Experimental validation of genome-environment associations in Arabidopsis

Yuxin Luo[1,2], Claire M. Lorts[1], Erica Lawrence-Paul[1], Jesse R. Lasky[1]

[1]Department of Biology, Pennsylvania State University

[2]Corresponding author

Abstract

Identifying the genetic basis of local adaptation is a key goal in evolutionary biology. Allele frequency clines along environmental gradients, known as genotype-environment associations (GEA), are often used to detect potential loci causing local adaptation, but GEA are rarely followed by experimental validation. Here, we tested loci identified in three different moisture-related GEA studies on *Arabidopsis*. We studied 44 GEA-identified genes using t-DNA knockout lines under drought and tested for effects on flowering time, an adaptive trait, and genotype-by-environment (GxE) interactions on performance and fitness. We found that *wrky38* mutants had significant GxE effects for fitness; *lsd1* plants had a significant GxE effect for flowering time, while 11 genes showed flowering time effects with no drought interaction. However, most GEA candidates did not exhibit GxE. In the follow-up experiments we found *wrky38* caused decreased stomatal conductance and specific leaf area under drought, indicating potentially adaptive drought avoidance. Additionally, we found that GEA identified natural putative LoF variants of WRKY38 associated with dry environments, as well as alleles associated with variation in *LSD1* expression. While only a few GEA putative drought-adapted genes were validated for GxE interactions for fitness under drought, we likely overlooked some genes because experiments might not well represent natural environments and t-DNA insertions might not well represent natural alleles. Nevertheless, GEAs apparently identified some genes contributing to local adaptation. GEA and follow-up experiments are straightforward to implement in model systems and thus demonstrate prospects for GEA discovery of new local adaptations.

INTRODUCTION

A substantial portion of genetic variation within species in ecologically important traits is likely due to local adaptation to the environment (Savolainen et al., 2013; Tigano and Friesen, 2016; Wadgymar et al., 2022). Local adaptation is defined by a genotype-by-environment (GxE) interaction where local genotypes have higher fitness than foreign genotypes (Kawecki & Ebert, 2004). Identifying locally adaptive alleles and traits is of great interest in plant breeding and conservation, where these alleles can be deployed for genetic improvement and prediction (Lasky et al., 2023). Central questions of local adaptation include the effect size and number of involved mutations (Yeaman and Whitlock, 2011), whether locally adapted mutations have tradeoffs among different environments (Lee et al., 2024), whether the same or different mutations are involved in local adaptation to the same environments in different populations (Ralph and Coop, 2015), and which environmental gradients drive locally adaptive genetic variation (Lasky et al., 2012).

Conducting multiple common garden experiments to map populations and genetic loci contributing to fitness tradeoffs has been a gold standard approach (Clausen et al., 1941). However, such experiments are logistically challenging or impossible in many systems. Alternative approaches using geographic patterns in allele frequency have emerged for identifying loci potentially underlying local adaptation. Loci showing substantial allele frequency differences between populations, i.e. certain environment variables, may indicate those driving local adaptation (Coop et al., 2010; Endler, 1973). With the increasing availability of high-throughput genomic data, Genome-Environment Association (GEA) studies have been widely employed to identify putative locally adapted loci (Lasky et al., 2023; Rellstab et al., 2015).

GEAs employ diverse methods to identify loci where allele frequency is closely associated with environmental conditions (Lasky et al., 2023; Rellstab et al., 2015). Methods differ depending on whether they are single-locus or multi-loci (e.g. genome-wide) models (Gehan et al., 2015; Hancock et al., 2011; Lasky et al., 2012; Lee et al., 2024), where the latter may capture groups of loci covarying with the environment that are potentially relevant for polygenic adaptation (Forester et al., 2018). Additionally, there are methods that control for genome-wide patterns of similarity among populations using random effects (Zoubarev et al., 2012). While these approaches may reduce spurious associations due to population structure-environmental covariation, they can also be too conservative (Alonso-Blanco et al., 2016). Thirdly, some approaches synthesize the evidence from GEA with additional evidence, such as multiple common garden experiments (Capblancq et al., 2023; Lasky et al., 2018). While these methods are now standard approaches in ecological genomics, the findings are rarely followed by experimental validation of variation at individual loci.

To experimentally test the hypothesis that variation at a given locus causes GxE for fitness, approaches that isolate experimental genetic variation to that locus are the most powerful. Creating near-isogenic lines (NILs) segregating at a given locus is ideal for testing GxE effects in a natural genetic background, but it is a time-consuming process. Gene knockout mutants in model species like *Arabidopsis* provide an alternative tool for testing GxE effects of variation at a particular gene. Although mutants may not fully represent natural allelic variation for local adaptation, using them may be a good option for bulk experimental screening of candidate loci (Chong and Stinchcombe, 2019; Monroe et al., 2018). Resequencing of large

numbers of natural *Arabidopsis* genotypes has yielded more detailed information on functional variation (Alonso-Blanco et al., 2016), which can be linked to loci of interest to make inferences about functional variants generating strong signals in GEA studies.

Arabidopsis thaliana is a small annual plant native to a wide range across Eurasia and Africa with diverse climates (Yim et al., 2024). Many studies have demonstrated potential drought adaptation of *A. thaliana* in quantitative traits (Dittberner et al., 2018), SNPs or genes (El-Soda et al., 2015; Exposito-Alonso et al., 2018; Hancock et al., 2011), and gene expression (Lasky et al., 2014). *A. thaliana* employs diverse strategies for drought adaptation such as drought escape via phenology, drought avoidance via traits reducing water loss, and drought tolerance via enhancing water acquisition or water-use efficiency (Lovell et al., 2013).

Here we focus on studying loci identified based on 3 different approaches, with the goal of testing whether identified genes indeed cause genotype-drought effects on fitness or other aspects of performance. We are also interested in whether the genes identified as drought-adapted in our experiments have allele frequency clines across the worldwide distribution of *A. thaliana*. For mutants identified as showing interesting genotype-environment interactions in our initial drought screen experiment, we followed up by testing their response to an additional, different drought treatment.

METHODS

Selecting genes from published genome-environment associations

We selected genes to test using five SNP lists from three published GEA studies that employed three different approaches. The first approach was based on mixed models of genome-wide associations with climate, where we included random effects to account for genomic similarity (Kang et al., 2008). For these mixed models, we calculated SNP associations with intra-annual monthly precipitation variability, growing season precipitation variability, or inter-annual growing season precipitation variability (i.e. 3 separate environmental GWAS, Lasky et al., 2014). We took the SNPs from all three climate associations and combined them into one list ranked by p-value for the mixed model association test. The second approach was a multivariate ordination where we identified SNPs most strongly associated with the first axis, which was associated with seasonality of temperature and precipitation, or the first axis after partialling out spatial variables, which was strongly associated with summer moisture (Lasky et al., 2012). We took SNPs with the strongest absolute loadings on each of these axes. The third approach was based on integrated information from four common gardens across Europe and also the broader ecotype panel to identify SNPs that showed the strongest SNP-climate associations and GxE for fitness favoring home climate alleles in monthly growing season precipitation variability or aridity index (i.e. annual precipitation/PET; Lasky et al., 2018). In total, we generated 1, 2, and 2 lists of SNPs from the three studies respectively.

From the top SNPs in each list, we only kept SNPs at least 250 kb apart. Starting from the SNP with the lowest p-values, we took up to 3 closest genes within 5kb of the SNP until we

had at least 15 genes for each of the four lists of SNPs, or in the case of the combined climate association study, until we had at least 25 genes.

This procedure resulted in a list of 90 candidate genes. We chose 44 SALK lines representing knockouts of 42 genes based on insertion location within exons and confirmed homozygosity using tDNA Express: Arabidopsis Gene Mapping Tool (Alonso et al., 2003; Table S1).

Main drought screen

Experimental design and growth conditions

To mimic more natural growing conditions, we programmed the temperature and photoperiod in the chamber based on that of the *Lip-0* ecotype that originated from southern Poland, representing a common field environment, following Lorts and Lasky (2020). Each tray represented either a drought or well-watered treatment, and seven replicates per mutant were randomly assigned positions for each treatment. Seeds were stratified in DI water for 4-5 days prior to planting.

All pots were well-watered by maintaining 2.5-5 cm of water at the bottom of each tray prior to drought treatments. Drought treatments began 19 days after planting when all seedlings had at least the first true leaves expanded. We removed the bottom water in the drought trays while maintaining a constant 2.5 cm of water in the well-watered trays. To prevent the topsoil from drying, well-watered pots were top-watered every other day with 2.5 ml of water. The soil in drought treatment plants was allowed to dry for five weeks after the bottom water was removed, then we top-watered each pot in both well-watered and drought treatments with 12ml 2x strength miracle grow solution every other day for three treatments total to maintain adequate nutrition. Trays were rotated and moved to a different bench in the growth chamber every other week. Percent volumetric water content was monitored every 30 minutes throughout the experiment using 3 drought-treatment pots and 2 well-watered pots containing *Columbia* (*Col*) and 5TE probes (METER Group, Inc., Pullman WA, USA).

Plant harvest and phenotyping at maturity

The fecundity of Arabidopsis plants that survived to reproduce was measured when all siliques reached maturity (dried and brown) and the rosette leaves had senesced. For each Arabidopsis plant, we measured inflorescence height and the number of secondary inflorescence branches (inflorescence branches arising from the inflorescence main axis). We also measured the total silique number and silique length of six siliques corresponding to 10, 20, 40, 60, 80, and 90th percentiles of silique positions along the inflorescence to represent silique length across the entire inflorescence (e.g. the 10th percentile silique was higher up the inflorescence than 10% of all siliques). After siliques were measured, we dried the mature inflorescence and rosette at 60 °C for 24 hours, weighed them separately, and then added their weights to get aboveground dry biomass. Finally, the rosette tissue was ground and used for $\delta^{13}C$ and $\delta^{15}N$ isotope analysis at the UC Davis Stable Isotope Facility.

