INTRODUCTION

Adaptation of natural populations to climatic conditions has been commonly documented within very diverse taxa (Fournier-Level et al., 2011; Hancock et al., 2011; Keller & Seehausen, 2012; Savolainen, Lascoux, & Merilä, 2013; Turner, Levine, Eckert, & Begun, 2008). However, climate is also a key factor in determining species range limits and hence it is an important aspect of the environment of organisms to which adaptation seems at some point constrained (Sexton, McIntyre, Angert, & Rice, 2009). Progress in understanding climate adaptation and its limits involves the study of the most relevant climatic parameters restricting the occurrence of a species and the genes and traits underlying adaptation to these climatic parameters across a species range. Also indepth analyses on how common parallel evolution is between different lineages are needed. This requires determining the genetic level at which parallelism occurs, i.e., whether the same genomic variants, or different variants of the same genes, or functionally related genes are involved. This can
indicate whether several routes of genomic change can lead to adaptation to similar climatic conditions or whether few options exist. Further analyses should target specific aspects of constraints, such as standing genetic variation, and the interference of linkage and demographic history.

If two independent lineages adapt to similar environments using the same genetic variants or genes, this is termed parallel evolution (Arendt & Reznick, 2008). The process is also called convergent evolution if the same genetic variants or genes are involved in adaptation but species and lineages are highly unrelated. In principle, parallel evolution on the level of the phenotype and genotype can be due to new mutations or selection on standing genetic variation (Chan et al., 2010), while introgression of adaptive variants from another lineage only mimics parallel adaptation. Whether or not evolution is repeatable has implications for adaptability under changing conditions, and knowledge of the genomic basis of this adaptation allows for predictions into the future. This has led to increasing interest in parallel evolution over the past years, and the development of new methods to quantify such parallelism (Yeaman, Gerstein, Hodgins, & Whitlock, 2018). Additionally, parallel detection of outliers in independent lineages strongly indicates adaptive value of the respective variants (Barrett & Hoekstra, 2011).

Theory considering only new mutations predicts that the chance of parallel evolution depends on the number of potential beneficial mutations (Orr, 2005). The chance of parallel evolution from standing genetic variation depends on the initial allele frequency, effect size and population history (MacPherson & Nuismer, 2017). The higher the initial allele frequency is, the greater the chance for parallel fixation of alleles experiencing similar selection. Likewise, the higher the effect size of the allele in question, the more likely the allele will become fixed in both lineages independently under similar selection regimes (MacPherson & Nuismer, 2017). Population demographic history also plays an important role because the larger the effective population size is, the more likely parallelism becomes, as random genetic drift is less likely to drive an allele to extinction (MacPherson & Nuismer, 2017). Finally, strong selection may also increase the chance of parallel adaptation if it acts in the same direction (MacPherson & Nuismer, 2017). Linked selection may further lead to the coupling of coadapted loci, increasing the regions under selection (Bierne, Welch, Loire, Bonhomme, & David, 2011; Butlin & Smadja, 2018). Adaptation from standing genetic variation has the advantage that adaptive alleles are already present in the populations, and adaptation can occur much more rapidly (Barrett & Schluter, 2008). In summary, parallelism on the level of genetic variants is expected when the number of variants increasing fitness is low, their frequency or effect size is large, selection is strong, and population history allows for maintenance of these variants. In contrast, its lack is expected when many variants carry small fitness benefits or when genetic diversity is lost through genetic drift.

Molecular studies have shown that parallel adaptation to environmental conditions is indeed not rare in nature. Typically, parallelism is assessed based on genomic data, such as through genome-wide association studies and the subsequent study of patterns of selection on the nucleotide level or in candidate genes associated with environmental factors being compared between different evolutionary lineages. Parallel adaptation from new mutations is most commonly found in experimental evolution in bacteria or viruses, and while some studies have found extensive parallelism at the nucleotide level (Wichman, 1999), parallelism at the level of genes seems much more common even in these organisms with small genomes (Bailey, Rodrigue, & Kassen, 2015; Barrick et al., 2009; Woods, Schneider, Winkworth, Riley, & Lenski, 2006). Also, selection from standing genetic variation is often detected as the source of rapid adaptation to environmental change, such as in resistance to drought in Arabidopsis thaliana (Exposito-Alonso et al., 2018). Within the same species, adaptation to the same environment might even be based on new mutations as well as on standing genetic variation, such as in the threespine stickleback, in which adaptation to freshwater has been facilitated by parallel adaptation from independent mutations (Chan et al., 2010) as well as from standing genetic variation (Colosimo et al., 2005; Jones et al., 2012). Adaptation to climatic conditions is probably complex because of the potentially diverse interplay of different climatic variables such as temperature, precipitation or seasonality across the distribution range. Nevertheless, parallelism in climate adaptation (on the level of genes) has been shown for species as distantly related as pine and spruce. Although these species diverged over 100 million years ago, patterns of convergence were identified in 47 genes, or between 10% and 18% of top candidate genes for adaptation to temperature and cold hardness (Yeaman et al., 2016). In contrast, within black spruce, adaptation to climatic conditions in Pleistocene lineages was found to be often lineage-specific (Prunier, Gérardi, Laroche, Beaulieu, & Bousquet, 2012). These contrasting results stress the need to better understand the conditions that lead to parallelism or its lack on the different levels of organismal organization.

Here, we studied parallelism in climate adaptation in the short-lived perennial plant Arabidopsis lyrata subsp. lyrata (A. lyrata from here on) of eastern North America. It is a close relative of the model plant A. thaliana (Hohmann et al., 2014; Novikova et al., 2016) and has become a model species itself (Bombíes & Weigel, 2007). Previous research identified two distinct genetic clusters in A. lyrata by microsatellite analysis and plastid DNA (Foxe et al., 2010; Griffin & Willi, 2014; Willi & Määttänen, 2010), which have been confirmed by pooled whole-genome sequence data more recently (Willi, Fracassetti, Zoller, & Van Buskirk, 2018). One cluster is located in the Midwest from Missouri to the Great Lakes region (subsequently called western cluster), and the other in the east, mainly the Appalachians from North Carolina to New York State (eastern cluster). Relatedness trees suggested that they each had their own refuge area during the last glaciation cycle close to the line of the maximum extent of the ice sheet and underwent post-glacial range expansion northwards but also eastwards, resulting in an admixture zone at Lake Erie (Griffin & Willi, 2014; Willi et al., 2018). Due to this range expansion, both ancestral clusters nowadays occur over fairly wide latitudinal and longitudinal ranges, with very diverse climatic conditions within ranges.
The ecological niche of *A. lyrata* across its distribution range has been studied recently, and the most important factors for determining the niche have been identified (Lee-Yaw, Fracassetti, & Willi, 2018). These were the average minimum temperature in early spring (March and April), precipitation of the wettest quarter, and moisture availability (Priestley–Taylor coefficient). Minimum temperature contributed most to the niche, while the other two variables had moderate importance.

In this study, we used sequencing data covering the complete genome of *A. lyrata* and the complete distribution range to address the following questions: (a) Which genes and gene ontology (GO) terms are involved in adaptation to niche-determining climatic factors? (b) Did the two distinct lineages adapt to parallel climatic conditions in parallel? If so, are the same genes or even the same SNPs involved in adaptation to the environment in the two postglacial lineages of *A. lyrata*, or is parallelism restricted to GO terms? (c) Could standing genetic variation, or the lack thereof, account for the observed patterns of (non-)parallelism?

