ORIGINAL RESEARCH

In the presence of population structure: From genomics to candidate genes underlying local adaptation

Nicholas Price[1,2] [iD] | Lua Lopez[3] | Adrian E. Platts[4,5] | Jesse R. Lasky[6]

[1]Department of Bioagricultural Sciences & Pest Management, Colorado State University, Fort Collins, CO, USA

[2]Department of Biological Sciences, University of Cyprus, Nicosia, Cyprus

[3]Department of Biology, Binghamton University (State University of New York), Binghamton, NY, USA

[4]Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

[5]Department of Biology, Center for Genomics and Systems Biology, New York University, New York, NY, USA

[6]Department of Biology, Pennsylvania State University, University Park, PA, USA

Correspondence
Nicholas Price, Department of Bioagricultural Sciences & Pest Management, Colorado State University, Fort Collins, CO 80523, USA.
Email: price4890@gmail.com

Abstract

Understanding the genomic signatures, genes, and traits underlying local adaptation of organisms to heterogeneous environments is of central importance to the field evolutionary biology. To identify loci underlying local adaptation, models that combine allelic and environmental variation while controlling for the effects of population structure have emerged as the method of choice. Despite being evaluated in simulation studies, there has not been a thorough investigation of empirical evidence supporting local adaptation across these alleles. To evaluate these methods, we use 875 *Arabidopsis thaliana* Eurasian accessions and two mixed models (GEMMA and LFMM) to identify candidate SNPs underlying local adaptation to climate. Subsequently, to assess evidence of local adaptation and function among significant SNPs, we examine allele frequency differentiation and recent selection across Eurasian populations, in addition to their distribution along quantitative trait loci (QTL) explaining fitness variation between Italy and Sweden populations and cis-regulatory/nonsynonymous sites showing significant selective constraint. Our results indicate that significant LFMM/GEMMA SNPs show low allele frequency differentiation and linkage disequilibrium across locally adapted Italy and Sweden populations, in addition to a poor association with fitness QTL peaks (highest logarithm of odds score). Furthermore, when examining derived allele frequencies across the Eurasian range, we find that these SNPs are enriched in low-frequency variants that show very large climatic differentiation but low levels of linkage disequilibrium. These results suggest that their enrichment along putative functional sites most likely represents deleterious variation that is independent of local adaptation. Among all the genomic signatures examined, only SNPs showing high absolute allele frequency differentiation (AFD) and linkage disequilibrium (LD) between Italy and Sweden populations showed a strong association with fitness QTL peaks and were enriched along selectively constrained cis-regulatory/nonsynonymous sites. Using these SNPs, we find strong evidence linking flowering time, freezing tolerance, and the abscisic-acid pathway to local adaptation.

1 | INTRODUCTION

Populations of a species may inhabit different environments where local selection pressures favor a combination of (multivariate) phenotypes (Conover, Duffy, & Hice, 2009; Hereford, 2009; Leimu & Fischer, 2008; Savolainen, Lascoux, & Merila, 2013). Local adaptation, by definition, occurs when the resident genotype is expected, on average, to have a higher relative fitness than a foreign genotype (Kawecki & Ebert, 2004). Despite the widespread evidence of local adaptation in many taxa (Arguello et al., 2016; Jeong & Di Rienzo, 2014; Leimu & Fischer, 2008), our understanding of the traits involved, its genetic basis, and its environmental underpinnings is still at an infant stage (Wadgymar et al., 2017, Tiffin & Ross-Ibarra, 2014, Savolainen et al., 2013).

In a variety of species, reciprocal transplant and common garden/laboratory experiments have showed significant adaptive differentiation between natural populations inhabiting different environments (Ågren & Schemske, 2012; Hendry, Taylor, & Mcphail, 2002; Kaufmann, Lenz, Kalbe, Milinski, & Eizaguirre, 2017; Phifer-Rixey et al., 2018; Savolainen, Pyhäjärvi, & Knürr, 2007; Via, 1991). Furthermore, in plants and animals, mapping experiments have uncovered quantitative trait loci (QTL) for traits that are thought to underlie local adaptation (Ågren, Oakley, Lundemo, & Schemske, 2017; Colosimo et al., 2004; Oakley, Agren, Atchison, & Schemske, 2014; Yang, Guo, Shikano, Liu, & Merila, 2016), in addition to QTL explaining fitness differences across environments (Ågren, Oakley, McKay, Lovell, & Schemske, 2013; Anderson, Lee, Rushworth, Colautti, & Mitchell-Olds, 2013). Despite the importance of QTL studies in providing direct evidence for local adaptation (Ågren et al., 2013), in many instances they provide a low resolution for its genetic basis, and in practical terms are time-consuming, expensive, and labor-intense (Joosen, Ligterink, Hilhorst, & Keurentjes, 2009).

With the advent of low-cost, and fast, next-generation sequencing (Henson, Tischler, & Ning, 2012), higher-resolution population genomic approaches have emerged as the new means for examining the genetic basis of local adaptation (Tiffin & Ross-Ibarra, 2014), (Lachance & Tishkoff, 2013; Savolainen et al., 2013; Sork, 2017). In brief, these methods include (a) identifying single nucleotide polymorphisms (SNPs) showing significant allele frequency differentiation between populations (F_{ST}) (Beaumont & Balding, 2004; De Villemereuil & Gaggiotti, 2015; Foll & Gaggiotti, 2008); (b) identifying genomic regions showing significant increases in linkage disequilibrium (Jacobs, Sluckin, & Kivisild, 2016) or composite likelihood ratios for recent sweeps (Degiorgio, Huber, Hubisz, Hellmann, & Nielsen, 2016; Huber, Degiorgio, Hellmann, & Nielsen, 2016); and (c) identifying genotype-by-environment associations (GEA; Caye, Jumentier, Lepeule, & Francois, 2019; Gunther & Coop, 2013; Lasky

et al., 2012; Luu, Bazin, & Blum, 2017; Zhou & Stephens, 2012). The latter approach has gained attention because it can be implemented based on individual (as opposed to population-based) sampling, and furthermore, it provides a direct link to ecologically relevant factors (e.g., climate).

An important hurdle that GEA and other methods need to overcome, is disentangling the effects of selection from demographic history (Hoban et al., 2016; Lotterhos & Whitlock, 2014). While nonadaptive processes (Bierne, Welch, Loire, Bonhomme, & David, 2011; Bonhomme et al., 2010; Hofer, Ray, Wegmann, & Excoffier, 2009) can generate population structure along the genome, and therefore lead to spurious genotype-by-environment associations, geographically varying environments can generate population structure in the regions of genes involved in adaptation (Mckay and Latta, (2002). To limit the number of spurious associations, studies usually estimate population structure using different methods (Price, Zaitlen, Reich, & Patterson, 2010), incorporate population structure (Kang et al., 2010; Wang et al., 2011; Yu et al., 2006; Zhou & Stephens, 2012) and/or geographic structure (Lasky et al., 2012) into statistical models, and finally test whether certain loci explain significantly higher variation in environment/climate than population structure itself (Fischer et al., 2013; Frachon et al., 2018; Hancock, Brachi, et al., 2011; Huber, Nordborg, Hermisson, & Hellmann, 2014; Lasky et al., 2014, 2012; Lasky, Forester, & Reimherr, 2018; Monroe et al., 2016; Price et al., 2018; Rellstab et al., 2017). While such approaches may limit the number of false positives, they can also lead to false negatives (Anderson, Willis, & Mitchell-Olds, 2011; Bergelson & Roux, 2010) since adaptive variants can also generate population structure, or can be randomly correlated with population structure.

To examine the performance of population genomic methods (primarily F_{ST}-based and GEA methods), there have been a multitude of simulation studies examining different demographic scenarios and selection regimes (De Mita et al., 2013; De Villemereuil, Frichot, Bazin, François, & Gaggiotti, 2014; Forester, Lasky, Wagner, & Urban, 2018; Lotterhos & Whitlock, 2014, 2015; Perez-Figueroa, Garcia-Pereira, Saura, Rolan-Alvarez, & Caballero, 2010; Yoder & Tiffin, 2017). Some of the general findings of these studies are (a) that the underlying demographic history can have a significant impact on which approach (F_{ST} or GEA) performs better (Lotterhos & Whitlock, 2015); (b) that under certain demographic scenarios the power of all methods can be low (De Mita et al., 2013; De Villemereuil et al., 2014); (c) the intensity/timing of selection, in addition to the number of loci involved, can have a significant impact on performance (De Villemereuil et al., 2014; Forester et al., 2018; Yoder & Tiffin, 2017); and (d) the sampling of individuals and markers can have a significant impact on the power to detect causative loci (De Mita et al., 2013; Yoder & Tiffin, 2017).

While simulation studies provide a controlled environment to evaluate methodologies, and determine the factors that can affect each method, we do not know how well they emulate real data and the complex demographic and selection forces underlying them. To compensate for that, the current study compares various empirical evidence of local adaptation, recent selection, and function, to provide a general evaluation for each approach. As a study system, we use the species *Arabidopsis thaliana*, hereafter mentioned as Arabidopsis. Arabidopsis populations are found all across Eurasia (1001 Genomes Consortium, 2016), traversing very different climates (Frachon et al., 2018; Lasky et al., 2012; Mojica et al., 2016; Monroe et al., 2018), and therefore have been thought to be locally adapted. Direct evidence of local adaptation has been observed between North Sweden and Central Italy populations (Ågren & Schemske, 2012; Fournier-Level et al., 2011), in which the fitness of genotypes was measured in a reciprocal transplant experiment at the native sites of each population (Ågren & Schemske, 2012); and recombinant inbred lines of these were used to map quantitative trait loci (QTL) explaining fitness variation (Ågren et al., 2013). Such QTL, can be used to evaluate population genomic methods by testing whether candidate SNPs are enriched within their confidence intervals. Furthemore, given that most of these QTL are of low resolution and span large genomic regions, a tight association between candidate SNPs and logarithm of the odds ratio (LOD) peaks should provide stronger support for an approach. In addition to a tight association to fitness QTL, when the underlying loci exhibit fitness trade-offs (i.e., when one allele is advantageous in one environment but deleterious in another) we would expect significant allele frequency differentiation between locally adapted populations (Tiffin & Ross-Ibarra, 2014). Allele frequency differentiation is also expected under conditional neutrality (i.e., when one allele is advantageous in one environment but neutral in another) but to a lesser degree (Tiffin & Ross-Ibarra, 2014). GEAs are examined using a large sample of individuals across multiple populations and environments to identify candidate SNPs underlying local adaptation. If a large proportion of these SNPs includes causal or linked loci, then we would expect them to show significant allele frequency divergence between pairs of locally adapted populations and significant enrichment along fitness QTL peaks. Furthermore, if selection is relatively recent, candidate SNPs of both F_{ST}- and GEA-based methods are expected to be within genomic regions that exhibit high linkage disequilibrium (Charlesworth, 2006; Kim & Nielsen, 2004) and a site frequency spectrum that is expected under a recent selective sweep (Kim & Stephan, 2002; Nielsen et al., 2005).