Follow-up drought and freezing experiments

In our main drought screen, we found that *wrky38* had significant GxE effects on several fitness-related traits, and *lsd1* had a significant GxE effect on flowering time. Both *WRKY38* and *LSD1* were top candidates associated with moisture in their originated studies (Lasky et al., 2018, 2012) Additionally, *WRKY38* was also found most strongly associated with winter low temperatures in the combined common garden study (Lasky et al., 2018), while knockouts of *LSD1* were found to affect chilling sensitivity (Huang et al., 2010) and allelic variation was associated with the second principal component of climate (moisture related, Lee et al., 2017) in *Arabidopsis*. Both genes happen to play a role in the salicylic acid (SA) pathway (Bernacki et al., 2019; Kim et al., 2008; Szechyńska-Hebda et al., 2016; Wituszynska et al., 2013), primarily known for microbial defense response but which can be activated by abiotic stress (reviewed in Miura and Tada, 2014; Wu et al., 2019). Incorporating previous studies and our results, we decided to use the two mutants to follow up with intermediate drought and freezing experiments, in order to test their possible adaptation in different drought and cold regimes in terms of their ecophysiological responses.

Drought experiment on *wrky38* and *lsd1*

We applied a constant temperature setting (20/14℃, 12h/12h) throughout this experiment and an intermediate drought by controlling the frequency that pots were saturated. Specifically, we bottom-watered pots every 3 days in the well-watered (WW) treatment while reducing the watering frequency for pots under drought (D) treatment to every 6 days starting Day 25, when the drought treatment started. At this time point, most plants had 10 leaves, which corresponds to the adult vegetative phase in *Col* (Lawrence-Paul et al., 2023). We grew 12 replicates of *Col*, *wrky38*, and *lsd1* each for WW and D treatments. On Day 40, we randomly picked 4 pots per genotype per treatment for destructive measurements. Rosettes were first cut and weighed for fresh weight, and then each of them was put into a petri dish with distilled water under 4 ℃ in the dark overnight to weigh their turgid weight. Next, all the fully expanded leaves were cut, scanned, and then dried separately with the rest of the rosette tissue in an oven at 60 ℃ for 2 weeks before leaf and rosette dry weights were both weighed. The total leaf area of each plant was analyzed by ImageJ, and the Specific Leaf Area (SLA) of each plant was calculated by (Total leaf area/Total leaf dry weight). Relative Water Content (RWC) was calculated by (Rosette fresh weight - Rosette dry weight) / (Rosette turgid weight - Rosette dry weight). We kept the dried materials and sent samples of lines that showed interactions with drought for $\delta^{13}C$ isotope analysis, an estimate for water use efficiency (WUE) along with leaf C:N and d15N.

Before the first plant flowered, we tracked stomatal conductance (g_{sw}) and Fv/Fm for 10 continuous days using a Li-600 porometer and fluorometer. Measurements were taken approximately one hour before lights were turned off. On every measurement, 4 pots of each genotype from each treatment were randomly picked with measurements conducted on the latest fully expanded leaves that were big enough to apply the device.

We recorded the flowering time for each non-destructively sampled plant. When all plants were matured and started senescing, we counted silique numbers and measured

inflorescence length and the lengths of 12 siliques (or all siliques if fewer than 12) from the top, middle, and bottom of the inflorescence to calculate the average silique length. Finally, we calculated the total silique length as the product of silique number and average silique length. Representing reproduction level, silique number, average silique length, and total silique length were considered fitness estimates.

Freezing experiment on *wrky38* and *lsd1*

To mimic the natural freezing stress that *Arabidopsis* could undergo in its growing season, we set the night (12h) temperature in the growth chamber to -2℃ and the day temperature to 10℃ as our freezing treatment, which started on Day 30 when plants were established and had been under cold acclimation at 10/4℃ (12/12h) for 2 weeks. We surrounded all pots with extra soil for insulation to mimic the circumstances in nature, preventing unnaturally cold soil temperatures.

On day 60, we measured diameter as an estimation of growth under freezing. Starting from Day 64, we shifted the temperature back to 20/14℃ until the end of the experiment to allow plants to flower and then recorded flowering time. We grew the plants until senescence, after which we measured aboveground biomass, silique numbers, inflorescence length, and average and total silique lengths as described above. We grew 12 replicates for each genotype and all of them survived until the end of the experiment.

Statistical analyses

In both the drought experiments, we were primarily interested in whether the knockouts had genotype-specific drought responses, i.e., whether there were genotype-by-drought interactions, as well as whether the knockouts altered average phenotypes under drought, i.e., whether genotype effects were significant. Thus, we applied linear mixed models (LMMs) to test traits we measured between *Col* and each mutant separately, considering genotypes as fixed effects and tray numbers as random effects. We also applied LMMs on the daily g_{sw} and Fv/Fm data to detect if the mutants differed with Col in ecophysiology on any of the 10 measuring days.

It was not feasible to control moisture in the freezing experiment as we did in the drought follow-up because soil lost water much more slowly than it did in higher temperatures, so we used *DunnettTest()* function in the R package *DescTools* (Signorell et al., 2023) to conduct Dunnett's Test on traits measured in the freezing experiment only, using *Col* as control and test for differences of the two mutants to control separately.

Functional variation in the screened loci in the 1001 Genomes

We analyzed putative functional variation (hereafter FV) in the 1001 Genomes resequencing data for all genes screened. We used the 1001 Genomes Polymorphism Browser (https://tools.1001genomes.org/polymorph/) to identify potential large-effect mutations in each gene, which were considered as putative lost-of-function (LoF) mutations. If more than one isoform existed for a gene, we recorded the case of the most common isoform except when no functional variant was found for this isoform, we instead used the most common isoform that

had functional variants. We summarized the counts of each type of functional variation and the positions where the variations occurred.

To follow up on the published GEA results, we asked whether the putative functional variants were associated with climate regimes of their origins. To do the latter we used mixed model tests of association that included random effects correlated according to a kinship matrix estimated from genome-wide identity by state among genotypes (Zoubarev et al., 2012).

Population genetic variation at *WRKY38*

Because we found some putative LoF for *WRKY38* were relatively common in the 1001 Genomes accessions, we further analysed the geographic and genetic structure of these variants. We used the imputed SNP matrix (hdf5 file) and short indels (vcf file) downloaded from the 1001 Genome Project website (https://1001genomes.org/accessions.html) for the following analyses.

We calculated Tajima's D across chromosome 5 (Chr 5) in 5-kb windows using VCFtools to test possible selection signals on *WRKY38* in natural accessions. The most common functional variant for *WRKY38* is a frameshift at position 7495793 on Chr 5 ($\text{Frameshift}_{7495793}$; Table S2).To detect linkage disequilibrium and potential haplotype structure between $\text{Frameshift}_{7495793}$ and adjacent SNPs, we calculated the square of Spearman correlation coefficient (r^2) between $\text{Frameshift}_{7495793}$ and each SNP within 5 kb upstream and downstream of *WRKY38* using the *cor()* function in base R.

To characterize the geographic and population structure of natural putative LoF mutations in 1001 Genomes accessions, we plotted mutations and their genetic clusters by different shapes and outline colors on a map. We also constructed neighbor-joining (NJ) trees built with the entire gene sequence of *WRKY38* (including coding and non-coding sequences), and a genome-wide NJ tree. We referred to the population structure of 1001 Genomes accessions in Alonso-Blanco et al. (2016), where the authors classified 9 genetic clusters and one admixed group. For simplicity of data presentation, we combined the 9 groups by 5 major regions — West Mediterranean (W-Med), North Mediterranean (N-Med), Central and Western Europe (CW-Euro), North Europe (N-Euro), and Asia. Accessions that were introduced, classified in the admixed group, or might be potential contaminations (Pisupati et al., 2017) were not included in the 5 groups.

We used the R package *maps* to visualize the distribution of the common putative LoF mutations of *WRKY38*. To build the genome-wide tree, we annotated the vcf file using SnpEff (Cingolani et al., 2012), extracted the synonymous sites, filtered out sites with minor allele frequency below 0.01 (--maf 0.01) and missing rate over 10% (--max-missing 0.9), and generated distance matrix using PLINK. Both NJ trees were constructed using the *nj()* function from package *ape*.

We hypothesized that the natural *WRKY38* putative LoF variants could underlie local adaptation. To identify which climate variables might best explain *WRKY38* functional variation and be most closely tied to mechanisms of selection, we used LMMs to scan for correlations across the 19 bioclimate variables. Models were implemented using *coxme* and *kinship2* packages to include kinship in the models as random effects. We focused on testing the local adaptation of intact *WRKY38*, $\text{Frameshift}_{7495793}$, the most common and widespread FV, and

Frameshift$_{7495793\&94}$, which accounted for a 3-bp frameshift that led back to the normal ORF thus we suspected to be potentially functional. Since mixed models have low power when causal variants are correlated with the genomic background, we also used Welch's t-test in R to compare the differences in bioclimate variables for accessions with/out the intact gene or the FVs.

Variation in *LSD1* expression

Only 5 of the total 1135 accessions had *LSD1* putative LoF mutations. Thus we suspected that LoF may not be the major cause of the strong association between climate and the *LSD1*-targeting SNP in Lasky et al. (2018), and the adaptation might exist due to amino acid changes or *cis*-regulatory variation impacting expression. Therefore, we scanned across 19 bioclimate variables and SNPs within 10kb around *LSD1* using univariate GEAs with GEMMA (Zoubarev et al., 2012) for 933 native accessions with known coordinates that were not potential contaminants (Pisupati et al., 2017). We found elevated correlations with climate at SNPs 1kb upstream *LSD1* and the start of the gene sequence, suggesting that the adaptation may occur at the promoter region. Thus, we further asked if these SNPs impact *LSD1* expression using the published 1001 genome transcriptome data (Kawakatsu et al., 2016). We compared the expression levels of *LSD1* between the two alleles of the top climate-associated SNP in the putative promoter region by Welch's t-test in R.

RESULTS

Putative functional variation in the experimentally screened genes in the 1001 Genomes accessions

Across all 1,135 accessions, 1973 putative frameshifts occurred over the 44 experimentally screened genes in 943 genotypes as well as 1116 stop-gained FVs in 842 genotypes, appearing to be the most prevalent types of putative high-impact variants (i.e. putative LoF) over the selected loci (Table S2). Splice donor variants were found in 303 accessions, and splice acceptor, stop-lost, and start-lost were found in 29, 5, and 4 accessions, respectively.

Putative frameshifts were most common in AT1G19410 (*FBD*), occurring in 639 accessions. Frameshift variants were also common in AT5G22570 (*WRKY38*), detected in 250 accessions. Stop-gained had the most counts in AT5G22900 (*CHX3*) with 492 accessions having the putative FVs. Splice-donor was also most prevalent in the *FBD* gene, exhibited by 220 accessions. The rest of the high-impact effects were rare among the screened loci, suggesting purifying selection on amino acid sequences. We found 14 accessions with splice-acceptor variants, 2 accessions with stop-lost, and 2 accessions with start-lost at AT5G10960 (*CAF1I*), AT3G45860 (*CRK4*), and AT5G22570 (*WRKY38*), respectively.

