Combining population genomics and fitness QTLs to identify the genetics of local adaptation in *Arabidopsis thaliana*

Nicholas Price[a,1], Brook T. Moyers[a], Lua Lopez[b], Jesse R. Lasky[b], J. Grey Monroe[a], Jack L. Mullen[a], Christopher G. Oakley[c,d], Junjiang Lin[a], Jon Ågren[e], Daniel R. Schrider[f], Andrew D. Kern[f], and John K. McKay[a]

[a]Department of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO 80523; [b]Department of Biology, Pennsylvania State University, University Park, PA 16802; [c]Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907; [d]Center for Plant Biology, Purdue University, West Lafayette, IN 47907; [e]Department of Plant Ecology and Evolution, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala, Sweden; and [f]Department of Genetics, Rutgers University, Piscataway, NJ 08854

Edited by Nils Chr. Stenseth, University of Oslo, Oslo, Norway, and approved April 3, 2018 (received for review November 15, 2017)

Evidence for adaptation to different climates in the model species *Arabidopsis thaliana* is seen in reciprocal transplant experiments, but the genetic basis of this adaptation remains poorly understood. Field-based quantitative trait locus (QTL) studies provide direct but low-resolution evidence for the genetic basis of local adaptation. Using high-resolution population genomic approaches, we examine local adaptation along previously identified genetic trade-off (GT) and conditionally neutral (CN) QTLs for fitness between locally adapted Italian and Swedish *A. thaliana* populations [Ågren J, et al. (2013) *Proc Natl Acad Sci USA* 110:21077–21082]. We find that genomic regions enriched in high F_{ST} SNPs colocalize with GT QTL peaks. Many of these high F_{ST} regions also colocalize with regions enriched for SNPs significantly correlated to climate in Eurasia and evidence of recent selective sweeps in Sweden. Examining unfolded site frequency spectra across genes containing high F_{ST} SNPs suggests GTs may be due to more recent adaptation in Sweden than Italy. Finally, we collapse a list of thousands of genes spanning GT QTLs to 42 genes that likely underlie the observed GTs and explore potential biological processes driving these trade-offs, from protein phosphorylation, to seed dormancy and longevity. Our analyses link population genomic analyses and field-based QTL studies of local adaptation, and emphasize that GTs play an important role in the process of local adaptation.

divergent selection | ecotype | FST | selective sweep | tradeoff

The study of local adaptation has a long history (1), yet its genetic basis is poorly understood (2). A major open question is whether local adaptation results from (*i*) genetic trade-offs (GTs), where alleles that maximize fitness in the home environment are deleterious in alternative environments, or (*ii*) conditionally neutral (CN) alleles that are advantageous in the home environment but neutral in alternative environments (3). Addressing the above will help enhance our understanding of how temporally or spatially varying selection maintains genetic variation (4, 5), which population genetic signatures can be used to identify local adaptation in the genome (6–8), and finally the biological process underlying local adaptation in different organisms (9).

Loci underlying GTs between populations are the result of variation in selection over space creating different local fitness optima (4). These loci should exhibit signals of divergent selection (6, 10) including the following: an increase in allele frequency differences between populations, an excess of intermediate frequency variants in the site frequency spectrum across all populations, and, if local adaptation is due to recent novel mutations that have fixed or gone to high frequency, evidence of a selective sweep in at least one population (11, 12). The identification of selective sweeps, like all methods used to detect local adaptation, can be influenced by population structure and demography (13), but also the strength of selection and whether it is acting on standing genetic variation (14). If the abiotic environment directly or indirectly creates the GT, we expect allele frequencies at these loci to be correlated with climate or other abiotic variables across the species' range (15–17). By definition, CN alleles are evolving neutrally in some subset of the species range, and thus population genetic signals at these loci should resemble drift more than GT alleles and therefore be more difficult to identify (6, 7).

Sessile organisms like plants are excellent systems for studies of local adaptation, and evidence for such local adaptation is abundant and well established (18). As in many species, reciprocal transplant experiments show substantial local adaptation in the model system *Arabidopsis thaliana* (19–21), and genome scans, climate associations, and genetic mapping of fitness in field experiments highlight genomic regions that may underlie local adaptation (8, 15, 16, 22). Genetic mapping, functional, and modeling studies have all illuminated the genetic basis of putatively adaptive traits (15, 23–27). However, evidence linking polymorphism at specific loci to variation in fitness across environments in natural populations is lacking.

Ågren et al. (19) identified six GT quantitative trait loci (QTLs) and three CN QTLs in a recombinant inbred (RIL) population derived from a cross between *A. thaliana* accessions from Italy and Sweden, evaluated in both parental environments for 3 y. At GT

Significance

Local adaptation can occur due to individual genetic variants that increase the fitness of individuals in their home environments but decrease fitness in other environments [genetic trade-offs (GTs)] or genetic variants that increase fitness in one environment but have no effect in other environments [conditional neutrality (CN)]. Here, we show that GT quantitative trait loci (QTLs) for fitness between Italian and Swedish *Arabidopsis thaliana* exhibit strong population genomic signatures of local adaptation, including elevated levels of allele frequency differentiation, correlations to climatic variables, and recent sweeps. Highly divergent genes between Italy and Sweden populations show evidence of more recent selection in Sweden than Italy, and the biological annotations of these genes suggest interesting mechanisms underlying local adaptation.

Author contributions: N.P. and J.K.M. designed research; N.P., L.L., and J.R.L. performed research; N.P., B.T.M., L.L., J.R.L., J.G.M., J.L.M., J.L., D.R.S., and A.D.K. analyzed data; and N.P., B.T.M., C.G.O., J.Å., and J.K.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence should be addressed. Email: price4890@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1719998115/-/DCSupplemental.

Published online April 23, 2018.