In conjunction with population genomic signatures of selection or significant associations with climate, genetic variation underlying local adaptation is expected to be enriched along sites that are functional and influence fitness. SNPs showing significant associations with climate were found to be significantly enriched among nonsynonymous, but also synonymous variation (Hancock, Brachi, et al., 2011; Lasky et al., 2012). The enrichment among synonymous variation (which largely evolve neutrally) may be the result of linkage disequilibrium due to neutral processes but also background

selection (Charlesworth, Morgan, & Charlesworth, 1993) and/or hitchhiking (Gillespie, 2000). A stricter enrichment test will be one that controls for sequence conservation along coding and noncoding sites. Sites that are highly conserved among species (Haudry et al., 2013; Hupalo & Kern, 2013; Miller et al., 2007) are assumed to be under functional constraint and selectively important—that is, due to purifying selection the number of tolerated mutations is limited (Graur, 2016). Therefore, SNPs showing significant evidence of local adaptation across highly constraint sites are more likely to be true positives.

To provide an evaluation of the approaches frequently used to study the genetic basis of local adaptation, we include two methods that identify genotype-by-environment associations while accounting for population structure [genome-wide efficient mixed model (GEMMA; Zhou & Stephens, 2012) and the latent factor mixed model (LFMM; Caye et al., 2019)] and a method [BAYESCAN (Foll & Gaggiotti, 2008)] that identifies SNPs showing higher allele frequency differentiation than expected under various neutral models of evolution. The two GEA methods (GEMMA and LFMM) were applied across a set of 875 Eurasian accessions (1001 Genomes Consortium, 2016) and four climate variables covering temperature, precipitation and photosynthetically active radiation (Lasky et al., 2012; Price et al., 2018). The 875 accessions excluded likely laboratory escapees or contaminants (Pisupati et al., 2017) and invasive lines that may reduce the signal of local adaptation (Lasky et al., 2012). BAYESCAN on the other hand was applied to a sample of accessions in North Sweden and South Italy.

After obtaining the results, we addressed the following questions: What is the overlap in significant SNPs identified by all methods? Do they show significant allele frequency differentiation and linkage disequilibrium between locally adapted Sweden and Italy populations? What are the derived allele frequency spectra and levels of linkage of disequilibrium of these SNPs across the whole Eurasian range? Do SNPs identified by any of the methods show a significant association/enrichment with and along fitness QTL peaks? Finally, do any of these SNPs show enrichment along cis-regulatory/nonsynonymous sites showing significant functional constraint?

Addressing the above questions allowed us to assess these methods and consider the best approach to examine candidate genetic variation underlying 20 fitness QTL and evidence tying flowering time to local adaptation. Flowering time is a life-history trait that is thought to play a significant role in local adaptation to climate (Ågren et al., 2017); Dittmar, Oakley, Agren, & Schemske, 2014; Hall & Willis, 2006; Sandring & Agren, 2009; Verhoeven, Poorter, Nevo, & Biere, 2008), and whose genetic basis has been thoroughly studied (Salomé et al., 2011; Sasaki et al., 2015). To re-examine evidence linking flowering time to climate adaptation, we used the following data: (a) a list of genes that were experimentally shown to affect flowering time; (b) high-confidence QTL explaining flowering time variation between Italy and Sweden populations (Ågren et al., 2017), and (c) flowering time estimates for Arabidopsis Eurasian accessions (1001 Genomes Consortium, 2016).

2 | MATERIALS AND METHODS

2.1 | Identifying SNPs showing significant associations with climate after accounting for population structure

To identify SNPs showing significant associations with climate, while accounting for population structure we used two prominent methods: GEMMA association (Zhou & Stephens, 2012) and LFMM (Caye et al., 2019), version 2. These methods were applied to four climate variables [Minimum Temperature of Coldest Month; Precipitation during Warmest Quarter; Soil moisture; and Photosynthetically Active Radiation during Fall (which is the time of germination for fall ecotypes such as in Italy and Sweden)] important to local adaptation (Lasky et al., 2012), and a SNP genotype matrix (1001 Genomes Consortium, 2016) derived from a set of 875 re-sequenced *Arabidopsis thaliana* Eurasian accessions (Table S1) that excluded laboratory escapees/contaminants (Pisupati et al., 2017) and accessions from outside the native Eurasian and African range of *A. thaliana* that may have weaker patterns of local adaptation (Lasky et al., 2012).

After we filtered for biallelic SNPs with minor allele frequency >0.05, we tested association with home climate of ecotype and tested for potential confounding effects of population structure using the software GEMMA (Zhou & Stephens, 2012) with a missingness threshold of 0.05. For the linear mixed model option, we used the Wald test (default) to test the null hypothesis that the mean climate occupied by the two alleles is equal (Lasky et al., 2014). To apply LFMM, we first used the R function "prcomp," to perform a principal component analysis (PCA) and estimate structure in the genotypic data. "prcomp" was applied over 100 random samples of 20,000 polymorphic loci. After determining how many components (K) explained most of the genotypic variance, we applied LFMM ("lfmm_test") in R and calibrated p-values using the "gif" option. To estimate a false discovery rate for both GEMMA and LFMM, we used the "qvalue" function (Storey, 2002) implemented in R.

2.2 | Estimating allele frequency differentiation and recent selection across Arabidopsis thaliana populations

Allele frequency differentiation between 25 South Italy and 40 North Sweden accessions (Tables S2 and S3) was estimated using an F_{ST}-based metric implemented in BAYESCAN (Foll & Gaggiotti, 2008) and a simple measure of absolute allele frequency differentiation ($|f_{N.Sweden} - f_{S.Italy}|$). $|f_{N.Sweden} - f_{S.Italy}|$ was used as an alternative measure, because from prior experience (Price et al., 2018) we noticed that significant F_{ST}s were limited to SNPs in which alternative alleles were fixed between populations (>0.95).

In addition to allele frequency differentiation between Italy and Sweden populations, we also estimated derived allele frequencies across the set of 875 Eurasian accessions (DAF$_{eurasia}$) for SNPs

that we could estimate the ancestral state. To infer an ancestral base probability, we used Phast suite's Prequel function (Hubisz, Pollard, & Siepel, 2011), and a MAF genome alignment of the species *Arabidopsis thaliana* (AT) (Berardini et al., 2015), *Arabidopsis lyrata* (AL) (Hu et al., 2011), *Arabidopsis haleri* (AH) (Briskine et al., 2017), *Capsella rubella* (CR) (Slotte et al., 2013), and *Neslia paniculata* (NP). Alignments were generated with LASTZ (Harris, 2007) and refined for high-confidence orthologs using the pipeline described by (Haudry et al., 2013). In order to remove reference bias, the *A. thaliana* genome was neutralized in the alignments before calling the ancestral base. The neutral tree used for ancestral inference was ((AT:0.0640717,(AH:0.0239032,AL:0.0287045):0.029485): 0.0306753,(NP:0.0654492,CR:0.0837745):0.0306753)ANC, where ANC is the location of the simulated ancestor. Derived allele frequencies were estimated when an ancestral state had a probability >0.6.

To estimate evidence of recent selection at the SNP level, we computed linkage disequilibrium (LD) using the coefficient of correlation (r^2). More specifically, using the package "PLINK" (Purcell et al., 2007) LD at a given SNP was estimated as the mean r^2 between it and neighboring SNPs within a 20-kb window ($\overline{r^2}$). LD was measured using the 65 Italy–Sweden accessions and the 875 Eurasian accessions ($\overline{r^2_{eurasia}}$). As an additional measure of local adaptation across Italy and Sweden populations, we used SNPs that showed a high $|f_{N.Sweden} - f_{S.Italy}| > 0.70$ and LD > 0.19 (hereafter referred to as AFD.LD SNPs). 0.70 and 0.19 represent the 95th percentiles of the respective genome-wide distributions.

For evidence of recent sweeps, we used previously calculated (Price et al., 2018) composite likelihood ratios (CLRs) that were computed using Sweepfinder2 (Degiorgio et al., 2016). We focused on CLRs in North Sweden, since preciously estimated CLR signals in other populations were very weak (Huber et al., 2014; Long et al., 2013; Price et al., 2018).

2.3 | Determining sites of functional importance

To narrow down SNPs to ones that are more likely to underlie differences in function/expression of protein-coding genes, we focused on cis-regulatory and nonsynonymous variation that was found along sites showing significant selective constraint. We regarded cis-regulatory SNPs as those found within 1 kb upstream from the transcriptional start site of a gene (Pass et al., 2017; Zou et al., 2011), unless these sites were found in genic regions of other genes (in which case they were excluded). To call nonsynonymous variation, we used biallelic sites, we used a publicly available python script (callSynNonSyn.py; archived at https://github.com/kern-lab/), and gene models downloaded from the TAIR database (TAR10 genome release) (Berardini et al., 2015). To annotate regions showing significant selective constraint across the *A. thaliana* genome, we used phastCons scores (Siepel et al., 2005) derived using a nine-way alignment of Brassicaceae species from the study by (Haudry et al., 2013). We defined conserved regions as those with a score ≥0.8 over blocks of ≥10 nucleotides.

2.4 | Fitness and flowering time QTL underlying Italy and Sweden populations

Quantitative trait loci explaining fitness variation between natural *A. thaliana* Italy and Sweden populations were retrieved by the study of (Ågren et al., 2013). These 20 fitness QTL (Table S4) were assembled into 6 genetic trade-off QTL (Ågren et al., 2013); however, we treated them as independent given the very long genetic distances between fitness QTL peaks (Table S4). Furthermore, high-confidence QTL explaining flowering time variation were retrieved between these populations were retried from (Ågren et al., 2017) (Table S5).

2.5 | Estimating enrichment of candidate SNPs along functional sites and fitness QTL

To test whether SNPs showing significant evidence of local adaptation are enriched along fitness QTL peaks and among cis-regulatory/nonsynonymous variation at sites showing significant selective constraint, we used 1,000 circular permutations. For example, when examining whether SNPs showing significant correlations to climate are enriched within a given distance from fitness QTL peaks or cis-regulatory/nonsynonymous sites we execute the following steps:

1. Shifting the *q*-values along the genome while keeping the genomic positions intact. During this step, we take the whole list of *q*-values ("qvalues") and shift them circularly from a random location ("random_loc") using the following R-code: shifted_qvalues < −qvalues[c(random_loc:total_no_SNPs, 1:(random_loc − 1))]. This shifts the tail-end *q*-values to the beginning of the list and vice versa.
2. We then choose significant SNPs using our threshold for significance (FDR/*q*-value <0.1) and estimate the proportion of SNPs within a given distance from a fitness QTL peak (e.g., 100 kb) or the proportion of SNPs at cis-regulatory/nonsynonymous sites showing significant functional constraint.
3. The above two steps are repeated 1,000 times to build a distribution of expected proportions. These distributions are compared to the observed proportion.

The same approximate procedure was followed when estimating the expected proportions of SNPs showing a significant F_{ST} according to BAYESCAN or SNPs showing a high absolute allele frequency divergence and linkage disequilibrium across Italy and Sweden populations (AFD.LD).

2.6 | Sliding window analysis of chromosomal variation SNPs showing evidence of local adaptation

To detect chromosomal regions with a high proportion of SNPs showing significant evidence of local adaptation, we used a sliding window approach. Specifically, for a window size of 20kb and a step size of 1kb we estimated the ratio of SNPs showing a specific requirement (e.g., $|f_{N.Sweden} − f_{S.Italy}| > 0.70$ and LD > 0.19) over the total number of SNPs within a 20-kb window.