Some mutations in the screened genes had potential dual effects and such dual effects were locus-specific. A 1-bp deletion or insertion at the same position in AT1G76090 (*SMT3*) caused both frameshift and start-loss, resulting in the most common dual effects observed in 95 accessions. Two insertions in AT3G45860 (*CRK4*) could each lead to dual effects of frameshift

and stop-gain, together occurring in 21 accessions. In rare cases, a 1-bp deletion in AT5G40830 (*ICA*) caused frameshift and stop-lost in 4 accessions, while a 3-bp deletion in AT5G65050 (*AGL31*/*MAF2*) caused stop-lost and disruptive-inframe-deletion effects in 1 accession.

Main drought screen

Drought treatment effects on phenotypes were strong and largely consistent across different t-DNA insertion lines (Figure 1). Specifically, the drought effects were significant with FDR=0.05 across all 44 tested lines, causing lower aboveground biomass, smaller inflorescence length and weight, fewer siliques, shorter average and total silique length, and delayed flowering, with line CS68738 (*lsd1*) being the only exception with no significant drought effect on flowering time (Figure 1). The drought effect on rosette weight was also common, significantly reducing rosette weights in 25 of 44 lines at FDR = 0.05.

Genotype effects varied across the 44 lines tested (Figure 1). Genotype effects were most common for flowering time, with 20 having nominally significant effects ($\alpha = 0.05$) on flowering, primarily accelerating flowering, except for *wrky38*, which exhibited delayed flowering under drought. Thirteen insertion lines were significant with FDR = 0.05 (12 accelerated flowering and 1 delayed flowering (*wrky38*). Rosette and inflorescence weights were the traits with the next most insertion lines with genotype effects, 14 lines having nominally significant effects and 5 significant at FDR = 0.05 that mostly led to increased rosette weight and reduced inflorescence weight. The effects of genotype on other traits measured were less common, with 5,5,5,3,1 lines having nominally significant effects and 1,0,0,1,1 lines significant with FDR = 0.05 for silique number, average and total silique length, inflorescence length, and aboveground biomass, respectively.

We found few genotype-by-treatment interactions caused by the t-DNA insertions. CS864818, a *WRKY38* (AT5G22570) knockout line, was the only one having a GxE effect that was significant at FDR = 0.05 on inflorescence length ($p_{adjusted} = 0.0274$, $\beta_{wrky38\times D} = 23.21$), silique numbers ($p_{adjusted} = 0.0024$, $\beta_{wrky38\times D} = 116.88$), total silique length ($p_{adjusted} = 0.0076$, $\beta_{wrky38\times D} = 1314.72$), aboveground biomass ($p_{adjusted} = 0.0189$, $\beta_{wrky38\times D} = 0.14$), and inflorescence weight ($p_{adjusted} = 0.0164$, $\beta_{wrky38\times D} = 0.13$). Generally, *wrky38* showed less drought sensitivity than *Col* while being smaller and less fecund in well-watered conditions except that the mutant had slightly longer inflorescence lengths than *Col* (*wrky38*: 40.1±3.010 cm; *Col*: 36.0±0.713 cm; Figure 2). Considering only genotype effects, the knockout of *WRKY38* caused significantly later flowering ($p_{adjusted} = 0.0002$, $\beta_{wrky38} = 15.86$), longer inflorescence ($p_{adjusted} = 0.0014$, $\beta_{wrky38} = 27.33$), more silique ($p_{adjusted} = 0.0344$, $\beta_{wrky38} = 14.36$), and greater rosette weights ($p_{adjusted} = 0.0190$, $\beta_{wrky38} = 0.05$, Figure 2).

Another mutant, *lsd1*, was one of only two having nominally significant genotype-by-treatment interaction on flowering time ($p = 0.0015$; the other had $p = 0.0479$). Notably, the mutant flowered earlier than *Col* under both WW and drought conditions (WW: 58.1±1.06 vs. 58.6±0.972 days; drought: 52.3±3.58 vs. 68.1±1.94 days) and the effect of drought on flowering time was not significant ($p = 0.3981$). The mutation significantly accelerated flowering compared to *Col* ($\beta_{lsd1} = -15.86$), and its interaction with drought exacerbated early flowering ($\beta_{lsd1\times D} = -15.43$). *lsd1* plants had significantly lighter inflorescence ($p = 0.0002$, $\beta_{lsd1} = -0.043$) and rosette

weights (p = 2.40e-6, β_{lsd1} = -0.009), thus lighter aboveground biomass (p = 2.14e-6, β_{lsd1} = -0.1328), compared to *Col* (Figure 2). Only drought had significant effects on fitness-related traits that reduced silique numbers (p =1.76e-11, $\beta_{drought}$ = -104.71), average silique length (p = 0.0115, $\beta_{drought}$ = -2.43), and total silique length (p = 5.28e-12, $\beta_{drought}$ = -1200.26).

Both genotypes had increased WUE under drought compared to WW conditions (Figure 1), with treatment effects on WUE significant in LMMs that include *Col* and each mutant (p_{wrky38} = 9.33e-18, p_{lsd1} = 1.15e-6). The genotype effect of *WRKY38* knockout also significantly increased its WUE compared to *Col* (p = 8.81e-8), while that of *lsd1* was not significant (p = 0.756; Figure 2). *wrky38* had an increased δ15N ratio under drought compared to the well-watered condition but its responses under both conditions were not substantially different from *Col* ($p_{Genotype}$ = 0.5010, $p_{Treatment}$ = 0.5575; Figure 2). *lsd1* mutants, however, had $\delta^{15}N$ being lower under drought while slightly higher under WW compared to *Col* ($p_{Genotype}$ = 0.0387, p_{GxE} = 4.22e-4) suggesting a possible nitrogen source shift under drought. C:N ratios had been overall reduced under drought, probably the result of reduced biomass under drought (Figure 2).

Figure 1. Reaction norms (lines) and average trait values of each genotype (boxes) of partial traits measured in the main drought screen of all 44 mutants and *Col*, with *Col*, *lsd1* (CS68738), and *wrky38* (CS864818) highlighted. Each line connects the average trait values of one genotype under well-watered (WW) and drought (D) treatments.

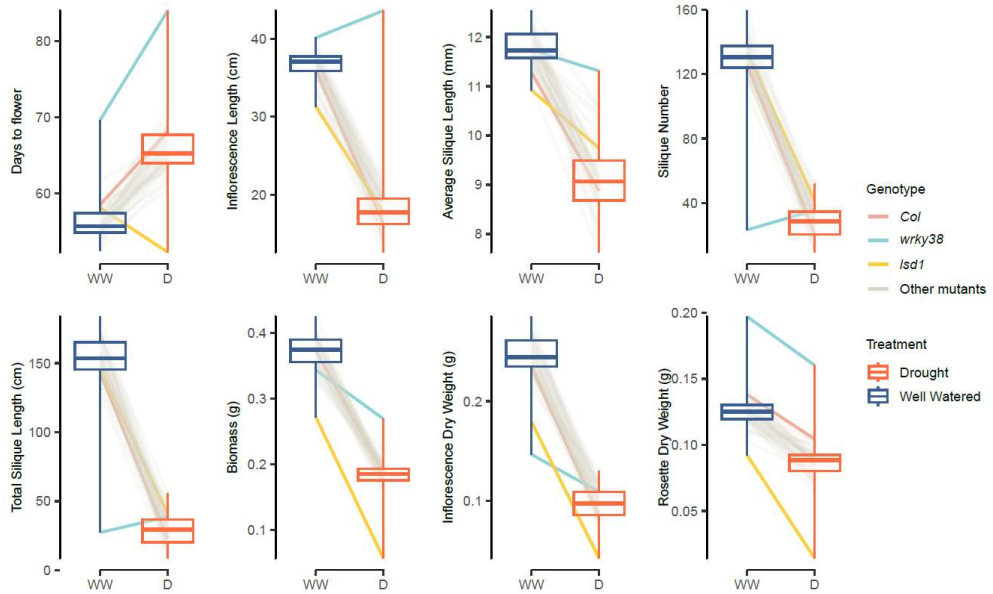
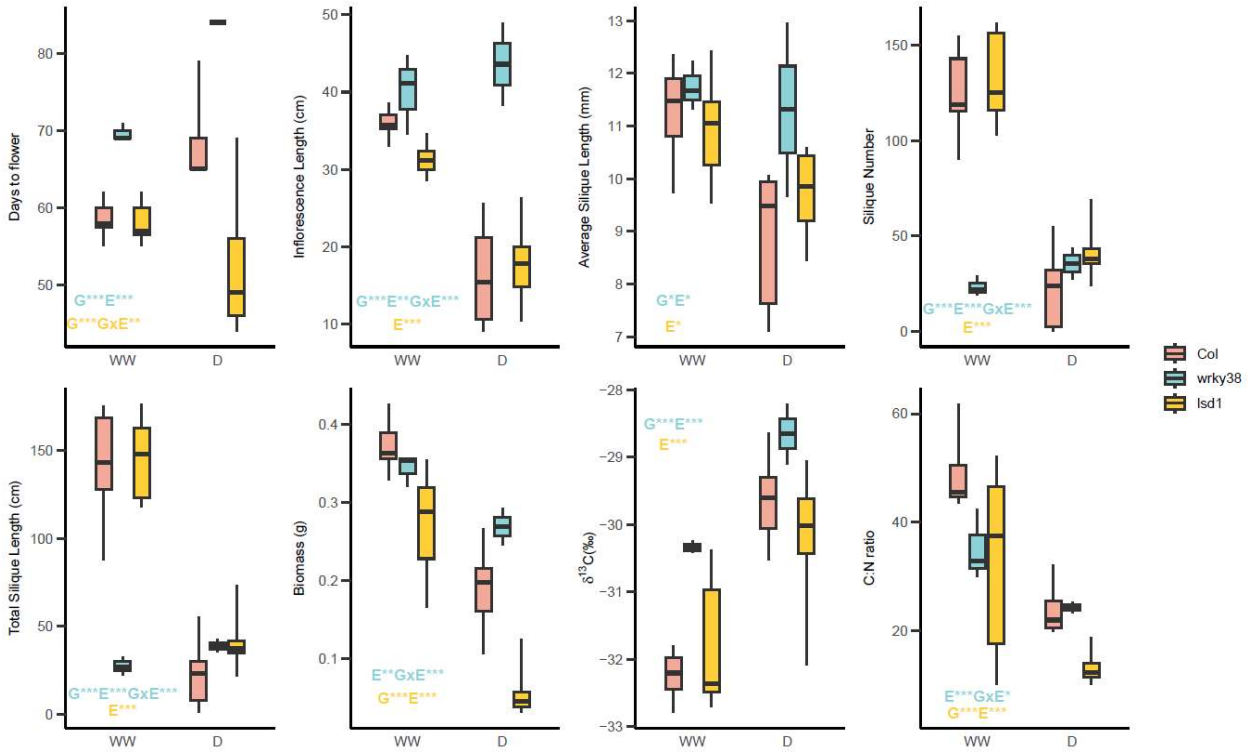


Figure 2. Part of the traits measured during the main drought screen of *Col*, *lsd1* (CS68738), and *wrky38* (CS864818). Text annotations on the box plots indicate the significance of genotype (G), treatment (E), and genotype-treatment interaction (G×E) effects at α = 0.05, based on LMMs comparing *Col* and one mutant genotype. The text color denotes the mutant genotype. Significance levels are represented as follows: ns, p>0.05; *, 0.01<p<0.05; **, 0.001<p<0.05; ***, p<0.001.