## 2 | MATERIALS AND METHODS

### 2.1 | Plant material and sampling strategy

We investigated parallel adaptation to climate in North American *Arabidopsis lyrata* subsp. *lyrata* (L.) O’Kane & Al-Shehbaz using pooled whole-genome sequences (pool-seq) of a subset of populations from a previously published study (Willi et al., 2018; GenBank/ENA project PRJEB19338). Plant material from 41 out of the 55 populations was used (GenBank/ENA accessions NC_000932 and NC_001284, respectively). For each population, all reads sequenced on different lanes were then merged, and only those mapping to scaffolds 1–8 were retained. Duplicate reads were removed using Picard 2.1.1 (http://broadinstitute.github.io/picard/), mappings were filtered to retain only properly mapped reads with mapping quality >20, and finally a single mpileup file was generated in SAMtools 1.3.1 (Li et al., 2009).

Variant calling was performed in VarScan 2.4.1 (Koboldt et al., 2009) on all populations together using the mpileup2snp command. On a population basis, only sites with depth ≥50 were considered, with ≥3 reads per allele and variant frequency ≥0.03 required to make a variant call. Indels were identified using Vcftools 0.1.14 (Danecek et al., 2011), and nucleotides were filtered for the following additional parameters: maximum read depth of 500 and minimum genotype quality of 20. Only biallelic sites were considered; indels and adjacent regions (5 bp in both directions) were removed. Additionally, sites in previously identified repetitive regions (Fracassetti et al., 2015) were removed using BEDTools 2.26.0 (Quinlan & Hall, 2010). The two sets of populations were then selected. Strand bias for both reference and alternative allele was removed in R 3.4.1 (R Core Team, 2016) on a population basis. Sites with a strand bias >90% were set to missing for each population, and subsequently the data sets were filtered again for sites with ≥50% missing data. Finally, our filtering resulted in a total number of 1,546,883 (east) and 1,371,857 (west) sites. A total of 750,998 sites were shared between clusters (48.55% of the eastern and 54.74% of the western cluster).

### 2.2 | Processing of pooled sequence data

We used a similar bioinformatics pipeline as described by Lucek, Hohmann, and Willi (2019). Trimming was performed for each library and lane separately using the Perl script trim-fastq.pl from the software package PoPOOLATION (Kofler et al., 2011) with a minimum base quality of 28 and a minimum length of 84 bp. All reads were mapped to a reference in BWA-MEM 0.7.13 (Li, 2013). Reference was the genome sequence of *A. lyrata* (Hu et al., 2011) with the updated annotation (Rawat et al., 2015) and the following modifications: Positions 8,746,475–8,835,273 and 9,128,838–9,212,301 on scaffold_2 were masked due to their high similarity with the plastid genome sequence, which may have been caused by erroneous assembly of the reference genome. Additionally, plastid and mitochondrial genome sequences of *A. thaliana* were included (GenBank/ENA accessions NC_000932 and NC_001284, respectively). For each population, all reads sequenced on different lanes were then merged, and only those mapping to scaffolds 1–8 were retained. Duplicate reads were removed using Picard 2.1.1 (http://broadinstitute.github.io/picard/), mappings were filtered to retain only properly mapped reads with mapping quality >20, and finally a single mpileup file was generated in SAMtools 1.3.1 (Li et al., 2009).

Variant calling was performed in VarScan 2.4.1 (Koboldt et al., 2009) on all populations together using the mpileup2snp command. On a population basis, only sites with depth ≥50 were considered, with ≥3 reads per allele and variant frequency ≥0.03 required to make a variant call. Indels were identified using Vcftools 0.1.14 (Danecek et al., 2011), and nucleotides were filtered for the following additional parameters: maximum read depth of 500 and minimum genotype quality of 20. Only biallelic sites were considered; indels and adjacent regions (5 bp in both directions) were removed. Additionally, sites in previously identified repetitive regions (Fracassetti et al., 2015) were removed using BEDTools 2.26.0 (Quinlan & Hall, 2010). The two sets of populations were then selected. Strand bias for both reference and alternative allele was removed in R 3.4.1 (R Core Team, 2016) on a population basis. Sites with a strand bias >90% were set to missing for each population, and subsequently the data sets were filtered again for sites with ≥50% missing data. Finally, our filtering resulted in a total number of 1,546,883 (east) and 1,371,857 (west) sites. A total of 750,998 sites were shared between clusters (48.55% of the eastern and 54.74% of the western cluster).

### 2.3 | Selection of climate variables and environmental association analysis

Eight environmental variables linked to climate, topography and vegetation had previously been used to model the ecological niche of *A. lyrata* in North America (Lee-Yaw et al., 2018). Of these, three climatic variables were identified to play a major role in determining the niche, i.e., showing a high percentage contribution to the model and/or high permutation importance in niche modeling (Lee-Yaw et al., 2018). We used these three variables in our genome-wide association study (GWAS), namely minimum temperature in early spring, i.e., the months of March and April \( T_{\text{min}} \)
precipitation of the wettest quarter ($P_{\text{wettest quarter}}$), and the Priestley–Taylor coefficient (Moisture). The latter is the ratio of annual actual evapotranspiration over annual potential evapotranspiration, a measure of moisture availability. We extracted values for $T_{\text{min spring}}$ and $P_{\text{wettest quarter}}$ from the WorldClim database version 1.4 (http://worldclim.org/; Hijmans, Cameron, Parra, Jones, and Jarvis (2005); containing data for 1960–1990) and for Moisture from the Consortium for Spatial Information (http://www.cgis.org/data; Trabucco and Zomer (2010); data for 1950–2000). Climatic variables are provided in Table S1 in Appendix S1.

Association of the three selected variables with the SNP sets of each cluster was tested by Environmental Association Analysis in BayPass version 2.1 (Gautier, 2015). In a first run, the BayPass core model was used with default parameters to generate the population covariance matrix from a set of 50,000 randomly selected sites, which was subsequently used to correct for population structure. Since population structure due to recolonization and similarity in climate where populations occur can show congruent patterns, we also assessed whether the population covariance matrices correlated with any of the climatic variables or latitude by Mantel tests (Mantel, 1967), using 10,000 bootstrap iterations to assess significance. The auxiliary covariate model was then assessed by Mantel tests (Mantel, 1967), using 10,000 bootstrap iterations to examine whether the population covariance matrix from a set of 50,000 randomly selected sites, which was subsequently used to correct for population structure, is correlated with any of the climatic variables or latitude.

2.4 Definition of outliers, GO enrichment analysis, and evaluation of shared SNPs, genes and GO terms

Parallel adaptation was studied at the 750,998 shared sites on the level of SNPs, genes and gene ontology (GO) terms, and we additionally aimed to narrow down the numbers of the latter two to outlier genes and enriched GO terms. To identify outlier SNPs, we used two different approaches: (a) We defined outlier SNPs as those SNPs whose average Bayes Factor (BF) over 10 runs was among the highest 1% for each cluster and climatic variable (7,510 SNPs, hereafter “BF outlier SNPs”). (b) Outlier SNPs were detected by a Hidden Markov Model (HMM) on mean BF values from 10 runs following Soria-Carrasco et al. (2014) (hereafter “HMM outlier SNPs”). Specifically, we modelled the distribution of BF among the highest 1% for each cluster and climatic variable (7,510 SNPs, hereafter “BF outlier SNPs”). Outlier genes were identified using the approach by Yeaman et al. (2016), which detects genes with a high number of outlier SNPs relative to the total number of SNPs using the 0.9999 percentile of the binomial distribution as threshold. We restricted this analysis to genes with at least one outlier SNP and at least three total SNPs. The A. thaliana orthologues were retrieved from the updated A. lyrata genome annotation (Rawat et al., 2015); gene
name and description were obtained from TAIR (The Arabidopsis Information Resource, www.arabidopsis.org). GO enrichment analysis was conducted using Snp2go 1.0.5 (Szkiba, Kapun, von Haeseler, & Gallach, 2014) for each cluster, climatic variable and SNP detection approach using a False Discovery Rate (FDR) of 0.05. This method is based directly on outlier SNPs rather than genes and accounts for clustering of outliers with respect to position and coverage of the genome, and for clustering of similarly annotated genes. The results of GO enrichment were visualized in REVIGO (Supek, Bošnjak, Škunca, & Šmuc, 2011) using the A. thaliana database with similarity set to 0.7 to discard redundant GO terms. For all analyses of GO terms, we restricted our analyses to terms of biological processes, and we report only outlier genes and enriched GO terms detected using both SNP outlier detection approaches.