2.7 | Flowering time estimates for A. thaliana Eurasian accessions and candidate flowering time genes

Estimates of flowering time for the 835 Eurasian *A. thaliana* accessions were downloaded from the study by the 1001 Genomes Consortium (1001 Genomes Consortium, 2016). In brief, plants were grown in growth chambers with the following settings: after 6 days of stratification in the dark at 4°C, constant temperature of 16°C with 16 hr of light/8 hr of darkness, and 65% humidity. Flowering time was scored as days until first open flower. See 1001 Genomes Consortium (2016) for further details. A set of genes that were experimentally verified to affect flowering time was downloaded from Prof. Dr. George Coupland website (Table S6). (https://www.mpipz.mpg.de/14637/Arabidopsis_flowering_genes).

2.8 | Constructing rooted gene trees

To build neighbor-joining trees of genes showing significant local adaptation, we downloaded 1:1 orthologs between *Arabidopsis thaliana* and outgroups *Arabidopsis lyrata* and *Capsella rubella* from the Phytozome database (Goodstein et al., 2012) and after aligning the coding sequences with MAFFT (Katoh & Toh, 2008) we used MEGA (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013) to build rooted gene trees.

3 | RESULTS

3.1 | Climate-associated SNPs show low allele frequency differentiation and linkage disequilibrium across Italy and Sweden populations

To identify SNPs showing significant associations with climate while accounting for population structure, we applied GEMMA (Zhou & Stephens, 2012) and LFMM 2 (Caye et al., 2019) on 875 Eurasian accessions (Table S1) and four climate variables (Minimum Temperature of Coldest Month; Annual mean temperature; Precipitation during warmest quarter; and Photosynthetically active radiation during fall [time of germination of Italy and Sweden ecotypes (Ågren & Schemske, 2012)]. In addition to identifying SNPs showing significant associations with climate, we also applied BAYESCAN (Foll & Gaggiotti, 2008) between 40 accessions North Sweden and 25 accessions in South Italy (Tables S2 and S3) (referred to as Sweden and Italy hereafter) to identify SNPs that showed significant allele frequency differentiation estimated by F_{ST} after accounting for different models of neutral evolution.

After examining the results, we focused on Minimum Temperature of Coldest Month (Min.Tmp.Cld.M) (Figure 1a) because the other climate variables resulted in a very small number of significant SNPs [similar to a previous study (Price et al., 2018)] and in many instances did not a show an enrichment of low *p*-values when examining their *p*-value histograms. Furthermore, winter temperatures, which are approximated by Min.Tmp.Cld.M, have been shown to play an important role in local adaptation of Arabidopsis (Ågren & Schemske, 2012; Gienapp et al., 2017; Lasky et al., 2018; Oakley et al., 2014).

As depicted in appendices 1 and 2, the p-value histograms produced by GEMMA and LFMM showed an enrichment of low p-values and relatively uniform distributions at larger values. LFMM was applied using eight latent factors ($K = 8$) to account for population structure (Appendix S2). Using a $K = 8$ showed the most uniformity across large p-values (Appendix S2) and resulted in a similar number of significant SNPs with GEMMA (Figure 1b). Smaller Ks resulted in a disproportionally larger number of significant SNPs and not a very uniform distribution across large *p*-values.

When examining significant SNPs (FDR < 0.1) after filtering out any significant associations with climate that did not segregate between Italy and Sweden populations, we find a very small overlap

between methodologies (Figure 1b). The two GEA methods used (GEMMA and LFMM) showed a 25% overlap in predicted SNPs (Figure 1b), in addition to a negligible overlap with SNPs that showed significant allele frequency differentiation (F_{ST}) between Italy and Sweden populations according to BAYESCAN (Figure 1b). It is quite surprising that LFMM and GEMMA captured an extremely small proportion of SNPs that showed a significant F_{ST} between Sweden and Italy populations (Figure 1b) given the strong adaptive divergence and evidence of genetic trade-offs underlying these populations (Ågren et al., 2013; Ågren & Schemske, 2012). This, however, may occur because BAYESCAN chooses a very different set of SNPs with a high F_{ST}.

To further examine the evidence of recent selection and genetic differentiation captured by the three methods in Italy and Sweden, we looked at absolute allele frequency differentiation (AFD: $|f_{N.Sweden} - f_{S.Italy}|$) and linkage disequilibrium (LD) estimated as the SNP level (r^2 (20 kb)—the average r^2 between a SNP and its neighboring SNPs within 20 kb). These measures were examined across different false discovery rates (FDRs), in order to study the relation between these signatures of adaptation and FDR (Figure 1c,d).

As expected, SNPs showing a significant F_{ST} (FDR < 0.1) according to BAYESCAN also showed a very high $|f_{N.Sweden} - f_{S.Italy}|$
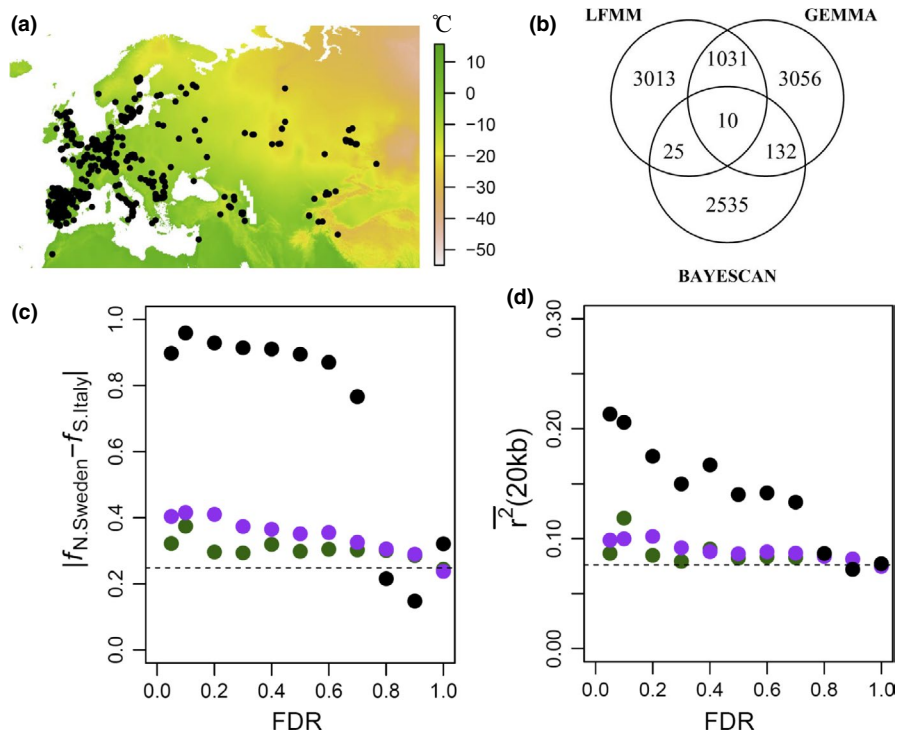

FIGURE 1 Low genetic divergence and recent selection underlying climate-associated SNPs in locally adapted Italy and Sweden populations (Ågren & Schemske, 2012; Ågren et al., 2013). (a) Map depicting Minimum Temperature of Coldest Month (Min.Tmp.Cld.M) across the native range of 875 re-sequenced *Arabidopsis thaliana* accessions (black dots). (b) A Venn diagram depicting the overlap in SNPs that showed significant associations with Min.Tmp.Cld.M according to LFMM and GEMMA, in addition to SNPs that showed significant allele frequency differentiation (F_{ST}) according to BAYESCAN. Significance was set at an FDR < 0.1, and climate-associated SNPs were filtered for ones that segregated between Italy and Sweden populations. (c) Examining the trend between absolute allele frequency divergence (AFD: $|f_{N.Sweden} - f_{S.Italy}|$) and false discovery rate (FDR) when using GEMMA (green dots); LFMM (purple dots); and BAYESCAN (black dots). The first two points are separated by an interval of 0.05 and the rest by an interval of 0.1. The dotted line depicts the genome average. (d) Similarly, to Figure 1c, this figure examines linkage disequilibrium (LD) across Italy and Sweden populations. LD was estimated at the SNP level by taking the average r^2 between a SNP and all other SNPs within a 20-kb window [r^2 (20 kb)]

—WILEY

(Figure 1c). A high $|f_{N.Sweden} - f_{S.Italy}|$ was also observed across SNPs that showed a higher FDR (Figure 1c), but that significantly dropped at an FDR > 0.7, where 98.5% of the SNPs were found. On the contrary, $|f_{N.Sweden} - f_{S.Italy}|$ associated with GEMMA and LFMM SNPs was low and near the genomic average across all FDRs (Figure 1c). The same pattern was observed when examining LD, where climate-associated SNPs showed a very low LD that was repeated across all FDRs (Figure 1d), while SNPs showing a significant F_{ST} according to BAYESCAN showed a significantly higher LD than the genome average (Figure 1d), and as expected, LD decreased as FDR increased. All in all, these results indicate that climate-associated SNPs capture very little evidence of local adaptation and recent selection across Italy and Sweden populations.

3.2 | Derived allele frequencies, linkage disequilibrium, and climatic differentiation, across Eurasian populations, significantly differ among candidate SNPs

The weak evidence of local adaptation observed across climate-associated SNPs may occur because they capture selection in other parts of the species range. To examine this possibility, we looked at their derived allele frequencies across the Eurasian range ($DAF_{eurasia}$), in association with linkage disequilibrium ($\overline{r^2_{eurasia}}$ (20 kb). Furthermore, we compared $DAF_{eurasia}$ and $\overline{r^2_{eurasia}}$ between climate-associated SNPs and SNPs that showed significant allele frequency differentiation and LD between Italy and Sweden populations. More specifically, the SNPs included were (a) all significant (FDR < 0.1) LFMM and GEMMA SNPs (i.e., unlike Figure 1b, we included SNPs that did not segregate between Italy and Sweden populations; (b) SNPs that exhibited significantly high F_{ST}s between Italy and Sweden populations according to BAYESCAN; and (c) SNPs that showed a high AFD and LD between Italy and Sweden populations (hereafter referred to as "AFD.LD"). This latter measure was included because 92% BAYESCAN SNPs showed an $|f_{N.Sweden} - f_{S.Italy}| \approx 1$, and we wanted to include SNPs that showed lower divergences and included high levels of linkage disequilibrium [comparisons of a similar measure of AFD and F_{ST} are discussed in Ref. (Berner, 2019)]. As a threshold for high, we used the 95th percentiles of the $|f_{N.Sweden} - f_{S.Italy}|$ and LD distributions ($|f_{N.Sweden} - f_{S.Italy}| > 0.70$ & LD > 0.19).