Follow-up drought experiment

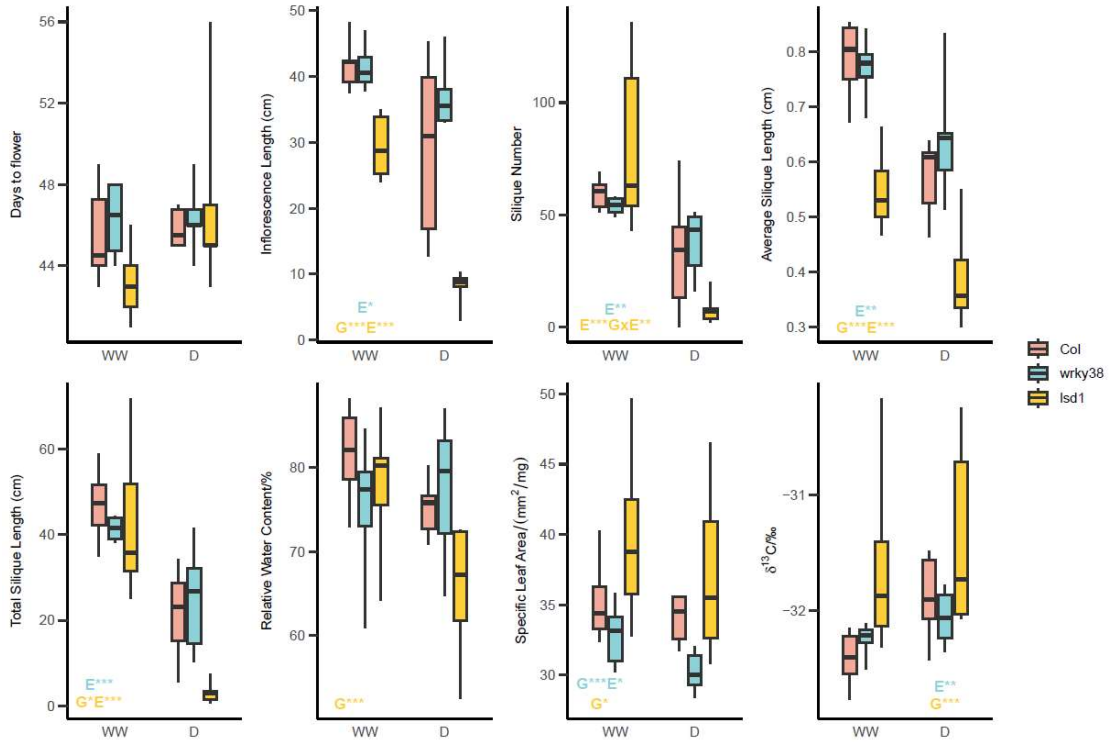
The mild drought significantly reduced fitness and inflorescence length across both *Col* and the mutants, as determined by mixed linear models that included *Col* and one mutant (*wrky38* or *lsd1*). Specifically, the drought significantly reduced fitness and inflorescence length for both *wrky38* (Silique number: $p_{Treatment}$ = 0.002, $\beta_{Drought}$ = -26.7; Average silique length: $p_{Treatment}$ =0.002, $\beta_{Drought}$ = -0.2; Total silique length: $p_{Treatment}$ = 1.0e-6, $\beta_{Drought}$ = -25.7) and *lsd1* (Silique number: $p_{Treatment}$= 7.1e-4, $\beta_{Drought}$ = -26.7; Average silique length: $p_{Treatment}$ =1.1e-4, $\beta_{Drought}$ = -0.2; Total silique length: $p_{Treatment}$ = 1.4e-11, $\beta_{Drought}$ = -25.7), while the mixed models also account for effects on *Col* in the same analyses. Flowering time, however, was not significantly impacted (*wrky38*: $p_{Treatment}$ = 0.97; *lsd1*: $p_{Treatment}$ = 0.19) (Figure 3). Plants of *wrky38* mutants did not exhibit significant genotype-specific responses to drought compared to *Col* across the traits we measured (Flowering time: $p_{Genotype}$ = 0.15; Inflorescence length: $p_{Genotype}$ = 0.24; Silique number: $p_{Genotype}$ = 0.96; Average silique length: $p_{Genotype}$= 0.50; Total silique length: $p_{Genotype}$ = 0.88). However, the direction of *wrky38* differences was consistent with the earlier drought screen; e.g. *wrky38* had higher fitness and biomass than wild type *Col* under drought, but the opposite under well-watered (Figure 3). In contrast, lsd1 mutants showed significant genotype effects for inflorescence length (p = 2.50e-7, β_{lsd1} = -21.1) and average silique length (p = 3.15e-11, β_{lsd1} = -0.18) and a significant GxE effect for silique number (p = 0.0081, $\beta_{lsd1\times D}$ = -48.5). This indicates that lsd1 mutants not only have reduced fitness under drought conditions, but the negative impact of drought was even more pronounced in lsd1 mutants compared to *Col*.

During the 10 days of ecophysiological tracking, we built LMMs to compare the daily g_{sw} and Fv/Fm of the two mutants with *Col*. The stomatal conductance (g_{sw}) of all three genotypes exhibited fluctuations that were associated with the regular watering schedule (Figure S1).

wrky38 had potentially adaptive reduced g_{sw} compared to *Col* and under drought, with genotype effects significant during the initial days of tracking, when drought-treated plants first experienced water limitation (9 days after treatment initiation; Table S3). A significant GxE interaction with drought was also observed (Table S3). The *lsd1* mutant also had reduced g_{sw} under drought, but the genotype effect was insignificant (Figure S1; Table S3). None of the Fv/Fm values we measured suggested major oxidative stress (Figure S2), coinciding with Zivcak et al. (2013), indicating a mild drought. No genotype or treatment pattern was detected for Fv/Fm.

The drought effect significantly reduced SLA when considering *Col* and *wrky38* (LMM, p = 0.0313) but was not significant in the model including *Col* and *lsd1* (Figure 3). Genotype effects on SLA were significant for both mutants, but *wrky38* had a lower SLA (potentially adaptive under drought) while *lsd1* had a higher SLA compared to *Col* (Figure 3). RWC exhibited little variation between genotypes and treatments except for lsd1 plants, which had a significant genotype effect resulting in lower RWC (p = 1.9e-4, β_{lsd1} = -0.1). The $\delta^{13}C$ response to treatment was generally lower in the follow-up than in the main drought screen, indicating a milder drought in the follow-up. The treatment effect on WUE was only significant for *wrky38* (p = 0.0029), whereas the genotype effect on WUE was only significant for *lsd1* (p = 3.7e-4). No significant GxE effects were detected for SLA, RWC, or $\delta^{13}C$.

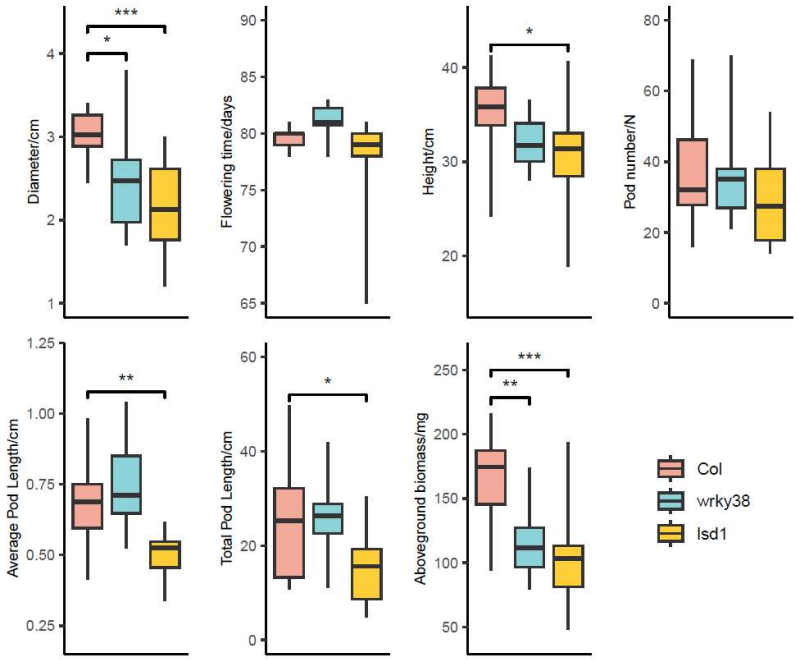
Figure 3 Traits measured in the follow-up intermediate drought experiment of *Col*, *lsd1* (CS68738), and *wrky38* (CS864818). Text annotations on the box plots indicate the significance of genotype (G), treatment (E), and genotype-treatment interaction (G×E) effects at α = 0.05, based on LMMs comparing *Col* and one mutant genotype. The text color denotes the mutant genotype. Significance levels are represented as follows: ns, p>0.05; *, 0.01<p<0.05; **, 0.001<p<0.05; ***, p<0.001.