Downloaded by guest on April 14, 2021

QTLs, the local allele was favored by selection in its home environment, but was associated with reduced fitness in the alternate environment during at least 1 y of their experiment. At CN QTLs, the local allele was favored in its home environment, but had no detectable effect on fitness in the other. Here, we use resequencing data from Eurasian *A. thaliana* accessions (25), including the two RIL parent accessions and nearby Swedish and Italian populations, to examine genomic signatures of local adaptation and selection across the whole genome and within these fitness QTLs.

Results

Population Genomic Signatures of Local Adaptation Along Fitness QTLs. The observed GT and CN QTLs contain thousands of genes that vary at single-nucleotide polymorphisms (SNPs) between the Italy and Sweden RIL parents (*SI Appendix*, Fig. S1), and so potentially underlie local adaptation in this system. To narrow down the list of genes, we used population genomic and sliding window approaches to identify regions along chromosomes with significant evidence of local adaptation. Using a sample of 40 accessions from Italy and 40 from Sweden (*SI Appendix*, Fig. S2), we first identified significantly high F_{ST} SNPs (implemented in BayesScan 2.0, q value < 0.2; Fig. 1*A* and *SI Appendix*, Table S1) and then chromosomal regions with an elevated proportion of high F_{ST}. Clusters of SNPs along a chromosome with high F_{ST} can result from genetic linkage, linkage disequilibrium due to selection, and/or nearby genes that experience divergent selection similarly, possibly due to shared function (28). Four of the six GT QTLs (2:2, 4:2, 5:1, 5:5) and none of the three CN QTLs overlapped with windows with significantly elevated proportions of high F_{ST} (Fig. 1*A*). This may be driven by the smaller size of CN QTLs intervals relative to GT QTLs, as the latter are made up of multiple site by year single QTLs.

We examined chromosomal regions containing high proportions of sites with significant correlations to climate (Fig. 1 *B* and *C* and *SI Appendix*, Fig. S3 *B* and *E*). We chose six variables to capture key aspects of climatic variation between Italy and Sweden (17) and looked for correlations with SNPs across 875 Eurasian accessions. We set our significance threshold to an false-discovery rate (FDR) of <0.05 after accounting for random associations due to population structure (29) and included only SNPs that varied between our two focal populations (*SI Appendix*, Fig. S2). We found the number of SNPs correlated with minimum temperature of the coldest month to be an order of magnitude larger than the number of SNPs correlated to any of the other six variables (*SI Appendix*, Table S1). Five GT QTLs (1:3, 2:2, 3:3, 4:2, 5:5) and one CN QTL (2:1) contained windows with significantly elevated proportions of SNPs correlated to climate (Fig. 1 *B* and *C* and *SI Appendix*, Fig. S3 *B–E*).

Using composite likelihood ratio (CLR) tests (implemented in SweepFinder2), we identified genomic regions with significant evidence of recent selective sweeps in Italy and Sweden (Fig. 1*D*). We also identified regions that showed significant increases in linkage disequilibrium (LD) in Italy and Sweden (Fig. 1*E*). We found strong signals of selective sweeps in Sweden across all chromosomes, but only two in Italy, both in centromeric regions (Fig. 1*D*). Four of six GT QTLs showed evidence of recent selective sweeps in Sweden (Fig. 1*D*). Only one CN QTL is near a recent selective sweep in Sweden (Fig. 1*D*). LD was significantly correlated between the two populations ($r \sim 0.60$), and sites with a high proportion of significantly high F_{ST} SNPs (Fig. 1*A*) or a CLR > 200 in Sweden (Fig. 1*D*) are within high average LD windows in both populations (*SI Appendix*, Fig. S4). To consider possible effects of substructure within the Italy and Sweden samples of accessions, we reestimated recent selective sweeps using a subset of accessions, following the approach used in Huber et al. (8) with accessions limited to those collected >2 km apart. The broad pattern of major sweeps in Sweden did not