After estimating the ancestral nucleotide states across the *A. thaliana* genome (explained in Materials and Methods), we estimated $DAF_{eurasia}$ for a subset of SNPs: LFMM (2,438), GEMMA (1,899), BAYESCAN (502), and *AFD.LD* (9,941). As shown in Figure 2a, compared to the DAF across the genome ("Genome"), GEMMA and LFMM are significantly enriched at low-derived frequency variants (0.05 ≤ $DAF_{eurasia}$ <0.1—SNPs with a MAF < 0.05 were filtered out before examination). Contrary to LFMM and GEMMA, BAYESCAN and *AFD.LD* SNPs showed a significant depletion in low-frequency variants and a higher proportion DAFs between 0.2–0.4 and 0.6–0.8 (Figure 2a). These results indicate that SNPs showing significant allele differentiation and LD between

Italy and Sweden populations show higher DAFs across the whole Eurasian range, when compared to SNPs showing significant associations with climate.

In addition to DAFs, AFD.LD and BAYESCAN SNPs showed a higher LD across Eurasian populations [$\overline{r^2_{eurasia}}$ (20 kb)] than climate-associated SNPs (Figure 2b). Compared to the genome-wide average LD (~0.12) and the 95th percentile (~0.34), the average LDs across the four sets of SNPs were BAYESCAN (~0.27); AFD.LD (~0.30); LFMM (~0.13); and GEMMA (~0.13).

Using the bins in which each set of SNPs showed the highest difference when compared to the genome average (Figure 2a), we found that low-frequency GEMMA and LFMM variants (green and purple points) were mostly found in North Sweden, Russia, and other parts of central Asia (Figure 2c). On the other hand, AFD.LD (blue) and BAYESCAN (black) SNPs with DAFs between 0.2–0.3 and 0.3–0.4 showed the highest frequency in North Sweden (Figure 2c), while SNPs with DAFs between 0.7–0.8 and 0.6–0.7 showed a depletion in North Sweden (Figure 2c) and higher frequencies in central Europe (Appendix S4). In addition to differences in location, significant LFMM and GEMMA alleles showed much higher differences in mean Min.Tmp.Cld.M (~10°C) than AFD.LD and BAYESCAN alleles (~3.6°C) (Figure 2d). This is not surprising since GEA methods identify SNPs that explain more climate variation than the genome average.

3.3 | High AFD and LD SNPs show a strong association with fitness QTL peaks and an enrichment at cis-regulatory and nonsynonymous sites showing significant selective constraint

As mentioned earlier, alleles underlying local adaptation are expected to be enriched along sites that show significant evidence of function, but most importantly along QTL explaining fitness variation between populations (which provide direct evidence of local adaptation). To test the above, we looked at the distribution of SNPs across nonsynonymous/cis-regulatory sites showing significant selective/functional constraint among Brassicaceae plants (Haudry et al., 2013) (phastCons > 0.8) and QTL explaining fitness variation between locally adapted Italy and Sweden populations (Ågren et al., 2013; Ågren & Schemske, 2012).

To examine direct evidence of local adaptation underlying SNPs identified by the four approaches, we first examined the proportion of SNPs at different distances from 20 QTL peaks (i.e., markers showing highest LOD score). Appendix S4 shows how the proportion of SNPs changes with distance from QTL peaks. Under the assumption that the proportion of SNPs showing significant evidence of local adaptation should be highest near LOD peaks, and decrease with distance, only LFMM and AFD.LD SNPs showed a negative association with distance (Appendix S4). The association of LFMM SNPs, however, was weaker (R^2 = .04, p-value = .56) in comparison with AFD.LD SNPs (R^2 = .2, p-value = .2), in addition to the proportion of AFD.LD SNPs being twice as high than LFMM SNPs
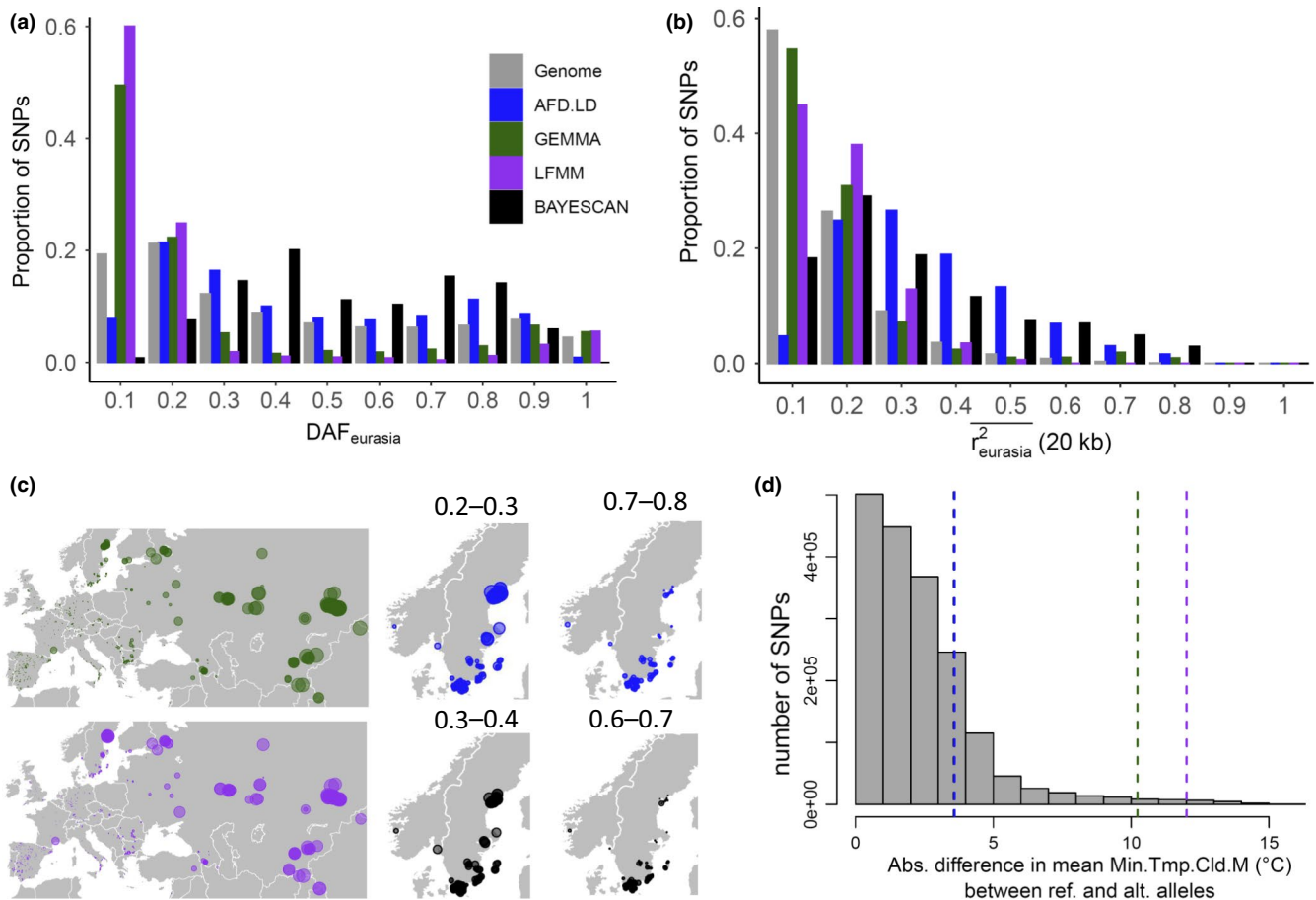
FIGURE 2 Candidate SNPs associated with local adaptation show contrasting patterns in genetic differentiation, selection, and climatic properties across Eurasia. (a) Comparing the derived allele frequency spectra (DAF$_{eurasia}$) of candidate SNPs underlying local adaptation (GEMMA, LFMM, BAYESCAN, AFD.LD) and a genome-wide set (Genome). (b) Comparing linkage disequilibrium [$r^2_{eurasia}$ (20 kb)] across the different classes of SNPs. (c) Examining the location of candidate SNPs with DAFs that showed the highest enrichment relative to the genome-wide set of SNPs (Figure 2a): AFD.LD (02–0.3 and 0.7–0.8); BAYESCAN (0.3–0.4 and 0.6–0.7); LFMM (<0.1); and GEMMA (<0.1). The size of the circles corresponds to the frequency of derived alleles at a specific location. (d) Absolute difference in mean Min.Tmp.Cld.M between reference and alternative alleles of candidate SNPs. AFD.LD and BAYESCAN SNPs show a smaller difference in Min.Tmp.Cld.M (~3.6°C) than GEMMA and LFMM SNPs (~10°C)

at a close distance from QTL peaks (0–100 kb) (Appendix S4). The nonsignificant negative association observed across AFD.LD SNPs could be caused by the small number of points (n = 10) (Appendix S4). In addition to the strongest association, the proportion of AFD.LD SNPs near QTL peaks was significantly (>95% percentile) higher than expected by chance (Figure 3c). The second largest proportion relative to expectation was observed across LFMM SNPs (Figure 3c). LFMM SNPs within 100 kb of fitness QTL peaks showed a much lower mean AFD (~0.36) and LD (~0.10) than AFD.LD SNPs, which by definition are high in AFD and LD (AFD > 0.70 & LD > 0.19). This is not surprising given the low mean AFD and LD across all significant LFMM SNPs (Figure 1c,d).

SNPs that were significant according to BAYESCAN, showed a high AFD and LD (Figure 1c,d), but nonetheless they did not show a strong association (Appendix S4) and significant enrichment along fitness QTL, such as, high AFD.LD SNPs (Figure 3c). This may occur because significant BAYESCAN SNPs cannot capture all selection events underlying fitness QTL.

To examine where AFD.LD and BAYESCAN differed along genetic trade-off QTL that were built using 20 fitness QTL (Ågren et al., 2013), we scanned for regions showing a high proportion of SNPs that were common between the two sets (AFD.LD and BAYESCAN) and SNPs that were unique to the AFD.LD set. As shown in Appendix S5, genetic trade-off QTL that included regions that show high composite likelihood ratios for recent sweeps we see a high proportion of SNPs that are common to both approaches; on the other hand, along genetic trade-off QTL where evidence of recent sweeps is almost absent (Appenices S6 and S7) we only observe SNPs that were unique to the AFD.LD set.

Finally, we examined the distribution of candidate SNPs underlying local adaptation across cis-regulatory/nonsynonymous sites showing significant selective constraint. As shown in Figure 3b,c, the observed proportions of GEMMA SNPs were within the random expectation, while the proportions of LFMM and AFD.LD SNPs were higher than the expectation at both conserved cis-regulatory and nonsynonymous sites. Finally, BAYESCAN SNPs were enriched

only along conserved nonsynonymous sites (Figure 3a). For methods that showed an enrichment at conserved sites (cis-regulatory and/or nonsynonymous sites), average LD in Eurasia ($\overline{r^2_{eurasia}}$) for BAYESCAN, LFMM, and AFD.LD SNPs was 0.27, 0.12, and 0.29, respectively. The average LD across LFMM SNPs was approximately the same as the genome-wide mean at conserved cis-regulatory and nonsynonymous sites (~0.12), while for BAYESCAN/AFD.LD SNPs was twice as high. Similarly, to the whole set of SNPs (Figure 2a), LFMM SNPs also showed a low mean DAF$_{eurasia}$ (~0.13).