Follow-up freezing experiment

Treated with overnight freezing for 30 days, both *lsd1* and *wrky38* had significantly smaller diameters than *Col* (p_{wrky38} = 0.0184, p_{lsd1} = 3.7e-4), representing their reduced vegetative growth under low temperatures (Figure 4). After returning to warmer temperatures, we found no variation in flowering time and silique number between *Col* and the two mutants. However, *lsd1* had a shorter inflorescence length (p = 0.0360), average silique length (p = 0.0023), and total silique length (p = 0.0218) than *Col*. Both *lsd1* and *wkry38* had smaller biomass than *Col*(p_{wrky38} = 0.0026, p_{lsd1} = 2.6e-4), which might also be due to reduced vegetative growth. Overall, *lsd1* plants had reduced growth and fitness compared to *Col* under freezing, consistent with findings of Huang et al. (2010) that *lsd1* is chilling sensitive. Although putative LoF variants were strongly associated with warm winter temperatures (Lasky et al. 2018), *WRKY38* knockouts did not exhibit fitness differences compared to *Col*, though their vegetative growth was inhibited by overnight freezing compared to *Col*.

Figure 4 Traits measured in the follow-up freezing experiment of *Col*, *lsd1* (CS68738), and *wrky38* (CS864818). Asterisks indicate significant trait differences between *Col* and each mutant based on Dunnett's test: *, 0.01<p<0.05; **, 0.001<p<0.05; ***, p<0.001.



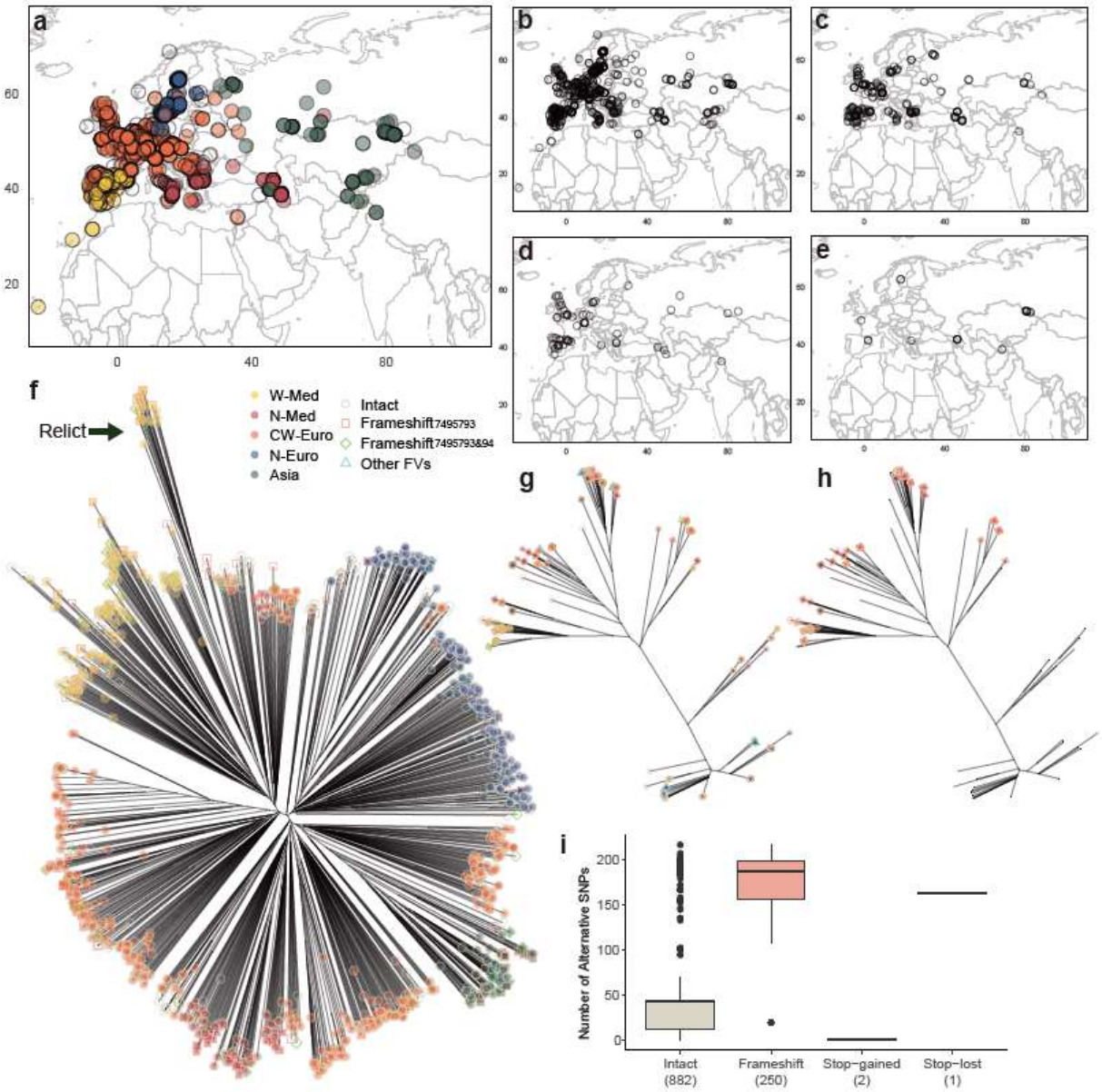
WRKY38 allelic variation

To evaluate potential selection on *WRKY38* and its surrounding genomic region, we analyzed SNP diversity patterns using Tajima's D. The Tajima's D of the 5-kb window containing the *WRKY38* gene was above 0 (Tajima's D = 1.42852) and is above the average of Chromosome 5 (average Tajima's D = 0.660), but it did not significantly deviate from the distribution of

Tajima's D across the chromosome (*z*-score = 0.659; Figure S3). The overall r^2 between SNPs in 10 kb around *WRKY38* and the most common functional variant (Frameshift$_{7495793}$) also indicated some haplotype structure, with the frameshift having elevated correlations (r^2 >0.4) for SNPs located within 2000 bp upstream of *WRKY38* and in the first 500 bp of the gene (where the frameshift was located, maximum r^2=0.5104; Figure S4). However, this was not a dramatically long region of linkage, suggesting no evidence for a recent sweep (cf. the strong correlation observed with chr4:10999188 and other SNPs in a 300 kb window around *LSD1,* Lee et al., 2017). Nevertheless, the putative LoF mutations carried large numbers of alternate SNP alleles across the locus, indicating some divergence and linkage (Figure 5i; Figure S5). The haplotype structure at *WRKY38* can also be seen in the *WRKY38* gene neighbor-joining tree. While the genome-wide NJ tree shows a star-like structure consistent with the recent expansion of Arabidopsis (Lee et al. 2017) with accessions clustered based on geography (Figure 5f), the *WRKY38* tree shows strong divergence between sequences with Frameshift$_{7495793}$ versus putative intact *WRKY38* sequences. This allelic variation is segregated across diverse lineages including the relicts (Figure 5a,b,c,f). Coincidentally, the rare FVs were all unique to their lineages, suggesting recent origins (Figure 5f; Table S3). In contrast to the widespread Frameshift$_{7495793}$, they might reflect newer local putative LoF mutations.

Although most N-Euro accessions are in the non-Frameshift$_{7495793}$ clade, all 12 N-Euro accessions having Frameshift$_{7495793}$ are in the Frameshift$_{7495793}$ clade, overlapping with W-Med, CW-Euro, and Asian accessions carrying this allele. The only accession having Frameshift$_{7495794}$ and the only one with Stop-lost$_{7495609}$ are both within Frameshift$_{7495793}$ clade, while the 6 accessions having Frameshift$_{7496041}$ and 2 accessions having Stop-gained$_{7496443}$, which are grouped at the same tip on the gene tree, only occurred in non-Frameshift$_{7495793}$ clade.

Figure 5 Natural variation of *WRKY38* putative LoF in the 1001 Genomes accessions. **a**, distribution of native *Arabidopsis* accessions from 1001 Genomes. Colors indicate ADMIXTURE genetic clusters based on Alonso-Blanco et al. (2016). **b**-**e**, distributions of natural *WRKY38* putative functional variation: **b**, distribution of accessions with intact *WRKY38*; **c**, distribution of accessions with Frameshift$_{7495793}$, the most common FV of *WRKY38* ; **d**, distribution of accessions with Frameshift$_{7495793\&94}$, a mutation that restores the normal open reading frame; **e**, distribution of accessions with other rare FVs of *WRKY38*. **f,** *WRKY38* FVs on the genomewide NJ tree, the relict lineage (Alonso-Blanco et al., 2016) was marked by an arrow. **g**, *WRKY38* FVs on the NJ gene tree, symbols as in (f). **h,** Frameshift$_{7495793}$ on the NJ gene tree. **i**, numbers of alternative SNPs in accessions with intact *WRKY38* and putative LoF FVs within 10 kb around *WRKY38*. The number of accessions with each type of functional variation is indicated in parentheses.