To evaluate whether the number of BF outlier SNPs shared between eastern and western cluster was higher or lower than expected, we used a resampling approach. We randomly sampled the number of outlier SNPs (7,510) from all SNPs (750,998) separately for the eastern and western cluster. We then calculated the overlap for a total of 10,000 replicates and compared the results with observed values.

### 2.5 | Signatures of directional selection in outlier genes

We estimated the strength of selection on SNPs in SelEstim 1.1.7 (Vitalis, Gautier, Dawson, & Beaumont, 2014), which implements an infinite island model to detect directional selection. The eastern and western SNP data sets were analyzed separately, with five pilot runs of 1,000 iterations each, followed by 10,000 iterations of sampling every 20 steps, of which the first 1,000 were discarded as burnin. The locus-specific selection coefficient $\delta_j$ was used as a measure of the strength of selection. To address whether outlier genes had higher selection coefficients than other genes we calculated mean selection coefficients for all genes with at least 10 SNPs for both clusters. We then tested for differences in mean selection coefficients within each cluster between (a) outlier genes, (b) those genes detected as outlier genes only in the respective other cluster and (c) all nonoutlier genes using ANOVA with a Tukey post hoc comparison in R.

To test whether outlier genes had a signature of elevated linkage disequilibrium (LD), we estimated LD for each population and gene from our read data. Because we worked with pool-seq data, we were restricted to assess LD within paired sequence reads (Feder, Petrov, & Bergland, 2012). Given the distribution of insert sizes, we focused on LD within 200 bp distance. We calculated the squared correlation coefficient $r^2$ between pairs of SNPs using the direct estimate method of LDx (Feder et al., 2012). We only considered sites with an intersecting read depth greater than five and a minor allele frequency $>0.15$ at either locus (Feder et al., 2012; Tuttle et al., 2016). We estimated LD weighted for the distance between SNPs for each gene and population and tested for significant differences in LD between outlier genes and all other genes within each cluster using a linear mixed-effects model implemented in the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). We used outlier class as fixed effect and population as random effect. This analysis was restricted to genes with at least 10 SNPs in the data set. A table of genes and their respective lengths, number of SNPs and relative SNP content is given in Appendix S4.

### 2.6 | Analysis of standing genetic variation in core and edge populations

In the case of nonparallelism in genomic patterns of adaptation, genetic drift could be a likely reason. Arabidopsis lyrata was shown to have been exposed to strong genetic drift and the loss of genomewide diversity along its expansion routes (Willi et al., 2018), and had possibly undergone bottlenecks in its refugia. Therefore, the species may have experienced a loss in the potential for parallel adaptation from standing genetic variation, which could account for lineage-specific adaptation. To assess whether the presence of standing genetic variation could have allowed for selection of the same loci in the expanding populations of both clusters, or whether a lack of standing genetic variation reduced the potential for parallelism, we focused on the core populations, from which recolonization occurred. Comparisons involved three populations for each core area (MD4, PA3 and NY1 in the east and WI1, WI2 and WI3 in the west, indicated in Figure 1), the reconstructed ancestral area from which the two lineages expanded postglacially (Willi et al., 2018). We evaluated standing genetic variation in each cluster by calculating mean minor allele frequency (MAF) for the three respective populations for outlier SNPs (BF and HMM) detected in the other cluster only. Note that while our cluster-wide filtering strategy only included polymorphic sites (with lowest per-population detection frequency of 0.03), alleles can be fixed across the three selected populations. An absence of polymorphism in the core area would indicate that there may not have been standing genetic variation from which selection could occur at the onset of post-glacial range expansion, thus limiting the potential for parallel adaptation. We furthermore compared the distribution of minor allele frequencies of all nonshared outlier SNPs using Wilcoxon tests for each climatic variable.

Fixation of alleles can occur through several processes, among them selection and drift. We tested whether nonsynonymous outlier SNPs were fixed more often than silent SNPs at core populations and the edge of the distribution range in both clusters. In addition to the core populations named above, three edge populations were chosen for each cluster (NC2, VA1 and NJ1 in the east and MO1, IL1 and ON7 in the west, see Figure 1), and minor allele frequencies were assessed. The observed fraction of fixed nonsynonymous outlier SNP sites (MAF = 0.0) was compared to a distribution generated through randomly sampling MAF (equivalent to the number of outlier SNPs of the respective data set) of silent nonoutlier sites using 10,000 replicates.
3 | RESULTS

3.1 | Population structure and climate variables within the study area

Two well separated genetic clusters were confirmed by PCA based on all overlapping SNPs (Figure 1b). The investigated climatic variables were not strongly correlated with each other or with latitude and longitude (Pearson’s correlation coefficient $\rho$ between $-0.7$ and 0.72), with the exception of minimum temperature in early spring, which showed a strong correlation with latitude ($\rho = 0.98$; Figure 2). Furthermore, none of the correlations between population structure based on BayPass covariance matrices and climatic variables were significant (see Table S3 in Appendix S1). For precipitation of the wettest quarter and the Priestley–Taylor coefficient (Moisture), parameter ranges were very similar in the eastern and western cluster (Table S1 in Appendix S1, Figure 2). However, the range of minimum temperature in early spring, $T_{\text{min spring}}$, was larger in the western cluster ($-8.2$ to $+3.2^\circ$C) compared to the eastern cluster $-2.8$ to $+4.6^\circ$C; temperature declined stronger towards the northern edge of the distribution in the west, in southern Canada.

3.2 | Environmental association analysis, outliers and GO term enrichment analysis

Results reported here concern the overlapping SNPs only. The 750,998 SNPs detected in both clusters covered large parts of the genome of Arabidopsis lyrata, including 27,023 of 32,069 annotated genes (84.3%). Approximately two thirds of the SNPs (498,305) were located in genes. The outlier SNPs were distributed across all eight chromosomes of the genome (Figures S1–S2 in Appendix S1). Of those, between 65.6% and 72.9% were located within genes for the twelve analyses (three climatic variables, two genetic clusters and two approaches of outlier detection) (Table S4 in Appendix S1).
S1). A total of 15.23% of SNPs were nonsynonymous, as were between 12.9% and 18.0% of outlier SNPs (Table S4 in Appendix S1). This lack of enrichment of nonsynonymous SNPs may be attributed to our strict filtering approaches, which may have led to the identification of linked rather than adaptive SNPs as outliers. The highest amounts of outlier SNPs located in genes and nonsynonymous SNPs (both absolute and relative to the percentage of SNPs located in genes) were detected in the western cluster.