3.4 | Genes underlying fitness QTL and controlling flowering time show significant evidence of local adaptation

Given the significant evidence of local adaptation and function underlying AFD.LD SNPs, we used them to detect potential genes that may underlie fitness QTL (note: We only focused on variants within 100 kb of their peaks), in addition to examining evidence of local adaptation underlying a list of ~170 genes affecting flowering time (Table S6). To further narrow down on SNPs that are more likely to underlie the fitness QTL examined, we only considered cis-regulatory/nonsynonymous variation at conserved sites that segregated between the parents used to derive the RILs (Ågren et al., 2013).

SNPs between the parental genomes were called in a previous study (Price et al., 2018).

Our analysis resulted in 24 genes within 100 kb of fitness QTL peaks and spanning three genetic trade-off QTL (2:2, 4:2, and 5:5) (Appendix S8). Many of these were involved in interesting biological processes such as response to different abiotic stress factors and the abscisic-acid signaling pathway which is important in abiotic stress response (Tuteja, 2007) (Appendix S8). Among these genes, two of them (AT4G33360 (*FLDH*) and AT4G33470 (*HDA14*)) showed strong expression G×E interactions [G×E interactions were identified in a previous study (Price et al., 2018)] when Italy and Sweden plants were grown under cold acclimation conditions (4°C) for two weeks (Gehan et al., 2015). Interestingly, *FLDH* is a negative regulator of the abscisic-acid signaling pathway (Bhandari, Fitzpatrick, & Crowell, 2010). As shown in Figure 4, this gene was within a region of a genetic trade-off QTL that showed a high proportion of AFD. LD SNPs. Furthermore, expression of *FLDH* under control and cold acclimation conditions was significantly lower in Sweden than Italy plants (Figure 4).

To examine whether genes affecting flowering time (Table S6) show significant evidence of local adaptation, we tested whether the number of such genes containing cis-regulatory and/or nonsynonymous SNPs with a high AFD (Figure 5a), and a high AFD and LD (AFD.LD) (Figure 5b) was significantly higher than expected by
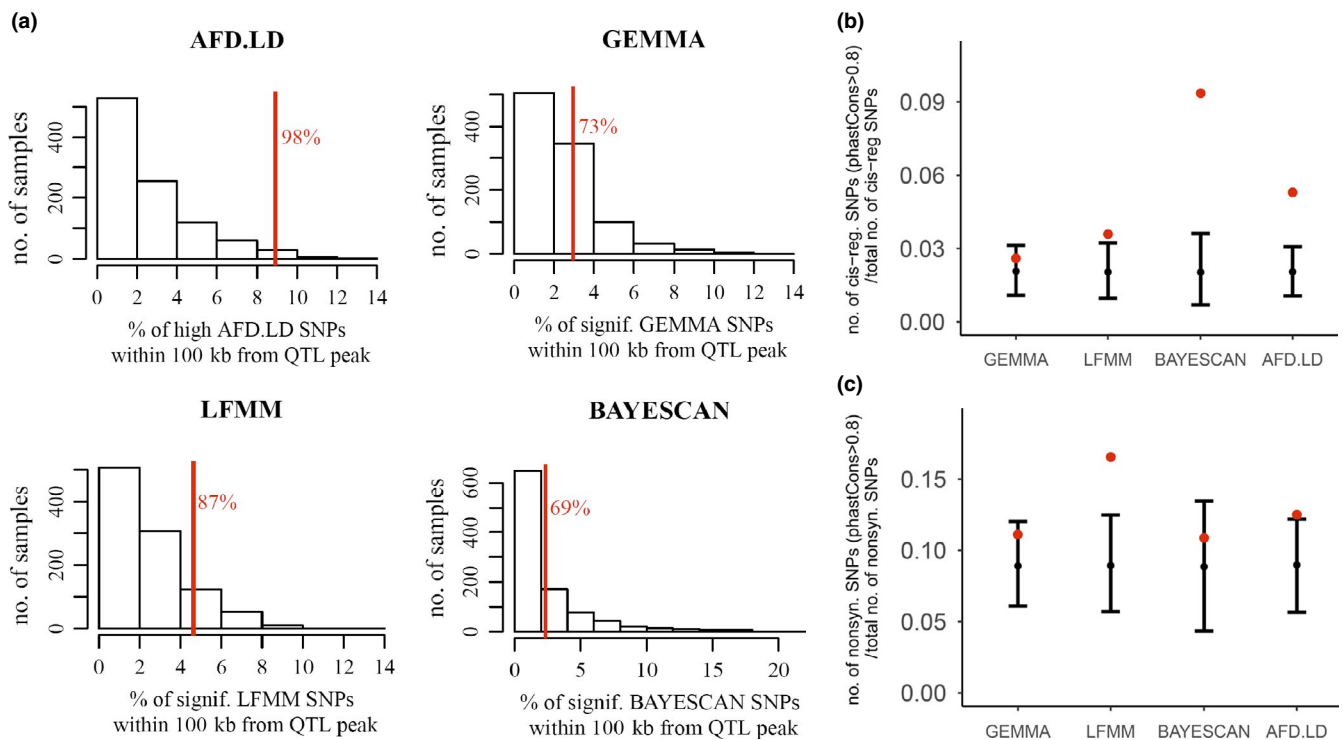

FIGURE 3 Testing for enrichment of candidate SNPs along QTL peaks explaining fitness variation between Italy and Sweden populations (Ågren et al., 2013) and cis-regulatory/nonsynonymous sites showing significant selective constraint across nine Brassicaceae species (Haudry et al., 2013). (a) Comparing the observed proportion of candidate SNPs (red lines) within 100 kb of fitness QTL peaks, to a distribution of expected proportions derived using circular permutations. Significance was set at the 95th percentile. (b) The observed proportion of candidate SNPs (red dots) along cis-regulatory sites showing significant selective constraint in relation to the expected proportion (in black are 95% intervals estimated using circular permutations). (c) The observed proportion of SNPs along nonsynonymous sites showing significant selective constraint
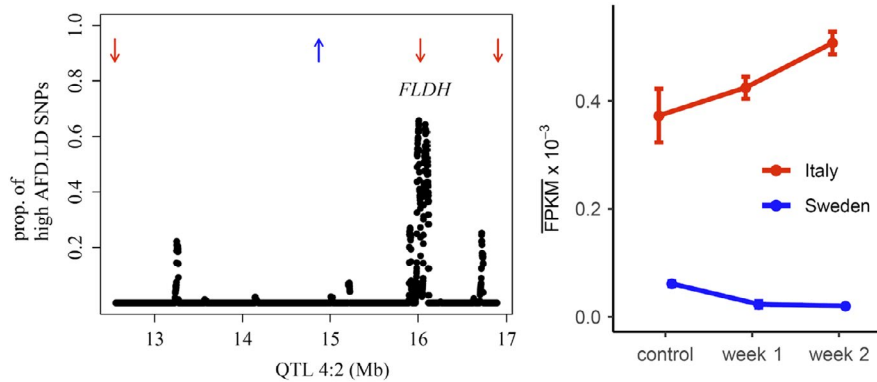
FIGURE 4 Significant evidence of local adaptation underlying the *FLDH* gene. *FLDH* is a negative regulator of ABA (Bhandari et al., 2010), found within genetic trade-off QTL 4:2 (Ågren et al., 2013) and 100 kb from a fitness QTL peak (red and blue arrows represent QTL where the Sweden genotype had lower fitness in Italy and higher fitness in Sweden, respectively). The region including *FLDH* showed a high proportion of high AFD and LD (AFD.LD) SNPs, and furthermore, in Italy and Sweden plants *FLDH* showed strong expression GxE interactions under control (22°C) and cold acclimation conditions (4°C) for two weeks (FPKM: fragments per kilobase million)

chance (>95%). As shown in Figure 5a,b, the observed number of genes is significantly higher than the expectation. Among the 12 genes with high AFD and LD SNPs, we identified three [AT1G09530 (*PIF3*), AT2G21070 (*FIO1*), and AT5G57660 (*COL5*)] that contained such SNPs along conserved nonsynonymous sites. Among the three genes, *PIF3* was found along a chromosomal region that showed the highest CLR for a recent sweep in Sweden and a high density of AFD. LD SNPs (Figure 5c). Eurasian accessions sharing a similar allele as the Sweden parent showed longer flowering time than accessions sharing the same allele as the Italy parent (Figure 5c). The same pattern was observed when examining *COL5* (Figure 5d), a flowering time gene which was also found within a flowering time QTL (FlrT-5:4, Table S5). According to FlrT-5:4, the Sweden genotype was associated with longer flowering time in both Italy and Sweden (Ågren et al., 2017). In conjunction, with its overlap to a genetic trade-off QTL (Ågren et al., 2017), it indicates a possible role in fitness trade-offs. Studies have attributed flowering time variation within FlrT-5:4 to *VIN3* (Ågren et al., 2017; 1001 Genomes Consortium, 2016). Although it may be an additional candidate, we did not find any significant genetic differentiation and selection along coding and cis-regulatory sites of *VIN3*.

When examining flowering time genes with high *AFD* and LD SNPs along cis-regulatory/nonsynonymous sites that did not show significant selective constraint, we identified an additional nine genes, four of which were found within flowering time QTL (FlrT): AT1G14920 (*GAI*); AT1G53090 (*SPA4*); AT1G79460 (*GA2*); AT2G22540 (*SVP*); AT2G28550 (*RAP2.7*); AT2G47700 (*RFI2*); AT4G32980 (*ATH1*-FlrT4:1); AT5G24470 (*PRR5*-FlrT5:2); and AT5G65060 (*MAF3*-FlrT5:5). ATH1 was found in genetic trade-off QTL 4:2, while gene MAF3 was found within genetic trade-off QTL 5:5 and within 100 kb of fitness QTL peaks.

4 | DISCUSSION

In the quest to study the genetic basis of local adaptation using genome-wide associations with environment, linear mixed models have

emerged as a powerful tool given their ability to account for population structure while testing for significant associations (Caye et al., 2019; Kang et al., 2010, 2008; Yu et al., 2006; Zhou & Stephens, 2012). Although they provide a robust statistical framework, the current study shows that such approaches may significantly limit our ability to understand the polygenic basis of local adaptation.

Both GEA methods (GEMMA and LFMM) resulted in SNPs that showed poor associations with QTL explaining fitness variation, in addition to low genetic differentiation and evidence of recent selection across locally adapted populations. The poor performance of GEA methods when examining populations isolated by distance has also been shown when using simulations (Lotterhos & Whitlock, 2015). In fact, F_{ST}-based approaches outperformed GEA methods in such instances (Lotterhos & Whitlock, 2015). In the current study, we show that SNPs exhibiting a significantly high F_{ST} according to BAYESCAN, capture higher population genomic evidence of recent selection but fail to show a strong association with fitness QTL and a significant enrichment along regions with a high LOD score. Using a more lenient FDR cutoff may capture some of the missing SNPs across fitness QTL that do not contain strong CLRs for recent sweeps (Appenices S6 and S7).

SNPs that show extreme AFD and LD are likely to contain many false positives, especially if no other information is considered. These genomic signatures, however, show a promising future in identifying recent local adaptation within a statistical framework (Kemppainen et al., 2015). Such an approach can avoid several complications of GEA methods, including (a) errors in measures of environmental variables; (b) an increase in false positives if more than one environmental variable is needed to capture the genetic basis of local adaptation; and (c) difficulties in choosing the correct minimum allele frequency to avoid spurious associations.