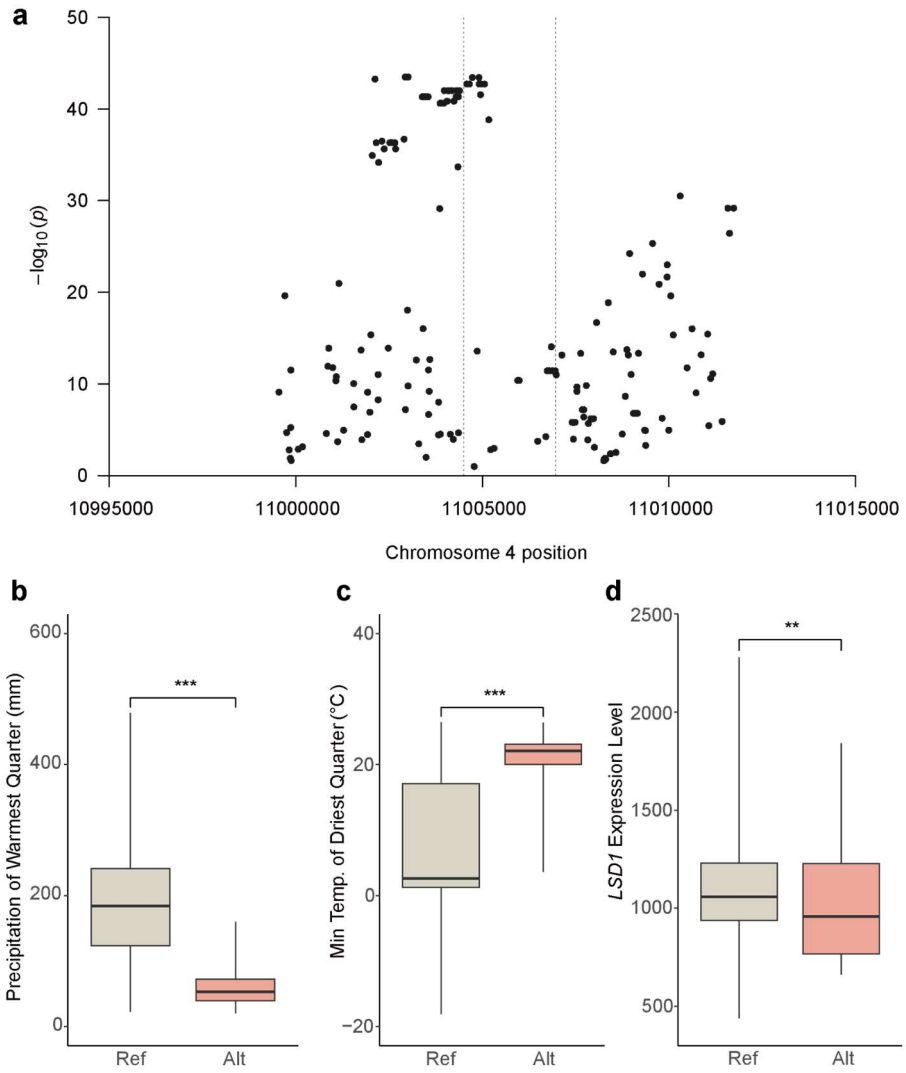
Our drought experiments showed evidence that *WRKY38* LoF alleles are drought-adapted. We next sought to identify which climate variables were most closely associated with putative LoF allele frequency in *WRKY38*. We did not find significant correlations at $\alpha = 0.05$ between the frequency of $\text{Frameshift}_{7495793}$ (the most common FV) or $\text{Frameshift}_{7495793\&94}$ (the FV that putative restores the normal ORF) and the bioclimatic variables when considering kinship (Table S4), which coincided with the findings of Alonso-Blanco et al. (2016) that almost no SNPs are significantly correlated with climate in environmental GWAS at FDR = 0.05 after considering population structure for 1001 Genomes accessions. Despite this, without accounting for kinship, significant differentiation was detected at $\alpha = 0.05$ for 11, 12, and 4 bioclimatic variables between accessions with and without intact *WRKY38*, $\text{Frameshift}_{7495793}$, and $\text{Frameshift}_{7495793\&94}$, respectively (Table S5). Among the significant differentiations, precipitation of the warmest quarter (bio18) had the lowest p-values for both intact *WRKY38* (p = 6.8e-10) and

Frameshift$_{7495793}$ (p = 8.8e-9) but in opposite directions. Specifically, intact *WRKY38* occurred more frequently in regions with wet summers while Frameshift$_{7495793}$ was more common in hot and dry regions, suggesting the drought adaptation of *WRKY38* putative LoF. Frameshift$_{7495793\&94}$ was most significantly correlated with higher isothermality (bio3; p = 3.7e-4) followed by higher mean temperature of the wettest quarter (bio8; p = 6.4e-4).

LSD1 allelic variation and expression

We also tested for climate correlations with *LSD1* SNPs. Mixed models (GEMMA) detected strong correlations between SNPs within 10kb around *LSD1*, particularly for SNPs within 1 kb upstream and the first 500 bp of *LSD1*, and multiple bioclimate variables (Figure 6a). We believed this suggested these SNPs might contribute to local adaptation via changes in gene expression. To validate the hypothesis, we picked a top climate-associate SNP (Chr4, position 11003866) from the 1 kb upstream region of *LSD1* and tested for *LSD1* expression differences between ref/alt alleles. This SNP had the strongest correlation, indicated by the lowest p-value, with precipitation of the warmest quarter (bio18; wald-p = 2.390e-41; Figure 6b) followed by the minimum temperature of the driest quarter (bio9; wald-p = 2.140e-36; Figure 6c). We found that the alternative allele corresponded to significantly lower *LSD1* expression compared to the reference allele (p = 0.003; Figure 6d), suggesting a connection between lower *LSD1* expression and hot and dry climates. Using linear models, we found that *LSD1* expression was significantly greater for genotypes from wetter summer climates (precipitation of the warmest quarter (p=8.683e-5) and precipitation of the wettest month (p=0.039)). Across the 10 kb region around *LSD1*, 68 of the 201 SNPs were most strongly associated with bio18 (precipitation of the warmest quarter). Moreover, when limiting to SNPs within 1kb upstream of *LSD1*, i.e. the region we thought to be linked with *LSD1*, the ratio became 25 of 35 SNPs. The results indicated that *LSD1* might play a role in local adaptation to moisture regimes by *cis*-regulatory variation.

Figure 6 Genotype-environment associations and *LSD1* expression. **a**, Scanning the *LSD1* region to identify potential SNP-environment associations using LMM within 10kb. Each point represents the smallest p-value from the 19 associations between each SNP and the 19 bioclimate variables. *LSD1* gene region is shown between the dashed lines. These associations are meant to identify where in the locus variation is most strongly associated with climate. **b**, the difference in precipitation of the warmest quarter (bio18) between reference and alternative alleles at a top climate-associated SNP (Chr4, position 11003866). **c**, the difference in the minimum temperature of the driest quarter (bio9) at an SNP (Chr4, position 11003866). **d**, the difference in *LSD1* expression at an SNP (Chr4, position 11003866). **b-d**, asterisks indicate significant trait differences from t-test: *, 0.01<p<0.05; **, 0.001<p<0.05; ***, p<0.001.



DISCUSSION

Validation of genotype-environment associations

Local adaptation is a major source of phenotypic variation within species, but traditional methods of identifying locally adapted genes are logistically challenging. Genotype-environment associations (GEA) offer an easy-to-implement tool to generate hypotheses about genes involved in local adaptation, but most published results remain as untested hypotheses. Here we screened 42 genes identified from genome-wide GEA studies and found at least two had effects on environmental responses and physiology linked to drought adaptation, suggesting potential local adaptations.

In our main drought experiment, among 42 potential drought-adapted genes detected with 3 different GEA approaches, the tDNA knockout of *WRKY38* (detected with RDA, Lasky et al., 2012) exhibited a significant GxE effect on fitness-related traits, and *lsd1* mutant (gene detected with combined GEA-common garden synthesis, Lasky et al., 2018) had a significant GxE effect on flowering time, which could potentially indicate variation in drought escape versus drought avoidance (Lovell et al., 2013). Coincidentally, these genes were both the top (#1) candidate genes with the strongest GEA in the studies used to select them. The physiological effect of *wrky38* identified in the follow-up screening suggests mechanisms of drought avoidance for *wrky38*, via reduced stomatal conductance and SLA. Furthermore, the high frequency of putative natural loss of function alleles of *WRKY38* in Arabidopsis populations found in the drier southern part of its range suggests potential local adaptation.

Since the advent of GEA a wide variety of methods have been proposed (reviewed by Rellstab et al., 2015). We tested candidates from three approaches here: multivariate ordination (RDA) of genotype and environment (Forester et al., 2018; Lasky et al., 2012), univariate climate mixed model associations (Lasky et al., 2014), and univariate climate models integrated with fitness data from multiple common gardens (Lasky et al., 2018). If the two candidates *WRKY38* and *LSD1* are taken as positives, the small number of positives remains too small to definitely compare these statistical GEA approaches. While Lotterhos (2023) recently criticized RDA as inappropriate for GEA due to a high false positive rate, the author used arbitrary and uncalibrated significance thresholds. By contrast, Lotterhos (2023) showed that the AUC-PR of RDA was higher than most methods. We propose that using ranked GEA results is more informative than focusing on arbitrary significance thresholds; as mentioned previously *WRKY38* and *LSD1* were the top-ranked loci in their respective studies (RDA and combined GEA-common garden mixed models, respectively; Lasky et al. 2012, 2018).

Although the screened genes were the ones with the largest effects in the original screen, we did GxE effects for fitness between well-watered and drought treatments were not significant in the follow-up experiment. Several factors could impact the results. First, the milder conditions of the follow-up drought experiment compared to the main drought screen may have reduced the G×E effect on fitness. Second, over-simplified laboratory environments could lead to systematic bias in plant fitness (Anderson et al., 2013). In natural environments, drought conditions are usually coupled with elevated temperatures, a variable not accounted for in controlled experiments designed to isolate the effects of a single environmental factor. In the laboratory, when only a single environmental variable is changed between treatments, a conditional neutrality, that a gene has fitness advantages under one treatment and is neutral under the other, can be more common (Anderson et al., 2013). Lastly, the knockout alleles of a gene may not be the mechanism of local adaptation in nature. Local adaptation may also occur through *cis*-regulatory effects on gene expression (Hämälä et al., 2020; Lasky et al., 2014) or changes in amino acid sequences.

Besides G×E effects on fitness, G×E effects on flowering time are also noteworthy when assessing local adaptation to drought stress, as early flowering is a critical drought escape strategy (Franks, 2011; Kenney et al., 2014). Chong and Stinchcombe (2019) reported systematic directional effects of flowering time among randomly selected SALK lines. In their study, 11 out of 27 lines under long-day conditions and 3 out of 27 lines under short-day conditions exhibited significantly different flowering times compared to *Col*, with all showing

delayed flowering. However, in our main drought screen, 19 out of 20 lines with significant genotype effects at $\alpha = 0.05$, and 12 out of 13 lines with significant genotype effects at FDR = 0.05, exhibited accelerated flowering (Figure 1). Among these, *wrky38* was the only line that showed delayed flowering (a potential drought avoidance strategy, Ludlow, 1989). Since the genes we selected were candidates for drought adaptation, the accelerated flowering observed in these lines may indicate their potential for drought adaptation via escape (Ludlow, 1989).

Though we only observed GxE for fitness in a small number of genes, evidence of associations between their natural functional variation and moisture-related bioclimate variables demonstrated the underlying biological bases of potential local adaptation of loci identified by GEA. Setting up this screen was relatively easy due to the availability of existing tDNA resources, making the approach an efficient way to identify and screen potential locally adapted genes in model organisms where such resources are available or relatively easily generated.

