect model testing for an association of plasticity in days to flower and predicted suitability while accounting for kinship among ecotypes using lmekin in R.

Linear mixed-effects kinship model fit by maximum likelihood

Data: ft.cluster

Log-likelihood = -3403.9

n=898

Model: FT_p ~ preds + (1 | id)

Fixed coefficients

	Value	Std Error	z	p
(Intercept)	7.64	15.96	0.48	0.63
preds	-1.6	3.67	-0.44	0.66

Random effects

Group	Variable	Std Dev	Variance
id	Vmat.1	18.61	346.3

Residual error= 6.2

Table S5. Tukey HSD post-hoc tests for genetic clusters' predicted suitability. Genetic clusters are: Admixed (N=116 ecotypes), Asia (67), Central Europe (173), Germany (58), Italy-Balkan-Caucasus (92), North Sweden (41), South Sweden (129), Relict (23), Spain (103), and Western Europe (96).

	diff	lwr	upr	p adj
asia-admixed	-0.4979256	-0.5526559	-0.4431953	0
central_europe-admixed	-0.0778336	-0.1206358	-0.0350315	0.0000005
germany-admixed	0.05570172	-0.0016571	0.11306056	0.065364
italy_balkan_caucasus-admixed	-0.1525927	-0.2023868	-0.1027987	0
north_sweden-admixed	0.03908031	-0.0257231	0.1038837	0.6603736
relict-admixed	-0.0563982	-0.1378092	0.02501283	0.4593667
south_sweden-admixed	0.13120376	0.08556561	0.1768419	0
spain-admixed	-0.0606979	-0.1089863	-0.0124094	0.002896
western_europe-admixed	0.08450361	0.03529149	0.13371573	0.000003
central_europe-asia	0.42009195	0.36876872	0.47141517	0
germany-asia	0.55362729	0.48965788	0.61759671	0
italy_balkan_caucasus-asia	0.34533287	0.28804851	0.40261722	0
north_sweden-asia	0.53700589	0.46628442	0.60772737	0
relict-asia	0.44152738	0.35533106	0.52772369	0
south_sweden-asia	0.62912934	0.57541817	0.6828405	0
spain-asia	0.43722772	0.38124716	0.49320828	0
western_europe-asia	0.58242919	0.52564995	0.63920843	0
germany-central_europe	0.13353535	0.07941785	0.18765284	0
italy_balkan_caucasus-central_europe	-0.0747591	-0.1207821	-0.0287361	0.000014
north_sweden-central_europe	0.11691394	0.05496118	0.17886671	0.0000001
relict-central_europe	0.02143543	-0.0577253	0.10059617	0.9975353
south_sweden-central_europe	0.20903739	0.16754634	0.25052843	0
spain-central_europe	0.01713577	-0.0272539	0.06152541	0.9684316
western_europe-central_europe	0.16233724	0.11694452	0.20772996	0

italy_balkan_caucasus-germany	-0.2082944	-0.2680952	-0.1484937	0
north_sweden-germany	-0.0166214	-0.0893961	0.0561533	0.999359
relict-germany	-0.1120999	-0.1999887	-0.0242112	0.0022898
south_sweden-germany	0.07550204	0.01911482	0.13188927	0.0009934
spain-germany	-0.1163996	-0.1749526	-0.0578466	0
western_europe-germany	0.02880189	-0.0305152	0.08811894	0.8756712
north_sweden-italy_balkan_caucasus	0.19167303	0.12469864	0.25864741	0
relict-italy_balkan_caucasus	0.09619451	0.01304497	0.17934405	0.0096571
south_sweden-italy_balkan_caucasus	0.28379647	0.23512481	0.33246813	0
spain-italy_balkan_caucasus	0.09189485	0.04072978	0.14305992	0.0000007
western_europe-italy_balkan_caucasus	0.23709632	0.18505861	0.28913403	0
relict-north_sweden	-0.0954785	-0.1883972	-0.0025599	0.0383232
south_sweden-north_sweden	0.09212344	0.02817846	0.15606842	0.0002389
spain-north_sweden	-0.0997782	-0.1656409	-0.0339155	0.0000789
western_europe-north_sweden	0.0454233	-0.0211196	0.11196616	0.4818301
south_sweden-relict	0.18760196	0.10687255	0.26833137	0
spain-relict	-0.0042997	-0.0865564	0.07795709	1
western_europe-relict	0.14090181	0.05809945	0.22370417	0.0000039
spain-south_sweden	-0.1919016	-0.2390318	-0.1447714	0
western_europe-south_sweden	-0.0467001	-0.0947763	0.00137599	0.0652042
western_europe-spain	0.14520147	0.09460257	0.19580037	0