Arabidopsis thaliana, however, is a simple, highly inbred species, and the populations we examined are separated by a very large geographic distance. GEA methods have been suggested to perform best across populations that do not exhibit a hierarchical population structure and isolated by large distances (De Villemereuil et al.,
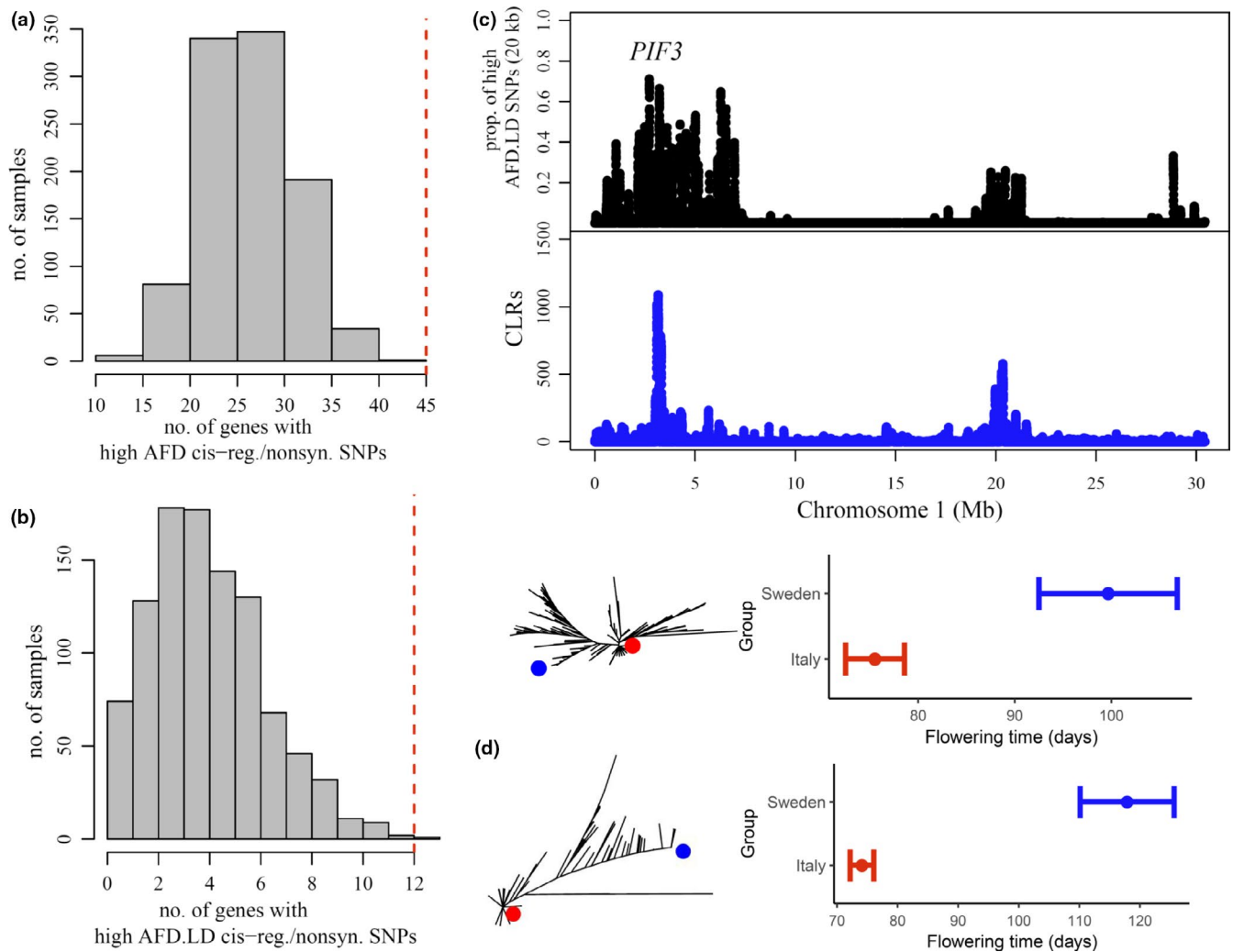
FIGURE 5 Genes known to affect flowering time show significant evidence of local adaptation along putative functional sites. (a) The number of flowering time genes containing cis-regulatory/nonsynonymous SNPs showing a high AFD was significantly higher (>95th percentile) than expected by chance. Distribution of random numbers was derived using circular permutations. (b) The number of flowering time genes with a high AFD and LD cis-regulatory/nonsynonymous SNPs was also significantly higher than expected by chance. (c) *PIF3* is a phytochrome interacting factor that has been found to affect flowering time (Oda et al., 2004) that was found underlying a region along chromosome 1 that showed the largest composite likelihood ratios (CLRs) for recent sweeps in Sweden, and windows with a high proportion of AFD.LD SNPs. A rooted phylogeny of the PIF3 coding region indicated that Eurasian accessions sharing the same allele as the Sweden parent (blue dot) show significantly higher flowering time than accessions sharing the same allele as the Italy parent (red dot). (d) *COL5* is another gene that has been found to affect flowering time (Hassidim, Harir, Yakir, Kron, & Green, 2009) and in which Eurasian accessions show significant genetic differentiation and segregation in flowering time. This gene is also found within previously identified flowering time QTL (FlrT-5:4) (Ågren et al., 2017) in which the Sweden genotype was associated with longer flowering time in both Italy and Sweden

2014; Lotterhos & Whitlock, 2015). An example of such populations is ones found in Northern and Southern Sweden, which follow a two-population island model (Huber et al., 2014). Examining the relation between SNPs showing significant associations with climate and other population genomic or field-based evidence of local adaptation can shed further light on the ability of GEA methods under such scenarios. Furthermore, such analyses can be expanded to include species with different life-history traits, such as the perennial European aspen, *Populus tremula* (Ingvarsson & Bernhardsson, 2019; Wang et al., 2018).

Despite the poor evidence of local adaptation and selection, SNPs identified by the GEA method LFMM showed an enrichment

along sites showing significant selective constraint. Given the significant enrichment of such SNPs among low-frequency-derived alleles that show poor direct and indirect evidence of local adaptation, we raise caution when interpreting such results. Enrichment of climate-associated SNPs along nonsynonymous sites (Hancock, Brachi, et al., 2011; Hancock, Witonsky, et al., 2011; Lasky et al., 2012) or the parallel occurrence of low-frequency loss-of-function mutations showing significant associations with climate (Monroe et al., 2016, 2018) has been interpreted as evidence of adaptation. Although such signals may represent instances of adaptation, they can also be explained by neutral evolution, in which relaxed selection across specific climates results in the enrichment of

independent loss-of-function mutations or nonsynonymous variation (Flowers, Hanzawa, Hall, Moore, & Purugganan, 2009; Zhen, Dhakal, & Ungerer, 2011; Zhen & Ungerer, 2008). In the context of local adaptation, these may represent instances of conditional neutrality, where in one environment expressing the gene has no significant impact on fitness. Some ways that could provide further support as to whether a recent loss-of-function mutation is adaptive are to compare LD or extended haplotype homozygosity (Sabeti et al., 2002) between individuals that have a loss-of-function mutation and ones that do not.

Among the genomic signatures of local adaptation examined, SNPs showing a high absolute allele frequency differentiation (AFD) and linkage disequilibrium (LD) between Italy and Sweden populations showed the strongest evidence of local adaptation and were enriched among nonsynonymous/cis-regulatory variation at sites showing significant selective constraint. Using these SNPs, we identified a list of candidate genes underlying fitness QTL. One of these was *FLDH*, a negative regulator of abscisic-acid signaling (Bhandari et al., 2010), that showed strong GxE interactions between Italy and Sweden plants under cold acclimation conditions. Abscisic-acid signaling is known to play an important role in abiotic stress response (Tuteja, 2007), with many studies supporting a key role in local adaptation to climate (Kalladan et al., 2017; Keller, Levsen, Olson, & Tiffin, 2012; Lasky et al., 2014; Ristova, Giovannetti, Metesch, & Busch, 2018). In addition to abscisic-acid signaling, our study provides further support for the important role of flowering time in local adaptation to climate. Among a list of genes that were experimentally shown to affect flowering time, we identified three genes (*PIF3*, *FIO1*, and *COL5*) that showed significant evidence of local adaptation and selective constraint along nonsynonymous sites. *FIO1* was previously shown to contain SNPs that showed significant associations with flowering time among natural Swedish lines (Sasaki, Zhang, Atwell, Meng, & Nordborg, 2015) and *COL5* was located within a QTL that explain flowering time variation among Sweden and Italy recombinant inbred lines (Ågren et al., 2017). Finally, *PIF3*, a transcription factor that interacts with phytochromes (Soy et al., 2012), has been implicated in multiple biological processes including early hypocotyl growth (Monte et al., 2004), photomorphogenesis (Dong et al., 2017), flowering time (Oda, Fujiwara, Kamada, Coupland, & Mizoguchi, 2004), and regulation of physiological responses to temperature (Jiang et al., 2017).

Interestingly, *PIF3* was found within a large region that showed significant evidence of local adaptation. Regions of high divergence may involve a single causative variant, or a group of linked genes that interact with *PIF3* and were under selection because they contributed to building an advantageous phenotype (Barton & Bengtsson, 1986; Yeaman & Whitlock, 2011). Although such "Islands of high divergence" can be the result of local adaptation, they can also be formed through nonadaptive processes (Pennisi, 2014), and therefore, caution should be exercised when drawing any conclusions.

When ignoring selective constraint, we identify a list of addition flowering time genes showing significant evidence of local adaptation along nonsynonymous/cis-regulatory sites. Genes such as SVP

and MAF3 were previously associated with flowering time variation among natural *Arabidopsis* accessions (Caicedo, Richards, Ehrenreich, & Purugganan, 2009; Sasaki et al., 2015), and MAF3 showed strong allelic variation along a multivariate climate gradient (Lasky et al., 2012). Although adaptation may involve sites that are not deeply rooted and/or under strong selective constraint, including additional plant genomes when estimating sequence conservation across species may increase our power to detect selectively important regions. As shown by studies examining adaptation in species ranging from bacteria (Maddamsetti et al., 2017) to birds (Sackton et al., 2019), addressing selective constraint can improve our understanding of its genetic basis.

To sum up, the current study identifies candidate genes and life-history traits that may underlie adaptation of Arabidopsis populations to local environments. Furthermore, it shows that understanding organismal adaptation to local environments is a very complex venture, where several lines of evidence are needed to obtain a comprehensive and well-supported picture.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

AUTHOR CONTRIBUTIONS

Nicholas Price designed research, performed research, analyzed data, and wrote the paper. Lua Lopez and Adrian E. Platts performed research and reviewed the paper. Jesse R. Lasky reviewed the paper.