Potential mechanisms of drought adaptation at *WRKY38* and *LSD1*

The two mutants presented opposite responses under drought. *lsd1* had a larger SLA than *Col* thus potentially more water loss (Niinemets, 2001), which may result in its reduced RWC. In contrast, *wrky38* had a significantly smaller SLA when compared to *Col* and when under drought, indicating how *wrky38* maintained stable RWC under well-watered and drought conditions. *wrky38* mutants have reduced g_{sw}, SLA, and leaf area (result not shown) under drought in the follow-up experiment and increased WUE in both drought experiments (Figure 2-3), suggesting a drought-avoidance strategy of natural *WRKY38* LoF genotypes. *WRKY38* negatively regulates plant basal pathogen defense by suppressing *PATHOGENESIS-RELATED GENE 1* (*PR1*) expression, a defense gene induced by salicylic acid (SA) (Kim et al., 2008). Additionally, *WRKY38* is also involved in drought- and cold-related responses in barley (Marè et al., 2004). In the *Arabidopsis WRKY* family, *AtWRKY54* and *AtWRKY70* were found to negatively impact osmotic tolerance by reducing stomatal closure (Li et al., 2013), while *AtWRKY46* was found to regulate light-dependent stomatal opening in guard cells (Ding et al., 2014). Based on previous studies and our findings, we hypothesize that *WRKY38* inhibits stomatal closure. If so, the lower g_{sw} we found in knockouts could reduce water loss through the stomata and enhance fitness under drought.

In 1001 Genome accessions, natural *WRKY38* LoF alleles were strongly associated with lower July moisture compared to intact alleles. Despite this strong association, these putative LoF alleles were not fixed or entirely absent in any region. Our population genetics analyses revealed that LoF *WRKY38* exists across the entire distribution region of *A. thaliana* and in every lineage group (Figure 5a-f). We suspected that the LoF of *WRKY38* might have originated before the migration across Eurasia of *A. thaliana* and colonized drier regions during this process. In both drought experiments, *wrky38* generally had lower fitness compared to *Col* under well-watered conditions. However, *wrky38* maintained relatively stable fitness under drought stress, while *Col* showed a substantial fitness decrease. The drought insensitivity of *wrky38* may explain how natural LoF *WRKY38* dominates drier regions during migration.

lsd1 plants responded differently in response to the long-term water deficiency of our main drought screen and periodical moisture fluctuation in our follow-up experiments (Figure 2-

3). Specifically, *lsd1* mutants had similar fitness to *Col* under well-watered and drought in the main experiment, but they were less fit than *Col* (i.e. genotype effect was significant) in the follow-up. Previous laboratory studies have shown that *lsd1* exhibits similar survival and seed production to the wild type under nonlethal water deficiency. However, under lethal drought stress that caused complete mortality in wild-type plants, *lsd1* demonstrated a significantly higher survival rate, albeit with reduced seed production. These findings suggest a trade-off between survival and fecundity under severe stress in controlled laboratory conditions (Szechyńska-Hebda et al., 2016; Wituszynska et al., 2013). When grown in the field, *lsd1* mutants did not exhibit seed yield differences compared to the wild type (Bernacki et al., 2019; Szechyńska-Hebda et al., 2016; Wituszynska et al., 2013). This could explain our finding that the *LSD1* gene had lower expression in the hotter and drier regions, as reduced expression could similarly mitigate fitness trade-offs, enhancing survival in stressful climates.

Freezing responses of *wrky38* and *lsd1*

We conducted the freezing experiment because both genes may also be involved in adaptation to freezing stress. A SNP within the *WRKY38* coding region (Chr. 5, pos:7496047) was the top candidate loci associated with temperature of the coldest month in Lasky et al. (2018), and *LSD1* is associated with chilling sensitivity (Huang et al., 2010). Moreover, these two genes are both in the salicylic acid (SA) pathway that can be activated by cold stress (Miura and Tada, 2014; Wu et al., 2019). *WRKY38* was positively regulated by SA (Kim et al., 2008) while *LSD1* conditionally regulated SA concentration with the existences of *ENHANCED DISEASE SUSCEPTIBILITY1* (*EDS1*) and *PHYTOALEXIN DEFICIENT4* (*PAD4*) (Bernacki et al., 2019; Szechyńska-Hebda et al., 2016; Wituszynska et al., 2013).

WRKY38 knockouts exhibited similar fitness to *Col* despite reduced vegetative growth after long-term overnight freezing (Figure 4). Considering that *WRKY38* was originally identified through RDA (Lasky et al., 2012), this finding suggests a potential role in multivariate adaptation. The LoF allele might be advantageous in drier and warmer climates, while the functional allele might be more beneficial in cooler and wetter climates. *LSD1* knockouts had significantly reduced vegetative growth and fitness under freezing, coincident with the findings of Huang et al. (2010). Both mutants exhibited some disadvantages under freezing conditions compared to *Col*, indicating possible links between the function of these genes and adaptation to cold climates. Although both drought and freezing conditions can limit water availability and there is overlap in some molecular pathways or responses to each set of conditions (Kim et al., 2024), here our results suggest the drought-adaptive variants are not cold-adapted.

Conclusion

In this study, we tested the potential role of 44 t-DNA knockout mutants of GEA-identified genes in adaptation to drought stress using common garden experiments. While most mutants did not exhibit significant G×E effects for flowering time, performance, or fitness, two knockouts, *wrky38* and *lsd1*, demonstrated evidence of drought adaptation in both the main drought screen and follow-up intermediate drought common gardens. Natural variation in the function or expression

of these genes across a moisture gradient further supports the utility of GEA approaches for generating hypotheses, though further experiments are required to test hypotheses. Our findings highlight the promise of GEA methods for uncovering novel local adaptations to environmental stressors.

Acknowledgments

We thank Cody Depew for assistance in ordering knockout lines, Amanda Penn for assistance with experiments, and Diana Gamba for assistance with computation. Funding was provided by NIH R35GM138300 to JRL.

References

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., Ecker, J.R., 2003. Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana. Science 301, 653–657. https://doi.org/10.1126/science.1086391

Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K.M., Cao, J., Chae, E., Dezwaan, T.M., Ding, W., Ecker, J.R., Exposito-Alonso, M., Farlow, A., Fitz, J., Gan, X., Grimm, D.G., Hancock, A.M., Henz, S.R., Holm, S., Horton, M., Jarsulic, M., Kerstetter, R.A., Korte, A., Korte, P., Lanz, C., Lee, C.-R., Meng, D., Michael, T.P., Mott, R., Muliyati, N.W., Nägele, T., Nagler, M., Nizhynska, V., Nordborg, M., Novikova, P.Y., Picó, F.X., Platzer, A., Rabanal, F.A., Rodriguez, A., Rowan, B.A., Salomé, P.A., Schmid, K.J., Schmitz, R.J., Seren, Ü., Sperone, F.G., Sudkamp, M., Svardal, H., Tanzer, M.M., Todd, D., Volchenboum, S.L., Wang, C., Wang, G., Wang, X., Weckwerth, W., Weigel, D., Zhou, X., 2016. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell 166, 481–491. https://doi.org/10.1016/j.cell.2016.05.063

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. Mol. Ecol. 22, 699–708. https://doi.org/10.1111/j.1365-294X.2012.05522.x

Bernacki, M.J., Czarnocka, W., Rusaczonek, A., Witoń, D., Kęska, S., Czyż, J., Szechyńska-Hebda, M., Karpiński, S., 2019. LSD1-, EDS1- and PAD4-dependent conditional correlation among salicylic acid, hydrogen peroxide, water use efficiency and seed yield in Arabidopsis thaliana. Physiol. Plant. 165, 369–382. https://doi.org/10.1111/ppl.12863

Capblancq, T., Lachmuth, S., Fitzpatrick, M.C., Keller, S.R., 2023. From common gardens to candidate genes: exploring local adaptation to climate in red spruce. New Phytol. 237, 1590–1605. https://doi.org/10.1111/nph.18465

Chong, V.K., Stinchcombe, J.R., 2019. Evaluating Population Genomic Candidate Genes Underlying Flowering Time in Arabidopsis thaliana Using T-DNA Insertion Lines. J. Hered. 110, 445–454. https://doi.org/10.1093/jhered/esz026

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide

polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6, 80–92. https://doi.org/10.4161/fly.19695
Clausen, J., Keck, D.D., Hiesey, W.M., 1941. Regional Differentiation in Plant Species. Am. Nat. 75, 231–250.
Coop, G., Witonsky, D., Di Rienzo, A., Pritchard, J.K., 2010. Using Environmental Correlations to Identify Loci Underlying Local Adaptation. Genetics 185, 1411–1423. https://doi.org/10.1534/genetics.110.114819
Ding, Z.J., Yan, J.Y., Xu, X.Y., Yu, D.Q., Li, G.X., Zhang, S.Q., Zheng, S.J., 2014. Transcription factor WRKY46 regulates osmotic stress responses and stomatal movement independently in Arabidopsis. Plant J. 79, 13–27. https://doi.org/10.1111/tpj.12538
Dittberner, H., Korte, A., Mettler-Altmann, T., Weber, A.P.M., Monroe, G., de Meaux, J., 2018. Natural variation in stomata size contributes to the local adaptation of water-use efficiency in Arabidopsis thaliana. Mol. Ecol. 27, 4052–4065. https://doi.org/10.1111/mec.14838
El-Soda, M., Kruijer, W., Malosetti, M., Koornneef, M., Aarts, M.G.M., 2015. Quantitative trait loci and candidate genes underlying genotype by environment interaction in the response of Arabidopsis thaliana to drought. Plant Cell Environ. 38, 585–599. https://doi.org/10.1111/pce.12418
Endler, J.A., 1973. Gene Flow and Population Differentiation. Science 179, 243–250. https://doi.org/10.1126/science.179.4070.243
Exposito-Alonso, M., Vasseur, F., Ding, W., Wang, G., Burbano, H.A., Weigel, D., 2018. Genomic basis and evolutionary potential for extreme drought adaptation in Arabidopsis thaliana. Nat. Ecol. Evol. 2, 352–358. https://doi.org/10.1038/s41559-017-0423-0
Forester, B.R., Lasky, J.R., Wagner, H.H., Urban, D.L., 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. Mol. Ecol. 27, 2215–2233. https://doi.org/10.1111/mec.14584
Franks, S.J., 2011. Plasticity and evolution in drought avoidance and escape in the annual plant Brassica rapa. New Phytol. 190, 249–257. https://doi.org/10.1111/j.1469-8137.2010.03603.x
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. Plant J. 84, 682–693. https://doi.org/10.1111/tpj.13027
Hämälä, T., Gorton, A.J., Moeller, D.A., Tiffin, P., 2020. Pleiotropy facilitates local adaptation to distant optima in common ragweed (Ambrosia artemisiifolia). PLOS Genet. 16, e1008707. https://doi.org/10.1371/journal.pgen.1008707
Hancock, A.M., Brachi, B., Faure, N., Horton, M.W., Jarymowycz, L.B., Sperone, F.G., Toomajian, C., Roux, F., Bergelson, J., 2011. Adaptation to Climate Across the Arabidopsis thaliana Genome. Science 334, 83–86. https://doi.org/10.1126/science.1209244
Huang, X., Li, Y., Zhang, X., Zuo, J., Yang, S., 2010. The Arabidopsis LSD1 gene plays an important role in the regulation of low temperature-dependent cell death. New Phytol. 187, 301–312. https://doi.org/10.1111/j.1469-8137.2010.03275.x
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. https://doi.org/10.1534/genetics.107.080101
Kawakatsu, T., Huang, S.C., Jupe, F., Sasaki, E., Schmitz, R.J., Urich, M.A., Castanon, R., Nery, J.R., Barragan, C., He, Y., 2016. Epigenomic diversity in a global collection of Arabidopsis thaliana accessions. Cell 166, 492–505.
Kenney, A.M., McKay, J.K., Richards, J.H., Juenger, T.E., 2014. Direct and indirect selection on flowering time, water-use efficiency (WUE, δ13C), and WUE plasticity to drought in