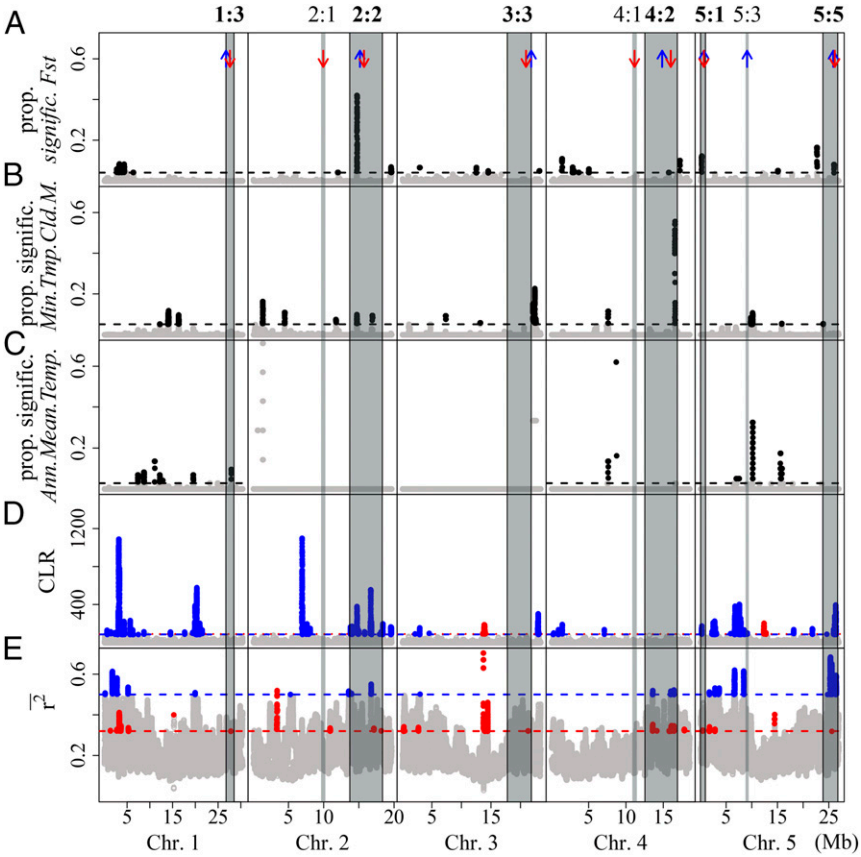

Fig. 1. Genome-wide patterns of local adaptation. Shaded regions represent GT QTLs identified by Ågren et al. (19) whose nomenclature is represented at the top of the figure (1:3, 2:2, 3:3, 4:2, 5:1, 5:5; bolded). CN QTLs (2:1, 4:1, 5:3) are shown as lines based on the location of the QTLs. Red and blue arrows along the *Top* represent site by year QTL peaks in Italy and Sweden, respectively. (*A*) Proportion of significantly high F_{ST} SNPs between Italy and Sweden populations within 50-kb windows. (*B* and *C*) Proportion of SNPs in 50-kb windows with significant correlations to minimum temperature of coldest month or annual mean temperature in 875 Eurasian accessions. (*D*) Composite likelihood ratio (CLR) test for selective sweeps, estimated every 1 kb within each population (red, Italy; blue, Sweden). (*E*) Average LD across SNPs within 100 kb. Significant thresholds for proportions (dashed lines) was set at the 99% percentile of all proportion across genome. Significant proportions are shown for chromosomes with at least 20 SNPs with significantly high F_{ST} or correlation to climate. Significance for CLR was estimated using simulations of a neutral model.

POPULATION BIOLOGY

Downloaded by guest on April 14, 2021

change, whereas in Italy we observe no significant sweeps with the smaller sample size (*SI Appendix*, Fig. S5).

High F_{ST} Regions and Recent Selective Sweeps in Sweden Colocalize with Climate-Correlated Regions. Since climate is an important factor in local adaptation, we expect selective sweeps and high F_{ST} regions to colocalize with climate-correlated regions of the genome. We estimated the mean and median distances from the midpoint of the 42 windows with elevated proportions of sites with significant climate correlations to (*i*) the midpoint of 20 regions with elevated proportions of high F_{ST} SNPs and (*ii*) the location of the highest CLR within 20 recent sweeps in Sweden. The mean and median distances for both high F_{ST} regions and selective sweeps were ∼1.60 and ∼0.80 Mb. The difference between mean and median indicates a skew in the distribution of distances. The observed median distance of ∼0.80 Mb is at approximately the 35th percentile of a permutated distribution of 10,000 random genomic samples of the same size (*SI Appendix*, Fig. S6), which indicates that certain sweeps or high F_{ST} regions (e.g., strongest sweep on Chr. 1, Fig. 1*D*) are not close to climate-correlated regions. To identify single high F_{ST} regions or sweeps that are found significantly close to regions with a significant proportion of climate correlated SNPs, we performed a single locus permutation test and identified three sweeps in Sweden and four high F_{ST} regions within the fifth percentile, or ≤0.08 Mb from a climate-correlated region (*SI Appendix*, Table S2). The proportion of high F_{ST} regions within ≤0.08 Mb from a climate-correlated region is significantly higher than expected by chance (binomial test *P* value < 0.05). Notably, two selective sweeps and a high F_{ST} region are within GT QTL 2:2, where the first sweep is associated with a minimum temperature of the coldest month correlated region and the second sweep with a soil moisture correlated region (*SI Appendix*, Fig. S7), potentially representing adaptation to different environmental variables.

GT QTLs Colocalize with High F_{ST} Regions. Given that divergent selection between Italy and Sweden populations generated GT and CN QTLs identified in the field (19), we expect an enrichment of population genetic signatures of local adaptation near QTL sites. GT QTLs span an average of ∼3 Mb, much larger than the length of population signatures of local adaptation they contain (Fig. 1). The large genomic regions of GT QTLs could be the result of the low resolution of QTL mapping and/or represent linked fitness QTLs that are difficult to disentangle (30). To account for these factors and the difference in length between CN and GT QTLs, we took pairs of single-year QTLs whose peaks were the shortest pairwise distance from each other within GT QTLs (Fig. 1*A*, red and blue arrows). Using the genomic midpoint between these pairs and the location of CN QTL peaks, we estimated mean and median distances to the nearest region with population genomic evidence of local adaptation including (*i*) elevated proportions of high F_{ST} SNPs, (*ii*) elevated proportion of climate-correlated SNPs, and (*iii*) highest CLR of sweep regions in Sweden. High F_{ST} and climate-correlated regions are closer to GT QTLs than selective sweeps (*SI Appendix*, Table S3). Climate-correlated regions are similarly close to CN QTLs and GT QTLs, while GT QTLs are closer to high F_{ST} regions than would be expected by chance alone (∼2% of the permuted median distance of 10,000 random genomic samples; Fig. 2).

Protein-Coding Genes Undergoing Adaptive Evolution in Italy and Sweden Populations. Following the colocalization of high F_{ST} regions with GT QTLs and correlations to climate, we examined the distribution of high F_{ST} SNPs along the genome in more detail. Approximately 67% of the 2,401 high F_{ST} SNPs are within transcribed genic regions, with ∼17% in *cis*-regulatory regions. Relative to the proportion of the genome covered by these regions (transcribed genic regions, 23%; promoter regions, 14%), both transcribed regions and promoter regions are significantly enriched in high F_{ST} SNPs (binomial test; *P* value < 0.0001 for each test).