DATA AVAILABILITY STATEMENT

In our dryad submission, we included:

1. The false discovery rate (FDR) associated with SNPs used to identify significant associations with Minimum Temperature of Coldest Month using GEMMA and LFMM
2. The FDR associated with SNPs used to identify instances of significant allele frequency differentiation (approximated by F_{ST}) between Italy and Sweden populations using BAYESCAN.
3. Absolute nonreference allele frequency differentiation (AFD) estimated between Italy and Sweden populations
4. Derived allele frequency divergence (DAF) across 875 Eurasian accessions
5. Linkage disequilibrium estimated using 65 Italy and Sweden accessions and 875 Eurasian accessions
6. Composite likelihood ratios (CLRs) for recent sweeps in North Sweden populations
7. Climate data for all the 1,135 *A. thaliana* re-sequenced genomes that have been recently published (1001 Genomes Consortium, 2016).
8. Conserved coding regions
9. Conserved noncoding regions

Private link: https://datadryad.org/stash/share/962vtPUZ9tTuwtr
RKA1px_WegXVYYEYsvxAWa_6lf_0

ORCID

Nicholas Price (iD) https://orcid.org/0000-0001-6672-2952

REFERENCES

1001 Genomes Consortium. (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

Ågren, J., Oakley, C. G., Lundemo, S., & Schemske, D. W. (2017). Adaptive divergence in flowering time among natural populations of Arabidopsis thaliana: Estimates of selection and QTL mapping. *Evolution, 71*, 550–564.

Ågren, J., Oakley, C. G., McKay, J. K., Lovell, J. T., & Schemske, D. W. (2013). Genetic mapping of adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 21077–21082.

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*, 1112–1122.

Anderson, J. T., Lee, C. R., Rushworth, C. A., Colautti, R. I., & Mitchell-Olds, T. (2013). Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology, 22*, 699–708.

Anderson, J. T., Willis, J. H., & Mitchell-Olds, T. (2011). Evolutionary genetics of plant adaptation. *Trends in Genetics, 27*, 258–266.

Arguello, J. R., Cardoso-Moreira, M., Grenier, J. K., Gottipati, S., Clark, A. G., & Benton, R. (2016). Extensive local adaptation within the chemosensory system following Drosophila melanogaster's global expansion. *Nature Communications, 7*, ncomms11855.

Barton, N., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity (Edinb), 57*(Pt 3), 357–376.

Beaumont, M. A., & Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology, 13*, 969–980.

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis, 53*, 474–485.

Bergelson, J., & Roux, F. (2010). Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nature Reviews Genetics, 11*, 867–879.

Berner, D. (2019). Allele frequency difference AFD(-)An intuitive alternative to F_{ST} for quantifying genetic population differentiation. *Genes (Basel), 10*.

Bhandari, J., Fitzpatrick, A. H., & Crowell, D. N. (2010). Identification of a novel abscisic acid-regulated farnesol dehydrogenase from Arabidopsis. *Plant Physiology, 154*, 1116–1127.

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*, 2044–2072.

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., & Sancristobal, M. (2010). Detecting selection in population trees: The Lewontin and Krakauer test extended. *Genetics, 186*, 241–262. https://doi.org/10.1534/genetics.110.117275

Briskine, R. V., Paape, T., Shimizu-Inatsugi, R., Nishiyama, T., Akama, S., Sese, J., & Shimizu, K. K. (2017). Genome assembly and annotation of *Arabidopsis halleri*, a model for heavy metal hyperaccumulation and evolutionary ecology. *Molecular Ecology Resources, 17*, 1025–1036.

Caicedo, A. L., Richards, C., Ehrenreich, I. M., & Purugganan, M. D. (2009). Complex rearrangements lead to novel chimeric gene fusion polymorphisms at the *Arabidopsis thaliana* MAF2-5 flowering time gene cluster. *Molecular Biology and Evolution, 26*, 699–711.

Caye, K., Jumentier, B., Lepeule, J., & Francois, O. (2019). LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Molecular Biology and Evolution, 36*, 852–860.

Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics, 134*, 1289–1303.

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics, 2*, e64.

Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D., & Kingsley, D. M. (2004). The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biology, 2*, E109.

Conover, D. O., Duffy, T. A., & Hice, L. A. (2009). The covariance between genetic and environmental influences across ecological gradients: Reassessing the evolutionary significance of countergradient and cogradient variation. *Annals of the New York Academy of Sciences, 1168*, 100–129.

De Mita, S., Thuillet, A. C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: Analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology, 22*, 1383–1399.

De Villemereuil, P., Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models: When and how much should we trust them? *Molecular Ecology, 23*, 2006–2019.

De Villemereuil, P., & Gaggiotti, O. E. (2015). A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution, 6*, 1248–1258.

Degiorgio, M., Huber, C. D., Hubisz, M. J., Hellmann, I., & Nielsen, R. (2016). SweepFinder2: Increased sensitivity, robustness and flexibility. *Bioinformatics, 32*, 1895–1897.

Dittmar, E. L., Oakley, C. G., Agren, J., & Schemske, D. W. (2014). Flowering time QTL in natural populations of Arabidopsis thaliana and implications for their adaptive value. *Mol Ecol, 23*, 4291–4303.

Dong, J., Ni, W., Yu, R., Deng, X. W., Chen, H., & Wei, N. (2017). Light-dependent degradation of PIF3 by SCFEBF1/2 promotes a photomorphogenic response in Arabidopsis. *Current Biology, 27*, 2420–2430.e6.

Fischer, M. C., Rellstab, C., Tedder, A., Zoller, S., Gugerli, F., Shimizu, K. K., … Widmer, A. (2013). Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Molecular Ecology, 22*, 5594–5607.

Flowers, J. M., Hanzawa, Y., Hall, M. C., Moore, R. C., & Purugganan, M. D. (2009). Population genomics of the *Arabidopsis thaliana* flowering time gene network. *Molecular Biology and Evolution, 26*, 2475–2486.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics, 180*, 977–993. https://doi.org/10.1534/genetics.108.092221

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology, 27*, 2215–2233.

Fournier-Level, A., Korte, A., Cooper, M. D., Nordborg, M., Schmitt, J., & Wilczek, A. M. (2011). A map of local adaptation in *Arabidopsis thaliana*. *Science, 334*, 86–89.

Frachon, L., Bartoli, C., Carrere, S., Bouchez, O., Chaubet, A., Gautier, M., … Roux, F. (2018). A genomic map of climate adaptation in *Arabidopsis thaliana* at a micro-geographic scale. *Frontiers in Plant Science, 9*, 967.

Gehan, M. A., Park, S., Gilmour, S. J., An, C., Lee, C. M., & Thomashow, M. F. (2015). Natural variation in the C-repeat binding factor cold response pathway correlates with local adaptation of Arabidopsis ecotypes. *The Plant Journal, 84*, 682–693.

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, 897–908.

Gillespie, J. H. (2000). Genetic drift in an infinite population. The pseudo-hitchhiking model. _Genetics_, 155, 909–919.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., … Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. _Nucleic Acids Research_, 40, D1178–D1186.

Graur, D. (2016). _Molecular and genome evolution_. Sunderland, MA: Sinauer.

Gunther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. _Genetics_, 195, 205–220.

Hall, M. C., & Willis, J. H. (2006). Divergent selection on flowering time contributes to local adaptation in Mimulus guttatus populations. _Evolution_, 60, 2466–2477.

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, 83–86.

Hancock, A. M., Witonsky, D. B., Alkorta-Aranburu, G., Beall, C. M., Gebremedhin, A., Sukernik, R., … Di Rienzo, A. (2011). Adaptations to climate-mediated selective pressures in humans. _PLoS Genetics_, 7, e1001375.

Harris, R. S. (2007). _Improved pairwise alignment of genomic DNA_. State College, PA: Pennsylvania State University.

Hassidim, M., Harir, Y., Yakir, E., Kron, I., & Green, R. M. (2009). Over-expression of CONSTANSLIKE 5 can induce flowering in short-day grown Arabidopsis. _Planta_, 230, 481–491.

Haudry, A., Platts, A. E., Vello, E., Hoen, D. R., Leclercq, M., Williamson, R. J., … Blanchette, M. (2013). An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. _Nature Genetics_, 45, 891–898.

Hendry, A. P., Taylor, E. B., & Mcphail, J. D. (2002). Adaptive divergence and the balance between selection and gene flow: Lake and stream stickleback in the Misty system. _Evolution_, 56, 1199–1216.

Henson, J., Tischler, G., & Ning, Z. (2012). Next-generation sequencing and large genome assemblies. _Pharmacogenomics_, 13, 901–915.

Hereford, J. (2009). A quantitative survey of local adaptation and fitness trade-offs. _American Naturalist_, 173, 579–588.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., … Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. _American Naturalist_, 188, 379–397. https://doi.org/10.1086/688018

Hofer, T., Ray, N., Wegmann, D., & Excoffier, L. (2009). Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. _Annals of Human Genetics_, 73, 95–108. https://doi.org/10.1111/j.1469-1809.2008.00489.x

Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J. F., Clark, R. M., … Guo, Y. L. (2011). The _Arabidopsis lyrata_ genome sequence and the basis of rapid genome size change. _Nature Genetics_, 43, 476–481.

Huber, C. D., Degiorgio, M., Hellmann, I., & Nielsen, R. (2016). Detecting recent selective sweeps while controlling for mutation rate and background selection. _Molecular Ecology_, 25, 142–156.

Huber, C. D., Nordborg, M., Hermisson, J., & Hellmann, I. (2014). Keeping it local: Evidence for positive selection in Swedish _Arabidopsis thaliana_. _Molecular Biology and Evolution_, 31, 3026–3039.

Hubisz, M. J., Pollard, K. S., & Siepel, A. (2011). PHAST and RPHAST: Phylogenetic analysis with space/time models. _Briefings in Bioinformatics_, 12, 41–51.

Hupalo, D., & Kern, A. D. (2013). Conservation and functional element discovery in 20 angiosperm plant genomes. _Molecular Biology and Evolution_, 30, 1729–1744.

Ingvarsson, P. K., & Bernhardsson, C. (2019). Genome-wide signatures of environmental adaptation in European aspen (_Populus tremula_) under current and future climate conditions. _Evolutionary Applications_, 13(1):132–142.

Jacobs, G. S., Sluckin, T. J., & Kivisild, T. (2016). Refining the use of linkage disequilibrium as a robust signature of selective sweeps. _Genetics_, 203, 1807–1825.

Jeong, C., & Di Rienzo, A. (2014). Adaptations to local environments in modern human populations. _Current Opinion in Genetics & Development_, 29, 1–8.

Jiang, B., Shi, Y., Zhang, X., Xin, X., Qi, L., Guo, H., … Yang, S. (2017). PIF3 is a negative regulator of the CBF pathway and freezing tolerance in Arabidopsis. _Proceedings of the National Academy of Sciences of the United States of America_, 114, E6695–E6702.

Joosen, R. V., Ligterink, W., Hilhorst, H. W., & Keurentjes, J. J. (2009). Advances in genetical genomics of plants. _Current Genomics_, 10, 540–549.

Kalladan, R., Lasky, J. R., Chang, T. Z., Sharma, S., Juenger, T. E., & Verslues, P. E. (2017). Natural variation identifies genes affecting drought-induced abscisic acid accumulation in Arabidopsis thaliana. _Proceedings of the National Academy of Sciences of the United States of America_, 114, 11536–11541.