We further identified between 16 and 39 outlier genes for each cluster and climatic variable, and by considering only overlapping outlier genes of the two outlier detection approaches (Figure S3 in Appendix S1). Notably, outlier genes using the HMM approach were almost a perfect subset of those detected using BF (between 51 and 65 genes), with either one or no outlier gene detected with HMM that was not also found with BF. Among the outliers were genes that had previously been identified to either play a direct role in adaptation to climate, or to be associated with traits and pathways potentially involved in climate adaptation (see Table 1 and Section 4). These included FLC (FLOWERING LOCUS C), one of the major regulators of flowering time, and four other genes related to flowering time or reproduction, as well as genes related to thermal tolerance, cold acclimation or water use efficiency. A detailed list including all outlier genes is given in Appendix S3.

The GO enrichment analysis detected between 35 and 110 significantly enriched terms for each cluster and climatic variable when restricting results to only overlapping terms of the two outlier detection methods (Appendix S2). When discarding redundant terms, 21–61 remained (Figures S4–S6 in Appendix S1). Significant GO terms comprised terms related to flower development, reproduction, response to biotic factors (virus/insect/biotic stimulus) and abiotic factors (temperature/light/radiation), circadian rhythm, DNA replication and cell cycle regulation.

### 3.3 | Evaluation of shared SNPs, genes and GO terms

To detect the amount of parallelism in climate adaptation between the eastern and western cluster, outlier SNPs detected in GWAS from both clusters based on the overlapping SNP data set of 750,998 SNPs were compared. Very few (0.05%–0.65%) of the outlier SNPs were shared between clusters for the three climatic variables (BF outliers: 53 SNPs for T_{min spring} and 61 for P_{wettest quarter} and 97 for moisture; HMM outliers: six SNPs for T_{min spring} and Moisture, three SNPs for P_{wettest quarter}; see Table S4, Figure S7 in Appendix S1). We compared the values from BF outlier SNPs with the expected number of shared SNPs determined by using a resampling approach. Observed values were marginally lower than expected by chance for T_{min spring}, marginally higher than expected for Moisture, and in the range of random expectation for P_{wettest quarter} (Figure 3). None of the outlier genes were shared between east and west. In contrast, multiple GO terms were shared between clusters (T_{min spring}; six GO terms/3.9% of enriched GO terms detected for this climate variable, P_{wettest quarter}; 7/5.07%, Moisture: 17/19.54%). These were mostly related to flower development, DNA replication, cell cycle and metabolism (Appendix S2).

Additionally, we counted the number of overlaps between outliers associated with the three climatic factors within each genetic cluster, as detecting the same outliers in GWAS runs of multiple climatic factors suggests more general responses involved in adaptation. The number of outlier SNPs shared between climatic variables was low and reflected the magnitude of correlation detected between variables. In the eastern cluster, where correlation between Moisture and T_{min spring} was higher than in the western cluster (Pearson’s correlation coefficient $r = 0.71$), 775 BF and 113 HMM outlier SNPs, representing 3.6% and 1.51% of the respective total number of outlier SNPs, were shared between those two variables (Figure S7 in Appendix S1). In contrast, correlation between T_{min spring} and Moisture was higher in the west with $r = 0.52$, and 602 BF outlier SNPs and 170 HMM outlier SNPs (2.8% and 1.83%) were shared between those two variables in the western cluster (Figure S7 in Appendix S1). No overlap in outlier SNPs was detected among all three climatic variables in the east, but 34 BF and two HMM outlier SNPs were shared among all variables in the west. No outlier genes were shared between all three variables in both clusters, and only two were shared between two variables in each cluster (P_{wettest quarter} and Moisture in the eastern cluster and T_{min spring} and P_{wettest quarter} in the western cluster). For enriched GO terms, a similar pattern was observed, where climatic variables having a higher correlation in the respective cluster also showed higher amounts of shared outliers, however, the relative proportions were higher (5.44% shared between Moisture and T_{min spring} in the east and 13.27% shared between T_{min spring} and P_{wettest quarter} in the west). While no GO terms were shared between all three variables in the east, nine GO terms (4.59%) were shared in the west. No GO term was enriched in response to all climate variables and in all clusters. However, among the 10 GO terms shared by at least four of the six cluster-environmental variable combinations, nine were related to flower development, with floral organ development and floral whorl development found in five of them (Appendix S2).

### 3.4 | Strength of selection and linkage disequilibrium

To detect whether outlier genes experienced stronger selection than other genes, we analyzed mean selection coefficients and LD values per gene (Figure 4). Mean selection coefficients (mean ± SD: 10.60 ± 9.53 for outliers in the eastern cluster, 8.41 ± 5.60 for outliers in the western cluster) across outlier genes were significantly higher (Tukey post hoc comparison, $p < .001$) than in nonoutlier genes (5.74 ± 3.12 and 6.26 ± 3.53, respectively). The same pattern was observed for LD (mean ± SE: 0.69 ± 0.02 for outliers in the eastern cluster and 0.74 ± 0.02 for outliers in the western cluster compared to 0.59 ± 0.02 and 0.66 ± 0.02 for nonoutlier genes in east and west, respectively). However, LD was also significantly ($p < .001$)
| TABLE 1 | Top candidate genes for climate adaptation in *Arabidopsis lyrata* |
|-----------------|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | A. *lyrata* ID | A. *thaliana* ortholog | Other gene names in A. *thaliana* | Potential involvement | Reference |
| \( T_{\text{min spring}} \) East | AL6G20630 | AT5G10140 | AGAMOUS-LIKE 25, AGL25, FLC, FLF, FLOWERING LOCUS C, FLOWERING LOCUS F, REDUCED STEM BRANCHING 6, RSB6 | Flowering time | Koornneef, Blanksetijn-de Vries, Hanhart, Soppe, and Peeters (1994) |
|                | AL8G36370 | AT5G59820 | ATZAT12, RESPONSIVE TO HIGH LIGHT 41, RHL41, ZAT12 | Cold acclimation | Vogel et al. (2004) |
|                | AL1G24630 | AT1G12980 | DORNROSCHEN, DRN, ENHANCER OF SHOOT REGENERATION 1, ESR1 | Floral organ initiation | Chandler and Werr (2017) |
|                | AL1G48900 | AT1G34350 | ARABIDOPSIS TRANSMEMBRANE PROTEIN 18, TMEM18 | Pollen tube growth | Dou et al. (2016) |
|                | AL1G56890 | AT1G49540 | ATEL2, ELONGATOR PROTEIN 2, ELONGATOR SUBUNIT 2, ELP2 | Biotic and abiotic stress response | Ding and Mou (2015) |
|                | AL3G11120 | AT3G01140 | ATMYB106, MYB DOMAIN PROTEIN 106, MYB106, NOECK, NOK | Cuticle development | Oshima et al. (2013) |
|                | AL3G11420 | AT3G02170 | LNG2, LONGIFOLIA2, TONI RECRUITING MOTIF 1, TRM1 | Cell elongation | Lee et al. (2006) |
| \( P_{\text{wettest quarter}} \) East | AL3G20880 | AT3G09670 | PDP2, PWWP DOMAIN PROTEIN 2 | Flowering time | Zhou et al. (2018) |
|                | AL7G16700 | AT4G34530 | CIB1, CRYPTOCHROME-INTERACTING BASIC-HEX-LOOP-HEX 1 | Flowering time | Liu et al. (2018) |
|                | AL6G39370 | AT5G27660 | ARABIDOPSIS PARK13 HOMOLOG, ATPARK13, DEG14, DEGRADATION OF PERIPLASMIC PROTEINS 14, PARK13 | Thermal tolerance | Basak et al. (2014) |
| Moisture East | AL3G16970 | AT3G06120 | MUTE | Stomatal differentiation | Han et al. (2018) |
|                | AL4G14110 | AT2G24060 | ATIF3-2, ATINFC-2, INITIATION FACTOR 3-2, SUPPRESSOR OF VARIEGATION9, SVR9 | Leaf development | Liu et al. (2019) |
|                | AL8G12700 | AT5G45870 | PYL12, PYR1-LIKE 12, RCAR6, REGULATORY COMPONENTS OF ABA RECEPTOR 6 | Water use efficiency | Yang et al. (2016) |