To test whether the unfolded site frequency spectrum (uSFS) is shifted for genes containing high F_{ST} SNPs in transcribed regions, we compared it to the uSFS across genes lacking high F_{ST} SNPs (Fig. 3). The uSFS for genes containing high F_{ST} SNPs in transcribed regions in Italy and Sweden populations are significantly different from those observed for other genes (two-tailed X^2: Italy, 429.90; Sweden, 1,113.23; df = 40 and P < 0.0001 for both tests). When comparing the uSFS spectrum of genes containing high F_{ST} SNPs within their transcribed regions to that of all other genes with 1:1 orthologs in *Arabidopsis lyrata* and *Capsella rubella*, in Sweden the pattern observed is that expected under a model of recent selection (31), including an increase in low-frequency variants (≤0.1), a lower proportion of intermediate-frequency variants (>0.1 and ≤0.9), and a higher proportion of high-frequency variants (>0.9; Fig. 3). This observation aligns with the high CLRs observed in Sweden in regions with high F_{ST} (Fig. 1 *A* and *D*). By comparison, in the Italian population, the pattern is much less pronounced, perhaps suggestive of older positive selection (31) (Fig. 3). This may explain the much lower CLRs observed in Italy (Fig. 1*D*).

Adaptive evolution can occur through changes at the amino acid level, in which we expect an increase in their frequency within a population. To detect such possible changes we examined derived amino acid frequency differences between Italy and Sweden populations ($x = fr_{Italy}[i] - fr_{Sweden}[i]$; *i*: polymorphic amino acid site). Among the 327 genes with high F_{ST} along transcribed regions, 90 genes contained sites with a high derived amino acid frequency in Sweden ($x < -0.90$), 56 genes contained sites with a high derived frequency in Italy ($x > 0.90$), and 30 genes contained sites with a high derived amino acid frequencies in both populations ($x < -0.90$ and $x > 0.90$). The proportion of high F_{ST} genes with high derived amino acid frequency differences in both populations was significantly higher (∼9%) than the rest of the genes in the genome (0.3%).

Candidate Genes Underlying GTs. Given the significant association between high F_{ST} SNPs and GT QTLs (Fig. 2), we identified a comprehensive list of putative candidate genes underlying these GTs using the following criteria: (*i*) genes within 0.6 Mb of the peaks of GT QTLs (approximately the median distance between

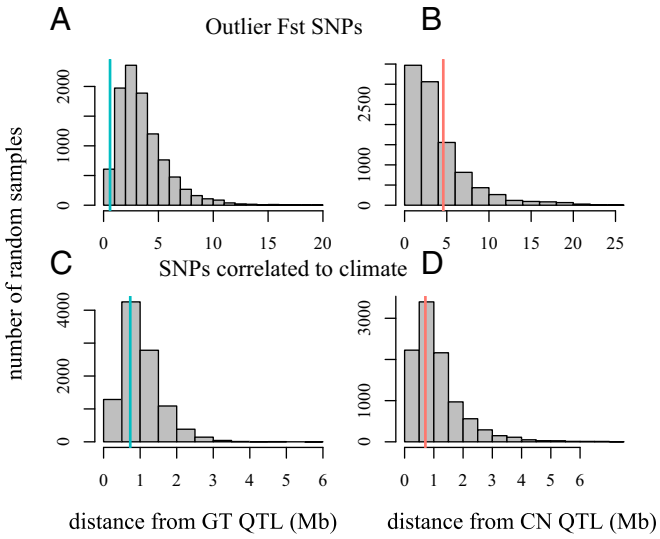

Fig. 2. Significant association between peaks of fitness QTLs and regions with population signatures of selection. (*A* and *B*) The median distance of the six GT QTLs (*A*; ∼2% of the randomly sampled distributions) and three CN QTLs (*B*; 73%) to the nearest region with a high proportion of significant outlier F_{ST} SNPs. (*C* and *D*) The median distance of the six GT QTLs (*C*; ∼32%) and the three CN QTLs (*D*; ∼30%) to the nearest region with a high proportion of SNPs with significant correlations to one or more of the six climate variables examined. The permuted distributions are of median distances from 10,000 random genomic location samples of the same size (3, 6).

Downloaded by guest on April 14, 2021

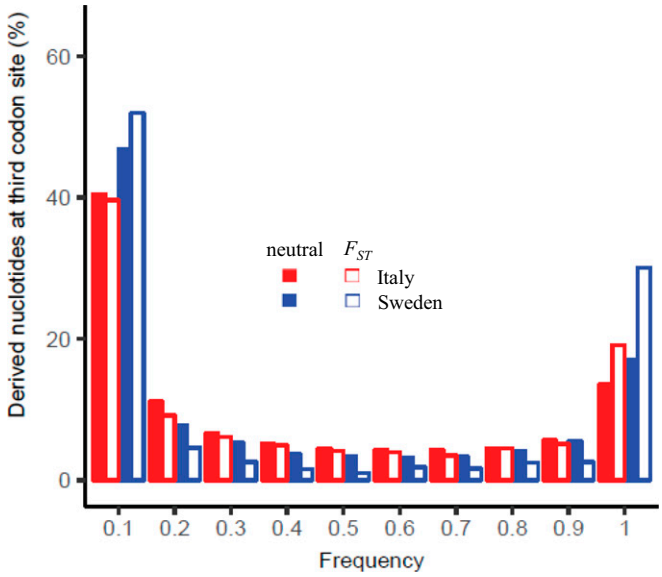

Fig. 3. The unfolded site frequency spectrum in Italy (red) and Sweden (blue) along genes containing high F_{ST} sites (empty bars) and all other genes (filled bars; neutral expectation) with 1:1 orthologs in *Arabidopsis lyrata* and *Capsella rubella*. Unfolded site frequency spectra are estimated from concatenated third codon positions.

high F_{ST} regions and GT QTLs), and (*ii*) containing high F_{ST} SNPs that varied between the two parent accessions and were within transcribed or promoter regions.