Arabidopsis thaliana. Ecol. Evol. 4, 4505–4521. https://doi.org/10.1002/ece3.1270
Kim, J.-S., Kidokoro, S., Yamaguchi-Shinozaki, K., Shinozaki, K., 2024. Regulatory networks in plant responses to drought and cold stress. Plant Physiol. 195, 170–189. https://doi.org/10.1093/plphys/kiae105
Kim, K.C., Lai, Z., Fan, B., Chen, Z., 2008. Arabidopsis WRKY38 and WRKY62 transcription factors interact with histone deacetylase 19 in basal defense. Plant Cell 20, 2357–2371. https://doi.org/10.1105/tpc.107.055566
Lasky, J.R., Des Marais, D.L., Lowry, D.B., Povolotskaya, I., McKay, J.K., Richards, J.H., Keitt, T.H., Juenger, T.E., 2014. Natural Variation in Abiotic Stress Responsive Gene Expression and Local Adaptation to Climate in Arabidopsis thaliana. Mol. Biol. Evol. 31, 2283–2296. https://doi.org/10.1093/molbev/msu170
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. Mol. Ecol. 21, 5512–5529. https://doi.org/10.1111/j.1365-294X.2012.05709.x
Lasky, J.R., Forester, B.R., Reimherr, M., 2018. Coherent synthesis of genomic associations with phenotypes and home environments. Mol. Ecol. Resour. 18, 91–106. https://doi.org/10.1111/1755-0998.12714
Lasky, J.R., Josephs, E.B., Morris, G.P., 2023. Genotype–environment associations to reveal the molecular basis of environmental adaptation. Plant Cell 35, 125–138. https://doi.org/10.1093/plcell/koac267
Lawrence-Paul, E.H., Poethig, R.S., Lasky, J.R., 2023. Vegetative phase change causes age-dependent changes in phenotypic plasticity. New Phytol. 240, 613–625. https://doi.org/10.1111/nph.19174
Lee, C.-R., Svardal, H., Farlow, A., Exposito-Alonso, M., Ding, W., Novikova, P., Alonso-Blanco, C., Weigel, D., Nordborg, M., 2017. On the post-glacial spread of human commensal Arabidopsis thaliana. Nat. Commun. 8, 14458. https://doi.org/10.1038/ncomms14458
Lee, G., Sanderson, B.J., Ellis, T.J., Dilkes, B.P., McKay, J.K., Ågren, J., Oakley, C.G., 2024. A large-effect fitness trade-off across environments is explained by a single mutation affecting cold acclimation. Proc. Natl. Acad. Sci. 121, e2317461121. https://doi.org/10.1073/pnas.2317461121
Li, J., Besseau, S., Törönen, P., Sipari, N., Kollist, H., Holm, L., Palva, E.T., 2013. Defense-related transcription factors WRKY70 and WRKY54 modulate osmotic stress tolerance by regulating stomatal aperture in Arabidopsis. New Phytol. 200, 457–472. https://doi.org/10.1111/nph.12378
Lorts, C.M., Lasky, J.R., 2020. Competition × drought interactions change phenotypic plasticity and the direction of selection on Arabidopsis traits. New Phytol. 227, 1060–1072. https://doi.org/10.1111/nph.16593
Lotterhos, K.E., 2023. The paradox of adaptive trait clines with nonclinal patterns in the underlying genes. Proc. Natl. Acad. Sci. 120, e2220313120. https://doi.org/10.1073/pnas.2220313120
Lovell, J.T., Juenger, T.E., Michaels, S.D., Lasky, J.R., Platt, A., Richards, J.H., Yu, X., Easlon, H.M., Sen, S., McKay, J.K., 2013. Pleiotropy of FRIGIDA enhances the potential for multivariate adaptation. Proc. R. Soc. B Biol. Sci. 280, 20131043. https://doi.org/10.1098/rspb.2013.1043
Ludlow, M.M., 1989. Strategies of response to water stress. In K. H. Kreeb, H. Richter, & T. M. Hinckley (Eds.), Structural and Functional Responses to Environmental Stresses (pp. 269–281). SPB Academic.
Marè, C., Mazzucotelli, E., Crosatti, C., Francia, E., Stanca, A. ?michele, Cattivelli, L., 2004. Hv-WRKY38: a new transcription factor involved in cold- and drought-response in barley. Plant Mol. Biol. 55, 399–416. https://doi.org/10.1007/s11103-004-0906-7
Miura, K., Tada, Y., 2014. Regulation of water, salinity, and cold stress responses by salicylic

acid. Front. Plant Sci. 5.
Monroe, J.G., Powell, T., Price, N., Mullen, J.L., Howard, A., Evans, K., Lovell, J.T., McKay, J.K., 2018. Drought adaptation in Arabidopsis thaliana by extensive genetic loss-of-function. eLife 7, e41038. https://doi.org/10.7554/eLife.41038
Niinemets, Ü., 2001. Global-Scale Climatic Controls of Leaf Dry Mass Per Area, Density, and Thickness in Trees and Shrubs. Ecology 82, 453–469. https://doi.org/10.1890/0012-9658(2001)082[0453:GSCCOL]2.0.CO;2
Pisupati, R., Reichardt, I., Seren, Ü., Korte, P., Nizhynska, V., Kerdaffrec, E., Uzunova, K., Rabanal, F.A., Filiault, D.L., Nordborg, M., 2017. Verification of Arabidopsis stock collections using SNPmatch, a tool for genotyping high-plexed samples. Sci. Data 4, 170184. https://doi.org/10.1038/sdata.2017.184
Ralph, P.L., Coop, G., 2015. Convergent Evolution During Local Adaptation to Patchy Landscapes. PLOS Genet. 11, e1005630. https://doi.org/10.1371/journal.pgen.1005630
Rellstab, C., Gugerli, F., Eckert, A.J., Hancock, A.M., Holderegger, R., 2015. A practical guide to environmental association analysis in landscape genomics. Mol. Ecol. 24, 4348–4370. https://doi.org/10.1111/mec.13322
Savolainen, O., Lascoux, M., Merilä, J., 2013. Ecological genomics of local adaptation. Nat. Rev. Genet. 14, 807–820. https://doi.org/10.1038/nrg3522
Signorell, A., Aho, K., Alfons, A., Anderegg, N., Aragon, T., Arppe, A., Baddeley, A., Barton, K., Bolker, B., Borchers, H., 2023. DescTools: tools for descriptive statistics. R package version 0.99. 26. Compr. R. Arch Netw 289–291.
Szechyńska-Hebda, M., Czarnocka, W., Hebda, M., Karpiński, S., 2016. PAD4, LSD1 and EDS1 regulate drought tolerance, plant biomass production, and cell wall properties. Plant Cell Rep. 35, 527–539. https://doi.org/10.1007/s00299-015-1901-y
Tigano, A., Friesen, V.L., 2016. Genomics of local adaptation with gene flow. Mol. Ecol. 25, 2144–2164. https://doi.org/10.1111/mec.13606
Wadgymar, S.M., DeMarche, M.L., Josephs, E.B., Sheth, S.N., Anderson, J.T., 2022. Local Adaptation: Causal Agents of Selection and Adaptive Trait Divergence. Annu. Rev. Ecol. Evol. Syst. 53, 87–111. https://doi.org/10.1146/annurev-ecolsys-012722-035231
Wituszynska, W., Slesak, I., Vanderauwera, S., Szechynska-Hebda, M., Kornas, A., Van Der Kelen, K., Mühlenbock, P., Karpinska, B., Mackowski, S., Van Breusegem, F., Karpinski, S., 2013. Lesion simulating disease1, enhanced disease susceptibility1, and phytoalexin deficient4 conditionally regulate cellular signaling homeostasis, photosynthesis, water use efficiency, and seed yield in Arabidopsis. Plant Physiol. 161, 1795–1805. https://doi.org/10.1104/pp.112.208116
Wu, Z., Han, S., Zhou, H., Tuang, Z.K., Wang, Y., Jin, Y., Shi, H., Yang, W., 2019. Cold stress activates disease resistance in Arabidopsis thaliana through a salicylic acid dependent pathway. Plant Cell Environ. 42, 2645–2663. https://doi.org/10.1111/pce.13579
Yeaman, S., Whitlock, M.C., 2011. THE GENETIC ARCHITECTURE OF ADAPTATION UNDER MIGRATION–SELECTION BALANCE. Evolution 65, 1897–1911. https://doi.org/10.1111/j.1558-5646.2011.01269.x
Yim, C., Bellis, E.S., DeLeo, V.L., Gamba, D., Muscarella, R., Lasky, J.R., 2024. Climate biogeography of Arabidopsis thaliana: Linking distribution models and individual variation. J. Biogeogr. 51, 560–574. https://doi.org/10.1111/jbi.14737
Zivcak, M., Brestic, M., Balatova, Z., Drevenakova, P., Olsovska, K., Kalaji, H.M., Yang, X., Allakhverdiev, S.I., 2013. Photosynthetic electron transport and specific photoprotective responses in wheat leaves under drought stress. Photosynth. Res. 117, 529–546. https://doi.org/10.1007/s11120-013-9885-3
Zoubarev, A., Hamer, K.M., Keshav, K.D., McCarthy, E.L., Santos, J.R.C., Van Rossum, T., McDonald, C., Hall, A., Wan, X., Lim, R., Gillis, J., Pavlidis, P., 2012. Gemma: a resource for the reuse, sharing and meta-analysis of expression profiling data.

Bioinformatics 28, 2272–2273. https://doi.org/10.1093/bioinformatics/bts430