*Note:* The involvement of the *A. thaliana* orthologues in pathways or traits important for adaptation to environment makes them top candidates in *A. lyrata*. Gene names in both species are given, as well as other names of the *A. thaliana* ortholog (from TAIR; www.arabidopsis.org) and the trait potentially involved in climate adaptation. A list of all outlier genes is given in Appendix S3.
elevated in outlier genes detected in the respective other cluster (0.69 ± 0.02 for outlier genes from the western cluster in the eastern data set and 0.73 ± 0.02 for outlier genes from the eastern cluster in the western data set), indicating that outlier genes were generally located in genomic regions of elevated LD. This was not the case for mean selection coefficient. In the eastern data set, mean selection coefficients for outlier genes of the western cluster (6.27 ± 4.03) were comparable to nonoutlier genes (5.74 ± 3.12), and significantly lower than those of outlier genes in the eastern cluster, indicating that the detected outliers are under stronger selection than other genes. Interestingly, while FLC was not among the genes included in this analysis, as it contained only six polymorphic sites shared by both clusters, mean selection coefficient of this gene was high in the east, where it was detected as an outlier, and low in the west (19.87 and 3.82, respectively).

3.5 | Differential loss of standing genetic variation

We compared minor allele frequencies (MAF) in the core of the distribution of the two genetic clusters as a proxy for original standing genetic variation, on which selection could have acted in both clusters during post-glacial range expansion. Alleles at those sites identified as outliers in one cluster were often polymorphic in core populations of the other cluster and had intermediate allele frequencies. This pattern was detected for all three climatic variables (Figure 5). While low MAFs (≤ 0.05) were common in both cluster cores for the respective outlier SNPs in the other cluster (11.8%–15.3% in the eastern cluster and 22.1%–25.1% in the western cluster), few sites were fixed for either allele (2.1%–4.2% in the eastern cluster and 4.5%–7% in the western cluster), indicating that standing genetic variation was present in the cores of distribution at outlier SNPs of the respective other cluster. We also detected significant differences in the distribution of minor allele frequencies of those SNPs between eastern and western cluster for all three climatic variables (Wilcoxon rank-sum test with continuity correction: W = 46,397,000 (T_{min spring}), W = 51,514,000 (P_{wettest quarter}), W = 44,049,000 (Moisture), p < .001 for all climatic variables), with more low frequency alleles in the western cluster (Figure 5). Fixation among nonsynonymous outlier SNPs was between 0.048 and 0.26 across the two clusters and core or edge of the distribution (Figure S8 in Appendix S1). The highest value was reached at edge populations of the western cluster, while the lowest was observed in western core populations. Compared to fixation levels at silent nonoutlier sites, fixation in nonsynonymous outliers was either higher (eastern core, western edge) or within the range of observation (eastern edge, western core). Notably, fixation levels of silent nonoutliers were lower in core populations than in edge populations for both clusters, and lower in the east than the west.

4 | DISCUSSION

4.1 | Genes and gene ontology terms of adaptation to climatic factors

A first aim of our study was to detect genes and biological processes involved in climate adaptation. Some of the outlier genes found are prime candidates of climate adaptation (Table 1) because previous studies documented their involvement in relevant regulatory pathways and their potential physiological implications. Among the most important pathways were those linked to coping with cold and reproductive development. An outlier for minimum temperature in early spring in the eastern cluster was the cold-acclimation related gene ZAT12. The gene is involved in the CBF pathway (Doherty, Van Buskirk, Myers, & Thomashow, 2009; Vogel, Zarka, Van Buskirk, Fowler, & Thomashow, 2004) (Table 1). Another gene of this pathway, ZAT10, was detected as a candidate for local adaptation in Scandinavian Arabidopsis lyrata subsp. petraea (Mattila, Tyrmi, Pyhäjärvi, & Savolainen, 2017). In Arabidopsis thaliana, the CBF/DREB1 (C-REPEAT/DRE BINDING FACTOR) genes are important regulators of cold acclimation (Thomashow, Gilmour, Stockinger, Jaglo-Ottosen, & Zarka, 2001) and play a role in local adaptation, including parallel adaptation to warmer climates by disruption of their function (Monroe et al., 2016). Another outlier gene detected in response to minimum temperature in early spring in the eastern cluster was FLC, for which our analysis also revealed that it was under strong selection. FLC is one of the central elements of flowering time control by repressing flowering in Arabidopsis thaliana (Michaels & Amasino, 1999) and A. lyrata (Kemi et al., 2013) while being negatively regulated by vernalization (Sheldon, Rouse, Finnegan, Peacock, & Dennis, 2000). Multiple studies have shown its crucial role for adaptation to climate along latitudinal and altitudinal clines in A. thaliana, together with its regulator FRI (FRIGIDA) and an epistatic interaction between the two.

**FIGURE 3** Expected and observed number of outlier SNPs shared between genetic clusters of Arabidopsis lyrata. Expectation was calculated using a random resampling approach with 10,000 replicates and is indicated in grey (the two 2.5% tails of the distribution in light red). Observed values are indicated in blue (T_{min spring}), green (P_{wettest quarter}) and orange (Moisture)
genes (e.g., Koornneef, Alonso-Blanco, Peeters, & Soppe, 1998; Korves et al., 2007; Méndez-Vigo, Pico, Ramiro, Martínez-Zapater, & Alonso-Blanco, 2011; Shindo et al., 2005; Stinchcombe et al., 2004). Regulators of flowering time were also found to be associated with precipitation in the eastern cluster. We detected PDP2 (PWWP DOMAIN PROTEIN 2) and CIB1 (CRYPTOCHROME-INTERACTING BASIC-HELIX-LOOP-HELIX 1), which are involved in the autonomous and photoperiod pathway of flowering time regulation, respectively (Liu et al., 2018; Zhou et al., 2018). Signatures of selection and local adaptation have also been detected in different genes from these two pathways in A. lyrata subsp. petraea (Mattila et al., 2016). Early flowering was shown to be an effective strategy of adaptation to coping with drought, via drought avoidance (Franks & Weis, 2008).

Top candidate genes of climate adaptation in the western cluster were not related to the well known pathways of flowering time; however, two outlier genes detected for minimum temperature in early spring were involved in other aspects of reproduction (pollen tube growth and floral organ initiation; see Table 1), and two genes were involved in leaf development (cell elongation and cuticle development). Common garden experiments had shown latitudinal clines for plant size and reproductive development in A. lyrata, with plants from the north growing larger rosettes under benign temperature conditions and having a higher propensity to flower early, independent of genetic cluster (Paccard, Fruleux, & Willi, 2014; Wos & Willi, 2015). Early spring temperature strongly correlates with latitude, particularly in the western cluster (Figure 2), and thus our GWAS may have detected the genetic basis for the observed phenotypes, making our outlier genes candidates for future studies. Particularly the fair number of flowering-related genes in our outliers in east and west suggests that pathways of reproductive phenology are important in climate adaptation – both along temperature and precipitation gradients. Further candidates of climate adaptation, particularly to precipitation and water availability, were genes involved in (thermal) tolerance, stomatal differentiation and water-use efficiency. In a complex climatic landscape involving multiple stress factors, adaptation to the local combination of such factors are thus facilitated by modulating different parts of the pathways responsible for physiological or developmental responses.