Our final list of 42 genes contains 21 genes in QTL 2:2, 1 gene in QTL 3:3, 5 genes in QTL 4:2, 10 genes in QTL 5:1, and 5 genes in QTL 5:5 (*SI Appendix*, Table S4). Of the 34 genes for which we were able to calculate Tajima's D, 24 have Tajima's $D > 1$ at synonymous sites, which is strong evidence of divergent selection (*SI Appendix*, Table S4). Six of the genes in *SI Appendix*, Table S4, show significant evidence of diversifying selection according to an Hudson–Kreitman–Aguade (HKA) test. Eighteen genes contain sites with high derived amino acid frequencies in one or both populations (*SI Appendix*, Table S4). Putatively mechanistic biological processes associated with our final list of genes include the following: cold acclimation/tolerance, respiratory metabolism, circadian rhythm, and seed longevity and dormancy (*SI Appendix*, Table S4).

Expression response of many genes to stressful conditions depends strongly on natural genotypic variation (32). Assuming that genes exhibiting expression genotype by environment interactions under stressful conditions underlie genotype by environments interactions at the fitness level, then these genes may represent candidates for local adaptation (15, 33, 34). To identify such candidates, we examined expression variation under control (22 °C) and cold conditions (4 °C) after 1 and 2 wk of cold (35). Among the genes in *SI Appendix*, Table S4, four genes showed a main effect in genotype (G); 14 showed additive genetic and environmental variance (G+E); and five showed genotype by environment interactions (G×E) after 1 and/or 2 wk of cold. Genes along QTL 2:2 and QTL 5:1 showed expression G×E interactions and therefore may underlie local adaptation (15).

GT QTL 2:2 is among the QTLs with the highest effect sizes in both Italy and Sweden (19) and shows high genetic differentiation between Italy and Sweden populations, correlations to minimum temperature of the coldest month, and recent selection in Sweden (Fig. 1 *A*, *B*, and *D*). It also contains gene AT2G35050, a predicted kinase showing significant amino acid divergence between Italy and Sweden populations, and highly significant expression G×E interactions (FDR $< 5.08 \times 10^{-5}$) under cold (*SI Appendix*, Table S4). Protein kinases are post-translational regulators that are involved in abiotic stress responses (36) and therefore may play a role in local adaptation. Using *A. lyrata* and *C. rubella* as outgroups, we examined allele divergence among Eurasian accessions and found two major allele groups, each containing one of the parent accessions (Fig. 4*A*). Eurasian accessions sharing a similar allele to the Sweden parent were found in colder regions than accessions sharing a similar allele to the Italy parent (Fig. 4 *B* and *C*). Fig. 4*D* depicts average expression ($\overline{FPKM}$) in fragments per kilobase of exon model per million mapped fragments of AT2G35050 in Italy and Sweden plants under control and cold conditions. The average under each condition is based on three replicates.

Additional genes along QTL 2:2 showing G×E interactions were AT2G34940, AT2G35010, AT2G35190. These genes are involved in seed germination, cell redox homeostasis, and abscisic acid (ABA) regulated root growth (*SI Appendix*, Table S4), therefore representing interesting candidates for local adaptation.

Discussion

Understanding the genetic basis of local adaptation is a complex but increasingly feasible challenge. The current study links direct evidence of genetic fitness trade-offs identified by field-based QTL experiments in native habitats (19) to population genomic signatures of local adaptation. We find substantial population evidence of selection underlying GT QTLs, including colocalization of regions with elevated F_{ST}, which in some instances overlap with regions showing evidence of recent sweeps in Sweden and climate-correlated SNPs in Eurasia. Examining the uSFS of genes containing high F_{ST} sites, we find evidence of recent selection in Sweden and potentially older selection in Italy. Finally, we identify a list of 42 genes that represent strong candidates underlying GTs between Italy and Sweden and point to potential mechanisms of local adaptation.

In contrast to GT QTLs, CN QTLs do not show any significant colocalization with any population genomic signatures of selection. CN QTLs are closest (but not closer than expected by chance) to regions enriched with climate-correlated SNPs (~0.70 Mb), approximately as close as GT QTLs. Despite filtering for sites segregating between Italy and Sweden populations, the close but not significant pattern may be driven by the presence of climate-correlated clusters of SNPs that are independent of local adaptation in the Italy and Sweden populations examined. Additionally, if some SNPs under selection and near QTL peaks are heavily confounded by population structure, they are likely to be eliminated after accounting for population structure.

Our observed colocalization of GT QTLs with high F_{ST} regions is predicted by simulation modeling (6). Although we find some high F_{ST} SNPs within CN QTLs, we find no significant enrichment in high F_{ST} SNPs, and no more colocalization of these regions than expected by chance. This may be linked to the expected duration of these signals: population genomic signals of divergence at CN QTLs are likely to erode more quickly over time if the advantageous allele in one population shifts to high frequency in the other due to gene flow or genetic drift (7). The results of Ågren et al. (19) suggest that GT may be more frequent than CN, but it is also possible that the observed GT QTLs actually are composed of multiple GT and CN QTLs, as is suggested by the example of QTL 2:2 where we identify multiple sweeps but only one high F_{ST} region (*SI Appendix*, Fig. S7).

Allele effect sizes are expected to follow an exponential distribution under adaptation without migration (37). Contrary to this expectation, the six GT QTLs explain a large proportion of genetic variance in fitness (19) and largely colocalize with closely linked, highly diverged alleles that show recent selection in Sweden and older positive selection in Italy. Fewer, larger, and tightly linked divergent alleles are expected when populations are connected by gene flow and experience selection toward different optima (38). A significant constraint that may lead to the evolution of such genetic architectures is the limitation in the number of possible beneficial mutations occurring in tight enough linkage to build a locally favored phenotype (38).