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., … Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. _Nature Genetics_, 42, 348–354.

Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., & Eskin, E. (2008). Efficient control of population structure in model organism association mapping. _Genetics_, 178, 1709–1723.

Katoh, K., & Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. _Briefings in Bioinformatics_, 9, 286–298.

Kaufmann, J., Lenz, T. L., Kalbe, M., Milinski, M., & Eizaguirre, C. (2017). A field reciprocal transplant experiment reveals asymmetric costs of migration between lake and river ecotypes of three-spined sticklebacks (_Gasterosteus aculeatus_). _Journal of Evolutionary Biology_, 30, 938–950.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. _Ecology Letters_, 7, 1225–1241.

Keller, S. R., Levsen, N., Olson, M. S., & Tiffin, P. (2012). Local adaptation in the flowering-time gene network of balsam poplar, _Populus balsamifera_ L. _Molecular Biology and Evolution_, 29, 3143–3152.

Kemppainen, P., Knight, C. G., Sarma, D. K., Hlaing, T., Prakash, A., Maung Maung, Y. N., … Walton, C. (2015). Linkage disequilibrium network analysis (LDna) gives a global view of chromosomal inversions, local adaptation and geographic structure. _Molecular Ecology Resources_, 15, 1031–1045.

Kim, Y., & Nielsen, R. (2004). Linkage disequilibrium as a signature of selective sweeps. _Genetics_, 167, 1513–1524.

Kim, Y., & Stephan, W. (2002). Detecting a local signature of genetic hitch-hiking along a recombining chromosome. _Genetics_, 160, 765–777.

Lachance, J., & Tishkoff, S. A. (2013). Population genomics of human adaptation. _Annual Review of Ecology Evolution and Systematics_, 44, 123–143.

Lasky, J. R., Des Marais, D. L., Lowry, D. B., Povolotskaya, I., Mckay, J. K., Richards, J. 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, 2283–2296.

Lasky, J. R., Des Marais, D. L., Mckay, J. K., Richards, J. H., Juenger, T. E., & Keitt, T. H. (2012). Characterizing genomic variation of Arabidopsis thaliana: The roles of geography and climate. _Molecular Ecology_, 21, 5512–5529.

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.

Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. _PLoS ONE_, 3, e4010.

Long, Q., Rabanal, F. A., Meng, D., Huber, C. D., Farlow, A., Platzer, A., … Nordborg, M. (2013). Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nature Genetics*, 45, 884–890.

Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, 23, 2178–2192.

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, 24, 1031–1046.

Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17, 67–77.

Maddamsetti, R., Hatcher, P. J., Green, A. G., Williams, B. L., Marks, D. S., & Lenski, R. E. (2017). Core genes evolve rapidly in the long-term evolution experiment with *Escherichia coli*. *Genome Biology and Evolution*, 9, 1072–1083

Mckay, J. K., & Latta, R. G. (2002). Adaptive population divergence: Markers, QTL and traits. *Trends in Ecology & Evolution*, 17, 285–291.

Miller, W., Rosenbloom, K., Hardison, R. C., Hou, M., Taylor, J., Raney, B., … Kent, W. J. (2007). 28-way vertebrate alignment and conservation track in the UCSC Genome Browser. *Genome Research*, 17, 1797–1808.

Mojica, J. P., Mullen, J., Lovell, J. T., Monroe, J. G., Paul, J. R., Oakley, C. G., & Mckay, J. K. (2016). Genetics of water use physiology in locally adapted *Arabidopsis thaliana*. *Plant Science*. 251, 12–22.

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, 3632–3644.

Monroe, J. G., Powell, T., Price, N., Mullen, J. L., Howard, A., Evans, K., … Mckay, J. K. (2018). Drought adaptation in *Arabidopsis thaliana* by extensive genetic loss-of-function. *eLife*, 7, e41038.

Monte, E., Tepperman, J. M., Al-Sady, B., Kaczorowski, K. A., Alonso, J. M., Ecker, J. R., … Quail, P. H. (2004). The phytochrome-interacting transcription factor, PIF3, acts early, selectively, and positively in light-induced chloroplast development. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 16091–16098.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research*, 15, 1566–1575.

Oakley, C. G., Agren, J., Atchison, R. A., & Schemske, D. W. (2014). QTL mapping of freezing tolerance: Links to fitness and adaptive trade-offs. *Molecular Ecology*, 23, 4304–4315.

Oda, A., Fujiwara, S., Kamada, H., Coupland, G., & Mizoguchi, T. (2004). Antisense suppression of the Arabidopsis PIF3 gene does not affect circadian rhythms but causes early flowering and increases FT expression. *FEBS Letters*, 557, 259–264.

Pass, D. A., Sornay, E., Marchbank, A., Crawford, M. R., Paszkiewicz, K., Kent, N. A., & Murray, J. A. H. (2017). Genome-wide chromatin mapping with size resolution reveals a dynamic sub-nucleosomal landscape in Arabidopsis. *PLoS Genetics*, 13, e1006988.

Pennisi, E. (2014). Disputed islands. *Science*, 345, 611–613.

Perez-Figueroa, A., Garcia-Pereira, M. J., Saura, M., Rolan-Alvarez, E., & Caballero, A. (2010). Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, 23, 2267–2276.

Phifer-Rixey, M., Bi, K., Ferris, K. G., Sheehan, M. J., Lin, D., Mack, K. L., … Nachman, M. W. (2018). The genomic basis of environmental adaptation in house mice. *PLoS Genetics*, 14, e1007672.

Pisupati, R., Reichardt, I., Seren, U., Korte, P., Nizhynska, V., Kerdaffrec, E., … Nordborg, M. (2017). Verification of Arabidopsis stock collections using SNPmatch, a tool for genotyping high-plexed samples. *Scientific Data*, 4, 170184.

Price, A. L., Zaitlen, N. A., Reich, D., & Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, 11, 459–463.

Price, N., Moyers, B. T., Lopez, L., Lasky, J. R., Monroe, J. G., Mullen, J. L., … Mckay, J. K. (2018). Combining population genomics and fitness QTLs to identify the genetics of local adaptation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 5028–5033.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81, 559–575.

Rellstab, C., Fischer, M. C., Zoller, S., Graf, R., Tedder, A., Shimizu, K. K., … Gugerli, F. (2017). Local adaptation (mostly) remains local: Reassessing environmental associations of climate-related candidate SNPs in Arabidopsis halleri. *Heredity (Edinb)*, 118, 193–201. https://doi.org/10.1038/hdy.2016.82

Ristova, D., Giovannetti, M., Metesch, K., & Busch, W. (2018). Natural genetic variation shapes root system responses to phytohormones in Arabidopsis. *Plant J*, 96, 468–481.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., … Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419, 832–837.

Sackton, T. B., Grayson, P., Cloutier, A., Hu, Z., Liu, J. S., Wheeler, N. E., … Edwards, S. V. (2019). Convergent regulatory evolution and loss of flight in paleognathous birds. *Science*, 364, 74–78.

Salomé, P. A., Bomblies, K., Laitinen, R. A., Yant, L., Mott, R., & Weigel, D. (2011). Genetic architecture of flowering-time variation in Arabidopsis thaliana. *Genetics*, 188, 421–433.

Sandring, S., & Agren, J. (2009). Pollinator-mediated selection on floral display and flowering time in the perennial herb Arabidopsis lyrata. *Evolution*, 63, 1292–1300.

Sasaki, E., Zhang, P., Atwell, S., Meng, D., & Nordborg, M. (2015). "Missing" G x E variation controls flowering time in *Arabidopsis thaliana*. *PLoS Genetics*, 11, e1005597.

Savolainen, O., Lascoux, M., & Merila, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, 14, 807–820.

Savolainen, O., Pyhäjärvi, T., & Knürr, T. (2007). Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics*, 38, 595–619.

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., … Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, 15, 1034–1050.

Slotte, T., Hazzouri, K. M., Agren, J. A., Koenig, D., Maumus, F., Guo, Y. L., … Wright, S. I. (2013). The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nature Genetics*, 45, 831–835.

Sork, V. L. (2017). Genomic studies of local adaptation in natural plant populations. *Journal of Heredity*, 109, 3–15.

Soy, J., Leivar, P., Gonzalez-Schain, N., Sentandreu, M., Prat, S., Quail, P. H., & Monte, E. (2012). Phytochrome-imposed oscillations in PIF3 protein abundance regulate hypocotyl growth under diurnal light/dark conditions in Arabidopsis. *The Plant Journal*, 71, 390–401.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 479–498.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30, 2725–2729.

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution*, 29, 673–680.

Tuteja, N. (2007). Abscisic acid and abiotic stress signaling. *Plant Signaling & Behavior*, 2, 135–138.

Verhoeven, K. J., Poorter, H., Nevo, E., & Biere, A. (2008). Habitat-specific natural selection at a flowering-time QTL is a main driver of local adaptation in two wild barley populations. *Mol Ecol*, *17*, 3416–3424.

Via, S. (1991). The genetic structure of host plant adaptation in a spatial patchwork: Demographic variability among reciprocally transplanted pea aphid clones. *Evolution*, *45*, 827–852.

Wadgymar, S. M., Lowry, D. B., Gould, B. A., Byron, C. N., Mactavish, R. M., & Anderson, J. T. (2017). Identifying targets and agents of selection: Innovative methods to evaluate the processes that contribute to local adaptation. *Methods in Ecology and Evolution*, *8*, 738–749.

Wang, J., Ding, J., Tan, B., Robinson, K. M., Michelson, I. H., Johansson, A., … Ingvarsson, P. K. (2018). A major locus controls local adaptation and adaptive life history variation in a perennial plant. *Genome Biology*, *19*, 72.

Wang, L., Jia, P., Wolfinger, R. D., Chen, X., Grayson, B. L., Aune, T. M., & Zhao, Z. (2011). An efficient hierarchical generalized linear mixed model for pathway analysis of genome-wide association studies. *Bioinformatics*, *27*, 686–692.

Yang, J., Guo, B., Shikano, T., Liu, X., & Merila, J. (2016). Quantitative trait locus analysis of body shape divergence in nine-spined sticklebacks based on high-density SNP-panel. *Scientific Reports*, *6*, 26632.

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution*, *65*, 1897–1911.

Yoder, J. B., & Tiffin, P. (2017). Effects of gene action, marker density, and timing of selection on the performance of landscape genomic scans of local adaptation. *Journal of Heredity*, *109*, 16–28.

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., … Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, *38*, 203–208.

Zhen, Y., Dhakal, P., & Ungerer, M. C. (2011). Fitness benefits and costs of cold acclimation in *Arabidopsis thaliana*. *American Naturalist*, *178*, 44–52.

Zhen, Y., & Ungerer, M. C. (2008). Relaxed selection on the CBF/DREB1 regulatory genes and reduced freezing tolerance in the southern range of *Arabidopsis thaliana*. *Molecular Biology and Evolution*, *25*, 2547–2555.

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, *44*, 821–824.

Zou, C., Sun, K., Mackaluso, J. D., Seddon, A. E., Jin, R., Thomashow, M. F., & Shiu, S. H. (2011). Cis-regulatory code of stress-responsive transcription in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 14992–14997.

SUPPORTING INFORMATION

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