Furthermore, many of the enriched gene ontology (GO) terms were related to transcription or metabolic processes (Appendix S2). Transcription factors, kinases and genes involved in hormone signaling had been identified to probably influence stress responses to multiple biotic and abiotic factors and facilitate crosstalk between them (Fujita et al., 2006). Other significant terms included flower development and shoot development, processes which had been shown to be important determinants of life history traits in two populations of A. lyrata originating from different climatic backgrounds (Remington, Figueroa, & Rane, 2015). Generally, biological processes involved in adaptation to environment seem to be comparable between different species of the

FIGURE 4 Selection and linkage in outlier genes. Mean strength of selection and LD was calculated per gene for all genes with at least 10 SNPs present. Boxplots for three groups per cluster are shown: Outlier genes, genes that were detected as outliers in the respective other cluster and all other (nonoutliers) genes. Eastern cluster is indicated in blue, western cluster in purple in both panels. Significance within clusters was calculated using linear mixed effect models; and significant differences are indicated with bars above the boxplots (p-value < .001, ***; < .01, **).
In the genus *Arabidopsis*, a comprehensive study across large parts of the native distribution identified biological processes enriched for various climate variables (Hancock et al., 2011). Many of these were related to development, transcription, DNA repair and metabolism and were also identified in our study. In *Arabidopsis halleri*, a closer relative of *A. lyrata*, 16 GO terms associated with adaptation to various climatic and topographic parameters were identified (Fischer et al., 2013), six of which (e.g., aging, DNA methylation, chromatin silencing, protein autophosphorylation) were also detected in our analyses. Interestingly, comparing the underlying genes in *A. halleri* with the outliers we found in *A. lyrata* did not yield any shared genes despite the apparent similarities in biological processes. The same pattern was observed comparing our outlier genes to those related to local adaptation in Scandinavian *A. lyrata* subsp. petraea (Mattila et al., 2017). Other studies focusing on particular stress responses also identified similar GO terms, although not necessarily in response to similar climate variables: Three out of 65 terms enriched in the eastern cluster in response to moisture availability, linked to circadian rhythm, response to reactive oxygen species and red or far-red light signalling, were also found to be among the most significantly enriched GO terms involved in freezing tolerance in *A. thaliana* (Horton, Willems, Sasaki, Koornneef, & Nordborg, 2016), but also in this case the outlier genes were not shared. Our results thus suggest that adaptation to climate in *Arabidopsis* often seems to involve similar biological processes, despite the different habitats the species naturally occur in. Whether these results are transferable to other plant species outside this clade is yet to be shown.

### 4.2 Parallelism in climate adaptation

Our second goal was to determine the extent to which the same SNPs, genes or GO terms were associated with climate in both lineages, interpreted as evidence for parallel evolution. We detected only few shared SNPs associated with climatic variables between the two clusters, and hardly more than expected by chance alone (Figure 3). The level of shared genes was even lower, with none shared between clusters, and only two shared between climates within each cluster. Nevertheless, selection was much higher in outlier compared to nonoutlier genes, corroborating their importance for adaptation. The reduced level of selection in the eastern data set for outlier genes detected in the western cluster further emphasized the importance of lineage-specific adaptation to climate in *A. lyrata* on the genomic level.

In contrast to the low number of shared outlier SNPs and genes, parallelism was higher at the level of GO terms, indicating that despite the prevailing lineage-specific adaptation on the genomic level, similar biological processes were involved in adaptation across the species range. An analogous pattern was found in experimental evolution studies in bacteria, where environmental adaptation was also shaped hierarchically with few shared nucleotide changes but many similar pathways affected (Bailey et al., 2015). In *A. halleri*, adaptation to environment in the Swiss Alps was found to be mostly local even on a small geographic scale, putatively due to the complex and heterogeneous environment in their study area (Rellstab et al., 2017).

**FIGURE 5** Standing genetic variation in core populations for the three climate variables. To assess whether a lack of standing genetic variation in the core populations (from which the species expanded post-glacially) could have led to the observed patterns of nonparallelism in outlier SNPs, we calculated minor allele frequency (in classes of 0.05) at outlier loci from the respective other genetic cluster for the three climate variables separately. This analysis was restricted to the three core populations of the target cluster; mean frequency across the three population in the core was calculated based on number of SNPs with no missing data. $T_{\text{min spring}}$ is indicated in blue, $P_{\text{wettest quarter}}$ in green and Moisture in red orange in both panels. Upper bounds of frequency classes are given below the barplots.
We also detected outliers shared between GWAS runs for the different climatic parameters but within clusters, suggesting that adaptation to different climatic factors also often involves the same biological processes, but different genes/SNPs. In line with our top candidate genes, floral whorl development and floral organ development were the two GO terms detected in five out of six climate-cluster combinations. Overall, patterns of outlier SNPs shared between climatic variables within clusters were consistent with the strength of the correlations between climatic variables. More outlier SNPs were shared between the two precipitation-related variables in the eastern cluster, where these two were more strongly correlated with each other, while in the western cluster the number of shared SNPs between precipitation of the wettest quarter and minimum temperature in early spring was higher, as was their correlation (Figure 2, Figure S7 in Appendix S1). The pattern was similar for outlier genes, although the proportion of shared genes was relatively low, indicating that adaptation is not only lineage-specific, but also environment-specific to some extent.

### 4.3 Reasons for a lack of parallelism on the level of genetic variants and genes

The third goal of our study was to address whether the presence or absence of standing genetic variation can account for the amount of parallel adaptation to climate in the two lineages of *A. lyrata*. Our analysis of allele frequencies at outlier SNPs indicated that standing genetic variation, on which selection could have acted, was present as intermediate allele frequencies in both core areas of the distribution, from where post-glacial range expansion started (Figure 5). The Pleistocene with its climate oscillations posed major challenges for many species, as they had to repeatedly move into new areas, adapt to new environments and survive in refuge areas during glacial periods (Hewitt, 2000). Isolation in refuge areas and subsequent recolonization often reduced population sizes and standing genetic variation available for adaptation, as for example found for North American *A. lyrata* (Willi et al., 2018). While standing genetic variation can be the basis for parallel adaptation to similar environmental gradients (Jones et al., 2012), allele frequencies may have been perturbed during demographic bottlenecks imposed by the glacial history, causing lineage-specific patterns of adaptation. This seems to probably have happened in *A. lyrata*: Populations in the western cluster were shown to have started from a genetically less diverse refuge area and expanded over longer distances since the withdrawal of ice (Willi et al., 2018). Concordantly, we detected more low frequency alleles in the core populations of the western cluster in those SNPs identified as outliers in the eastern cluster, which could reflect stronger bottleneck effects in the west. This pattern was also observed in the higher amounts of fixed sites in the western cluster in both core and edge populations. In particular at the western distribution edge, a high proportion (~26%) of nonsynonymous outlier SNPs were fixed, probably through drift, leading to a loss of genetic variation along the expansion route, which was particularly long in the west (Willi et al., 2018). In addition, strong selection may have led to an excess of fixed nonsynonymous outliers compared to silent nonoutliers. Low genetic diversity at candidate loci in recently colonized areas was also found in *A. lyrata* subsp. *petraea* in Scandinavia, where it was associated with selective sweeps (Mattila et al., 2017). These patterns of selection may be associated with recent colonization and the required adaptation to new environments. In core populations, however, where genetic diversity is higher, more outliers in the east were fixed compared to nonoutliers, potentially due to selection. Interestingly, while we detected higher selection specifically in outlier genes, a different pattern was detected for LD, suggesting that the high LD we detected was not generally a result of selection. Outlier genes might be located in regions of high LD. A similar pattern was found in *A. lyrata* subsp. *petraea*, where outliers related to local adaptation were located in regions of low recombination rate and high gene density (Hämälä & Savolainen, 2019). Lower recombination rates may contribute to population divergence through reducing the effects of gene flow (e.g., Renaut et al., 2013) and thus be important for local adaptation through coupling of coadapted alleles (Bierne et al., 2011; Butlin & Smadja, 2018).