Downloaded by guest on April 14, 2021

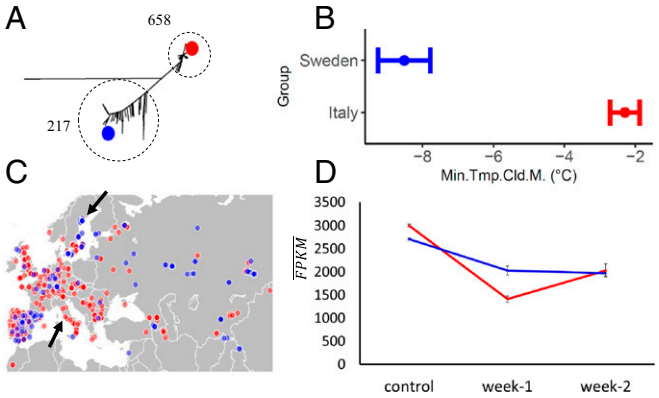

Fig. 4. Highly diverged alleles of AT2G35050 segregate to locations across Eurasia with significantly different temperatures during the coldest month of the year. (*A*) The rooted genealogy (outgroups: *A. lyrata* and *C. rubella*) of AT2G35050 using 875 Eurasia accessions. The blue and red dots indicate the topology of the Sweden and Italy RIL parent accessions, with the number of accessions in each major clade labeled. (*B*) Map of accessions sharing the same allele as the Sweden (blue) or Italy parent (red). Arrows indicate approximate location of Italy and Sweden parents. (*C*) Accessions sharing the same allele as the Sweden parent grow in regions with significantly lower temperatures during the coldest month of the year (Min.Tmp.Cld.M) than accessions sharing the same allele as the Italy parent (95% CI). (*D*) Average expression ($\overline{FPKM}$) and corresponding SEs of AT2G35050 under control conditions (22 °C) and 2 wk of cold (4 °C). According to *DESeq2* (57), AT2G35050 showed strong G×E interactions (*SI Appendix*, Table S4).

Similar to our study, Long et al. (22) identified 22 recent sweeps in North Sweden, and almost none in South Sweden. Subsequently, Huber et al. (8) analyzed these sweeps using various models to account for complex demographic history and identified nine new local sweeps in South Sweden, with only three of the original sweeps remaining significant in North Sweden. The three sweeps in North Sweden were all on chromosome 5 and did not include the highly significant and well-supported sweeps that were observed in the present study (e.g., along QTL 2:2) (*SI Appendix*, Fig. S7) even when accessions were subsampled to limit the effects of substructure (*SI Appendix*, Fig. S5). Our results suggest a lower false-positive rate in recent sweeps than estimated by Huber et al. (8). While there is reason to worry about the false-positive rate of SweepFinder2 under certain demographic models, our evidence of sweeps at GT QTLs is buttressed by the colocalization of climate correlations and regions enriched for high F_{ST} sites.

Among the candidate genes underlying GT QTLs (*SI Appendix*, Table S4), we find some interesting prospects for the biological mechanisms of local adaptation between Italy and Sweden. We described the potential role of AT2G35050 (QTL 2:2) in local adaptation to temperature above (Fig. 4). Using the same RIL population, Postma and Ågren (39) had found seed dormancy QTLs overlapping GT QTL 2:2 and 5:1. Seed dormancy is one mechanism for avoiding germination during stressful conditions (40, 41). Genes AT2G34900, AT2G34940, and AT5G01560 within these QTLs were found to regulate (42, 43) or affect germination (44), with genes AT2G34900 and AT5G01560 acting through the ABA response pathway (42, 43) and AT2G34940 showing expression G×E interactions under cold (*SI Appendix*, Table S4). AT5G65410 (QTL 5:5) loss-of-function mutants decrease seed longevity (45), so it may be a candidate for trade-offs between seed longevity and dormancy. Seed dormancy and longevity can be negatively correlated (46) and have shown G×E interactions (47).

This study suggests that GTs are a major component of local adaptation in *A. thaliana* and points to climate as one of the selective agents driving these trade-offs. Fitness trade-offs are of particular importance in evolution, because such spatial variation in the fitness of alleles can stably maintain genetic polymorphism across the species range (48, 49). However, fitness trade-offs have also proven difficult to observe, requiring careful experimentation across environments, often over multiple years, and so such studies are rare relative to studies of local adaptation more broadly. From our analyses, it is clear this effort will be needed to fully understand patterns of local adaptation and the maintenance of genetic variation.

Methods

Whole-Genome Sequencing of Italy and Sweden Parent Accessions. We used resequencing data of the RIL parent accessions described in Ågren et al. (19) to identify SNPs between the Italy and Sweden parent accessions (details in *SI Appendix*, *SI Method 1*).

Population Genetic Sampling. We used three sets of accessions downloaded from the *Arabidopsis thaliana* 1001 Genomes database (25): (*i*) 41 accessions in North-Central Italy; (*ii*) 49 accessions in North Sweden (*SI Appendix*, Fig. S2); and (*iii*) 875 accessions from across Eurasia (25). Collectively, the three sets included 880 unique accessions whose latitude and longitude coordinates are listed in Dataset S1. We aggregated the genotype files for these accessions and filtered for biallelic, nonindel sites. To test for selective sweeps, we used all 41 and 49 accessions in Italy and Sweden, and a subset of 17 accessions in Sweden and 8 accessions in Italy to compare our results to Huber et al. (8). For other population genetic analyses including F_{ST}, Tajima's *D*, and HKA, we chose a subset of 40 accessions from each region (*SI Appendix*, Fig. S2).

F_{ST} Peaks Along Chromosomes. We used BayeScan2.0 (50) to estimate F_{ST}: a method that incorporates a multinomial Dirichlet likelihood (51) to account for variation in allele frequencies due to various neutral population genetic models. We used two populations within our BayeScan analysis (Italy and Sweden) and analyzed each chromosome separately for computational practicality. We kept the prior odds for the neutral model at the default value 10. We set the individual test threshold at q value < 0.2 since the highest F_{ST} values fall within q values of 0–0.2 (*SI Appendix*, Fig. S8). To identify regions containing a high proportion of significant F_{ST} values, we used a sliding window of 50 kb and a step size of 1 kb. We estimated the proportion of significant F_{ST} values as the number of significant F_{ST} values within a window over the total number of F_{ST} values along the whole chromosome. We set significance for proportions at the 99th percentile of all proportions across the genome.

Selective Sweeps. We used Sweepfinder2 (52) and the largest sample of accessions from each region (41 from Italy and 49 from Sweden; *SI Appendix*) to calculate CLR for recent selective sweeps. We estimated CLRs every 1 kb along the genome within each population, and used *A. lyrata* as an outgroup to polarize SNPs (alignments downloaded from the VISTA database; ref. 53). We called polarity when the *A. lyrata* and *A. thaliana* nucleotide states were the same. To arrive at a significance threshold, we used a similar method as Long et al. (22): standard neutral simulations in the program ms (54). We scaled diversity (θ) to a sequence of length of 1 Mb and set recombination rate to five times smaller than the value of θ. Our significance thresholds were estimated at 85 CLRs for Sweden and 89 CLRs for Italy.

LD. To estimate LD (r^2) across the genome, we used the program plink 1.07 (55) with a window size of 1 kb ("–ld-window-kb 1"). To further normalize LD across the genome, we used a sliding window of 100 kb and a step size of 1 kb and estimated the mean of all r^2 within the window. We used the 99th percentile of all windows across the genome as our significance threshold.

Testing for Divergent Selection at the Genic Level. To estimate divergent selection at the genic level, we used the HKA test and Tajima's *D* test, and examined unfolded site frequency spectra at third codon positions or derived amino acid frequencies at the amino acid level ($x = fr_{Italy}[i] - fr_{Sweden}[i]$; *i*: polymorphic amino acid site). Details in estimating all of the above statistics are explained in *SI Appendix*, *SI Method 2*.

Correlations to Climate. We downloaded a SNP genotype matrix for a panel of 1,135 globally distributed accessions from the 1001 Genomes database. From this panel, we filtered out accessions from outside the native Eurasian and North African range of *A. thaliana*, as these accessions may have weaker patterns of local adaptation (17). We also filtered out accessions that were likely laboratory escapees or contaminants (56), leaving 875 accessions (*SI Appendix*). We filtered for biallelic SNPs with minor allele frequency >0.05. For each SNP, we tested association with home climate of ecotype and tested

Downloaded by guest on April 14, 2021

for potential confounding effects of population structure using the software "gemma" (29). Details on the methodology and climate variables are examined in *SI Appendix*, *SI Method 3*.

Estimating Distance Between QTLs and Regions Showing Population Signatures of Selection. To estimate distances between CN and GT QTLs, we first defined regions under selection by merging overlapping 50-kb windows with significantly high proportions of high F_{ST} sites and correlations to climate (Fig. 1 *A* and *B*). In the case of sweeps (Fig. 1*C*), we used a window of 500 kb as a single sweep event (22). We then estimated the physical distance between CN or GT QTLs and the nearest genomic region containing population genetic signatures of selection. Details on how distance was estimated is explained in *SI Appendix*, *SI Method 4*.

Examining Expression Variation Between Italy and Sweden Under Cold Conditions. To detect differences in expression between Italy and Sweden plants under cold, we used RNA data from the study by Gehan et al. (35). Details on the methodology is explained in *SI Appendix*, *SI Method 5*.

ACKNOWLEDGMENTS. We thank the lab groups of Graham Coop, Jeff Ross-Ibarra, and Annie Schmitt for helpful discussions on scans for local adaptation, and Christian D. Huber for discussions on simulations. In addition, we thank two anonymous reviewers for helpful suggestions on this manuscript. This study was financially supported by National Science Foundation Awards DEB 1022202, 1523752, and 1556262, and a grant from the Swedish Research Council (to J.Å.). A.D.K. was supported by NIH Grant R01GM078204.