The chance for parallel evolution from new mutations is predicted to be low when many beneficial mutations are possible (Orr, 2005). This is likely to be the case in polygenic traits such as adaptation to environment in the complex genomes of eukaryotes and can be explained by redundant pathways leading to adaptation in highly polygenic traits, including multiple SNPs affecting the same gene in a similar way, and several genes involved in the same pathway. Considering the relatively short time for adaptation after the glacial retreat and the highly polygenic nature of adaptation to climate, it seems plausible that many of the alleles affected by adaptation needed to be recruited from standing genetic variation, and the drifting of allele frequencies probably lowered the chances of parallel adaptation on the level of SNPs and genes.

Additionally, some differences in SNPs/genes detected in the two clusters might be the result of differences in climate, such as the differences in the range of climatic variables covered by the populations, and differences in the variance of climatic conditions within sites between east and west. Difference in climatic range was most pronounced for minimum temperature in early spring, as much lower temperatures are reached in the west compared to the east (Figure 2). Furthermore, climate is the result of a complex interplay between different climatic variables. Differences in the combination of climatic variables and the resulting selective pressures may have influenced the divergent adaptation in the two areas of *A. lyrata*. Our results are similar to those found in black spruce in North America. In this species postglacial range expansion combined with climate adaptation was found to be mostly lineage-specific and was hypothesized to be based predominantly on selection from standing genetic variation (Prunier et al., 2012), and the authors speculated that both environmental heterogeneity and stronger drift associated with loss of genetic diversity in one of the lineages could account for lineage-specificity.

In conclusion, using whole-genome pool-seq data covering the entire distribution range of *A. lyrata* subsp. *lyrata* in North America,
we identified genes and biological processes involved in adaptation to niche-determining climate variables. Gene ontology terms related to flower development were found repeatedly and regulators of flowering time were among our outlier genes, emphasizing the crucial role of the timing of reproduction for local adaptation. We found little evidence for intraspecific parallel adaptation to a comparable climatic gradient on the level of SNPs and genes between two postglacial lineages. Nevertheless, our results indicate a genetic basis of adaptation to the climatic factors that shape the distribution of this species, i.e., minimum temperature in early spring and two factors depicting water availability. This was most prominently shown by the significantly elevated selection coefficients found for outlier genes, and the fact that similar biological processes were affected in both lineages. We hypothesize that the highly polygenic nature of adaptation to climate and a Pleistocene population history with significant action of genetic drift limited the potential for parallel evolution.

ACKNOWLEDGEMENTS
Collection permits were granted by: Clinton County Conservation Board IA, Commonwealth of Pennsylvania, Cornell University, Fort Leonard Wood Army Base, Illinois Department of Natural Resources, Iowa Department of Natural Resources, Maryland Department of Natural Resources, Michigan Department of Natural Resources, Nature Conservancy of Eastern New York, Nature Conservancy of Maryland, Nature Conservancy of Ohio, New York State Office of Parks, Ontario Parks, Palisades Interstate Park Commission NY, The United States National Park Service, Virginia Department of Conservation and Recreation, Wisconsin Department of Natural Resources, and the Ojibways of the Pic River First Nation, the owners of the Rock Island Lodge, and John Haataja. This work was supported by the Swiss National Science Foundation (PP00P3_123396, PP00P3_146342, 31003A_140979, 31003A_166322), and the Fondation Pierre Mercier pour la Science, Lausanne. Sequencing was done at the Genetic Diversity Centre ETH Zürich, and at the Department of Biosystems Science and Engineering of ETH Zürich in Basel and the University of Basel. Marco Fracassetti and Olivier Bachmann prepared the libraries for sequencing. Calculations were performed at sciCORE (http://scicore.unibas.ch/) scientific computing core facility of the University of Basel.

AUTHOR CONTRIBUTIONS
N.W. and Y.W. designed the research. Y.W. conducted the field sampling. N.W. and K.L. analysed the data. N.W. wrote the manuscript.

DATA AVAILABILITY STATEMENT
Results of GWAS (mean BayesFactor for each cluster and climatic variable), table of gene length and SNP content, as well as detailed tables with results of outlier gene and enriched GO term detection (Appendix S2–S4) are available at Dryad (https://doi.org/10.5061/dryad.qfftdz0cq). Raw reads from sequencing data are available at the European Nucleotide Archive (ENA) under study accession PRJEB19338.

ORCID
Nora Walden https://orcid.org/0000-0002-3078-6791
Kay Lucek https://orcid.org/0000-0002-2253-2556
Yvonne Willi https://orcid.org/0000-0001-7654-3261

REFERENCES
Arendt, J., & Reznick, D. (2008). Convergence and parallelism reconsidered: What have we learned about the genetics of adaptation? Trends in Ecology & Evolution, 23(1), 26–32. https://doi.org/10.1016/j.tree.2007.09.011
Bailey, S. F., Rodrigue, N., & Kassen, R. (2015). The effect of selection environment on the probability of parallel evolution. Molecular Biology and Evolution, 32(6), 1436–1448. https://doi.org/10.1093/molbev/msv033
Barrett, R. D. H., & Hoekstra, H. E. (2011). Molecular spandrels: Tests of adaptation at the genetic level. Nature Reviews Genetics, 12(11), 767–780. https://doi.org/10.1038/nrg3015
Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. Trends in Ecology & Evolution, 23(1), 38–44. https://doi.org/10.1016/j.tree.2007.09.008
Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., … Kim, J. F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature, 461(7268), 1243–1247. https://doi.org/10.1038/nature08480
Basak, I., Pal, R., Patil, K. S., Dunne, A., Ho, H.-P., Lee, S., … Møller, S. G. (2014). Arabidopsis AtPARK13, which confers thermotolerance, targets misfolded proteins. Journal of Biological Chemistry, 289(21), 14458–14469. https://doi.org/10.1074/jbc.M114.548156
Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01
Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. The Annals of Mathematical Statistics, 41(1), 164–171. https://doi.org/10.1214/aoms/1177691796
Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: Why genome scans may fail to map local adaptation genes. Molecular Ecology, 20(10), 2044–2072. https://doi.org/10.1111/j.1365-294X.2011.05080.x
Bombles, K., & Weigel, D. (2007). Arabidopsis—a model genus for speciation. Current Opinion in Genetics & Development, 17(6), 500–504. https://doi.org/10.1016/j.gde.2007.09.006
Butlin, R. K., & Smadja, C. M. (2018). Coupling, reinforcement, and speciation. The American Naturalist, 191(2), 155–172. https://doi.org/10.1086/695136
Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal, G., Shapiro, M. D., Brady, S. D., … Kingsley, D. M. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science, 327(5963), 302. https://doi.org/10.1126/science.1182213
Chandler, J., & Werr, W. (2017). DORNRÖSCHEN, DORNRÖSCHEN-LIKE, and PUCHI redundantly control floral meristem identity and organ initiation in Arabidopsis. Journal of Experimental Botany, 68(13), 3457–3472. https://doi.org/10.1093/jxb/erx208
Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., … Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly, 6(2), 80–92. https://doi.org/10.4161/fly.19695
Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., … Kingsley, D. M. (2005). Widespread parallel
evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science, 307(5717), 1928. https://doi.org/10.1126/science.1107239