1. Clausen J, Keck DD, Hiesey WM (1941) Regional differentiation in plant species. *Am Nat* 75:231–250.
2. Wadgymar SM, et al. (2017) Identifying targets and agents of selection: Innovative methods to evaluate the processes that contribute to local adaptation. *Methods Ecol Evol* 8:738–749.
3. Anderson JT, Lee CR, Rushworth CA, Colautti RI, Mitchell-Olds T (2013) Genetic trade-offs and conditional neutrality contribute to local adaptation. *Mol Ecol* 22:699–708.
4. Levene H (1953) Genetic equilibrium when more than one ecological niche is available. *Am Nat* 87:331–333.
5. Gillespie JH (1991) *The Causes of Molecular Evolution* (Oxford Univ Press, New York).
6. Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. *Trends Ecol Evol* 29:673–680.
7. Yoder JB, Tiffin P (2017) Effects of gene action, marker density, and timing of selection on the performance of landscape genomic scans of local adaptation. *J Hered* 109:16–28.
8. Huber CD, Nordborg M, Hermisson J, Hellmann I (2014) Keeping it local: Evidence for positive selection in Swedish *Arabidopsis thaliana*. *Mol Biol Evol* 31:3026–3039.
9. Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nat Rev Genet* 14:807–820.
10. Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet* 2:e64.
11. Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genet Res* 23: 23–35.
12. Kaplan NL, Hudson RR, Langley CH (1989) The "hitchhiking effect" revisited. *Genetics* 123:887–899.
13. Hoban S, et al. (2016) Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *Am Nat* 188:379–397.
14. Schrider DR, Kern AD (2016) S/HIC: Robust identification of soft and hard sweeps using machine learning. *PLoS Genet* 12:e1005928.
15. Lasky JR, et al. (2014) Natural variation in abiotic stress responsive gene expression and local adaptation to climate in *Arabidopsis thaliana*. *Mol Biol Evol* 31:2283–2296.
16. Fournier-Level A, et al. (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science* 334:86–89.
17. Lasky JR, et al. (2012) Characterizing genomic variation of *Arabidopsis thaliana*: The roles of geography and climate. *Mol Ecol* 21:5512–5529.
18. Hereford J (2009) A quantitative survey of local adaptation and fitness trade-offs. *Am Nat* 173:579–588.
19. Ågren J, Oakley CG, McKay JK, Lovell JT, Schemske DW (2013) Genetic mapping of adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 110:21077–21082.
20. Ågren J, Schemske DW (2012) Reciprocal transplants demonstrate strong adaptive differentiation of the model organism *Arabidopsis thaliana* in its native range. *New Phytol* 194:1112–1122.
21. Wilczek AM, Cooper MD, Korves TM, Schmitt J (2014) Lagging adaptation to warming climate in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 111:7906–7913.
22. Long Q, et al. (2013) Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat Genet* 45:884–890.
23. Oakley CG, Ågren J, Atchison RA, Schemske DW (2014) QTL mapping of freezing tolerance: Links to fitness and adaptive trade-offs. *Mol Ecol* 23:4304–4315.
24. Ågren J, Oakley CG, Lundemo S, Schemske DW (2017) Adaptive divergence in flowering time among natural populations of *Arabidopsis thaliana*: Estimates of selection and QTL mapping. *Evolution* 71:550–564.
25. 1001 Genomes Consortium (2016) 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* 166:481–491.
26. Murphey M, et al. (2015) DOG1-imposed dormancy mediates germination responses to temperature cues. *Environ Exp Bot* 112:33–43.
27. Oakley CG, et al. (2018) Genetic basis of photosynthetic responses to cold in two locally adapted populations of *Arabidopsis thaliana*. *J Exp Bot* 69:699–709.
28. Michalak P (2008) Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* 91:243–248.
29. Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 44:821–824.
30. Ronin YI, Korol AB, Nevo E (1999) Single- and multiple-trait mapping analysis of linked quantitative trait loci. Some asymptotic analytical approximations. *Genetics* 151:387–396.
31. Nielsen R (2005) Molecular signatures of natural selection. *Annu Rev Genet* 39: 197–218.
32. Des Marais DL, Hernandez KM, Juenger TE (2013) Genotype-by-environment interaction and plasticity: Exploring genomic responses of plants to the abiotic environment. *Annu Rev Ecol Evol Syst* 44:5–29.
33. Whitehead A, Crawford DL (2006) Neutral and adaptive variation in gene expression. *Proc Natl Acad Sci USA* 103:5425–5430.
34. Des Marais DL, et al. (2012) Physiological genomics of response to soil drying in diverse *Arabidopsis* accessions. *Plant Cell* 24:893–914.
35. Gehan MA, et al. (2015) Natural variation in the C-repeat binding factor cold response pathway correlates with local adaptation of *Arabidopsis* ecotypes. *Plant J* 84:682–693.
36. Wang H, Chevalier D, Larue C, Ki Cho S, Walker JC (2007) The protein phosphatases and protein kinases of *Arabidopsis thaliana*. *Arabidopsis Book* 5:e0106.
37. Orr HA (1998) The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution* 52:935–949.
38. Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. *Evolution* 65:1897–1911.
39. Postma FM, Ågren J (2015) Maternal environment affects the genetic basis of seed dormancy in *Arabidopsis thaliana*. *Mol Ecol* 24:785–797.
40. Postma FM, Lundemo S, Ågren J (2016) Seed dormancy cycling and mortality differ between two locally adapted populations of *Arabidopsis thaliana*. *Ann Bot* 117: 249–256.
41. Huang X, et al. (2010) The earliest stages of adaptation in an experimental plant population: Strong selection on QTLS for seed dormancy. *Mol Ecol* 19:1335–1351.
42. Duque P, Chua NH (2003) IMB1, a bromodomain protein induced during seed imbibition, regulates ABA- and phyA-mediated responses of germination in *Arabidopsis*. *Plant J* 35:787–799.
43. Xin Z, Wang A, Yang G, Gao P, Zheng ZL (2009) The *Arabidopsis* A4 subfamily of lectin receptor kinases negatively regulates abscisic acid response in seed germination. *Plant Physiol* 149:434–444.
44. Laval V, et al. (2003) Seed germination is blocked in *Arabidopsis* putative vacuolar sorting receptor (atbp80) antisense transformants. *J Exp Bot* 54:213–221.
45. Bueso E, et al. (2014) *ARABIDOPSIS THALIANA HOMEOBOX25* uncovers a role for Gibberellins in seed longevity. *Plant Physiol* 164:999–1010.
46. Nguyen TP, Keizer P, van Eeuwijk F, Smeekens S, Bentsink L (2012) Natural variation for seed longevity and seed dormancy are negatively correlated in *Arabidopsis*. *Plant Physiol* 160:2083–2092.
47. Hanin N, Quaye M, Westberg E, Barazani O (2013) Soil seed bank and among-years genetic diversity in arid populations of *Eruca sativa* Miller (Brassicaceae). *J Arid Environ* 91:151–154.
48. Mitchell-Olds T, Willis JH, Goldstein DB (2007) Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nat Rev Genet* 8:845–856.
49. Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet Res* 70:155–174.
50. 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.
51. Balding DJ (2003) Likelihood-based inference for genetic correlation coefficients. *Theor Popul Biol* 63:221–230.
52. DeGiorgio M, Huber CD, Hubisz MJ, Hellmann I, Nielsen R (2016) SweepFinder2: Increased sensitivity, robustness and flexibility. *Bioinformatics* 32:1895–1897.
53. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: Computational tools for comparative genomics. *Nucleic Acids Res* 32:W273–W279.
54. Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.
55. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575.
56. Des Marais DL, Guerrero RF, Lasky JR, Scarpino SV (2017) Topological features of a gene co-expression network predict patterns of natural diversity in environmental response. *Proc Biol Sci* 284:20170914.
57. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with *DESeq2*. *Genome Biol* 15:550.

Downloaded by guest on April 14, 2021