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... 1000 Genomes Project Analysis Group (2011). The variant call format and VCF tools. Bioinformatics, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Ding, Y., & Mou, Z. (2015). Elongator and its epigenetic role in plant development and responses to abiotic and biotic stresses. Frontiers in Plant Science, 6, 296. https://doi.org/10.3389/fpls.2015.00296

Doherty, C. J., Van Buskirk, H. A., Myers, S. J., & Thomashow, M. F. (2009). Roles for Arabidopsis CAMTA transcription factors in cold-regulated gene expression and freezing tolerance. The Plant Cell, 21(3), 972. https://doi.org/10.1105/tpc.108.063958

Dou, X.-Y., Yang, K.-Z., Ma, Z.-X., Chen, L.-Q., Zhang, X.-Q., Bai, J.-R., & Ye, D. (2016). At7MEM18 plays important roles in pollen tube and vegetative growth in Arabidopsis. Journal of Integrative Plant Biology, 58(7), 679–692. https://doi.org/10.1111/jipb.12459

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), 352–358. https://doi.org/10.1038/s41559-017-0423-0

Feder, A. F., Petrov, D. A., & Bergland, A. O. (2012). LDEx: Estimation of linkage disequilibrium from high-throughput pooled resequencing data. PLoS ONE, 7(11), e48588. https://doi.org/10.1371/journal.pone.0048588

Fischer, M. C., Rellstab, C., Tedder, A., Zoller, S., Gugerli, F., Shimizu, K., ... Tedder, A., Haudry, A., Wright, S. I., & Mable, B. (2014). Evolutionary shifts to self-fertilisation rewire the range of Arabidopsis thaliana. Plant, Cell & Environment, 39(11), 2570–2579. https://doi.org/10.1111/pce.12521

Fracassetti, M., Griffin, P. C., & Willi, Y. (2015). Validation of pooled parallel sequencing of individual and pooled samples. PLoS ONE, 10(10), e0140462. https://doi.org/10.1371/journal.pone.0140462

Fuxe, J. P., Stift, M., Tedder, A., Haudry, A., Wright, S. I., & Mable, B. K. (2010). Reconstructing origins of loss of self-incompatibility and selfing in North American Arabidopsis lyrata: A population genetic context. Evolution, 64(12), 3495–3510. https://doi.org/10.1111/j.1558-5646.2010.01094.x

Futschik, A., ... Schlötterer, C. (2011). PoPoolation: A toolbox for the single symmetric division to create stomata. Developmental Cell, 19(6), 907–913. https://doi.org/10.1016/j.devcel.2010.07.004

Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. Genetics, 204(1), 1555–1579. https://doi.org/10.1534/genetics.115.181453

Griffin, P. C., & Willi, Y. (2014). Evolutionary shifts to self-fertilisation restricted to geographic range margins in North American Arabidopsis lyrata. Ecology Letters, 17(4), 484–490. https://doi.org/10.1111/ele.12248

Hämälä, T., & Savolainen, O. (2019). Genomic patterns of local adaptation under gene flow in Arabidopsis lyrata. Molecular Biology and Evolution, 36(11), 2557–2571. https://doi.org/10.1093/molbev/msz2149

Han, S.-K., Qi, X., Sugihara, K., Deng, J. H., Endo, T. A., Miller, K. L., ... Torii, K. U. (2018). MUTE directly orchestrates cell-state switch and the single symmetric division to create stomata. Developmental Cell, 45(3), 303–315.e5. https://doi.org/10.1016/j.devcel.2018.04.010

Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G., ... Bergelson, J. (2011). Adaptation to climate across the Arabidopsis thaliana genome. Science, 334(6052), 83–86. https://doi.org/10.1126/science.1209244

Harte, D. (2017). Package ‘HiddenMarkov’. Bioinformatics, 33(21), 4305–4314. https://doi.org/10.1093/bioinformatics/btx260
Trabucco, A., & Zomer, R. (2010). Global soil water balance geospatial database. CGIAR consortium for spatial information.

Turner, T. L., Levine, M. T., Eckert, M. L., & Begun, D. J. (2008). Genomic analysis of adaptive differentiation in Drosophila melanogaster. Genetics, 179(1), 455–473. https://doi.org/10.1534/genetics.107.083659

Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., ... Balakrishnan, C. N. (2016). Divergence and functional degradation of a sex chromosome-like supergene. Current Biology, 26(3), 344–350. https://doi.org/10.1016/j.cub.2015.11.069

Vitalis, R., Gautier, M., Dawson, K. J., & Beaumont, M. A. (2014). Detecting and measuring selection from gene frequency data. Genetics, 196(3), 799–817. https://doi.org/10.1534/genetics.113.152991

Vogel, J. T., Zarka, D. G., Van Buskirk, H. A., Fowler, S. G., & Thomashow, M. F. (2004). Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of Arabidopsis. The Plant Journal, 41(2), 195–211. https://doi.org/10.1111/j.1365-313X.2004.02288.x

Wichman, H. A. (1999). Different trajectories of parallel evolution during viral adaptation. Science, 285(5426), 422–424. https://doi.org/10.1126/science.285.5426.422

Willi, Y., Fracassetti, M., Zoller, S., & Van Buskirk, J. (2018). Accumulation of mutational load at the edges of a species range. Molecular Biology and Evolution, 35(4), 781–791. https://doi.org/10.1093/molbev/msy003

Willi, Y., & Määttänen, K. (2010). Evolutionary dynamics of mating system shifts in Arabidopsis lyrata. Journal of Evolutionary Biology, 23(10), 2123–2131. https://doi.org/10.1111/j.1420-9101.2010.02073.x

Woods, R., Schneider, D., Winkworth, C. L., Riley, M. A., & Lenski, R. E. (2006). Tests of parallel molecular evolution in a long-term experiment with Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America, 103(24), 9107–9112. https://doi.org/10.1073/pnas.0602917103

Wos, G., & Willi, Y. (2015). Temperature-stress resistance and tolerance along a latitudinal cline in North American Arabidopsis lyrata. PLoS ONE, 10(6), e0131808. https://doi.org/10.1371/journal.pone.0131808

Yang, Z., Liu, J., Tischer, S. V., Christmann, A., Windisch, W., Schnyder, H., & Grill, E. (2016). Leveraging abscisic acid receptors for efficient water use in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America, 113(24), 6791–6796. https://doi.org/10.1073/pnas.1601954113

Yeaman, S., Gerstein, A. C., Hodgins, K. A., & Whitlock, M. C. (2018). Quantifying how constraints limit the diversity of viable routes to adaptation. PLOS Genetics, 14(10), e1007717. https://doi.org/10.1371/journal.pgen.1007717

Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C., ... Aitken, S. N. (2016). Convergent local adaptation to climate in distantly related conifers. Science, 353(6306), 1431–1433. https://doi.org/10.1126/science.aaf7812

Zhou, J.-X., Liu, Z.-W., Li, Y.-Q., Li, L., Wang, B., Chen, S., & He, X.-J. (2018). Arabidopsis PWWP domain proteins mediate H3K27 trimethylation on FLC and regulate flowering time. Journal of Integrative Plant Biology, 60(5), 362–368. https://doi.org/10.1111/jipb.12630

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Walden N, Lucek K, Willi Y. Lineage-specific adaptation to climate involves flowering time in North American Arabidopsis lyrata. Mol Ecol. 2020;29:1436–1451. https://doi.org/10.1111/mec.15338