Environmental response in gene expression and DNA methylation reveals factors influencing the adaptive potential of *Arabidopsis lyrata*

Tuomas Hämälä[1,*], Weixuan Ning[1,2], Helmi Kuittinen[1], Nader Aryamanesh[1,3,4], and Outi Savolainen[1,4]

[1] Department of Ecology and Genetics, University of Oulu, 90014 Oulu, Finland

[2] Current address: School of Natural Sciences, Massey University, Palmerston North 4442, New Zealand

[3] Current address: Embryology Research Unit, Bioinformatics Group, Children's Medical Research Institute, University of Sydney, Westmead NSW 2145, Australia

[4] These authors contributed equally to this work

* Corresponding author: tuomas.hamala@gmail.com

Abstract

Understanding what factors influence plastic and genetic variation is valuable for predicting how organisms respond to changes in the selective environment. Here, using gene expression and DNA methylation as molecular phenotypes, we study environment-induced variation among *Arabidopsis lyrata* plants grown at lowland and alpine field sites. Our results show that gene expression is highly plastic, as many more genes are differentially expressed between the field sites than between populations. The environmentally responsive genes evolve under strong selective constraint – the strength of purifying selection on the coding sequence is high, while the rate of adaptive evolution is low. We find, however, that positive selection on *cis*-regulatory variants has likely contributed to genetically variable environment-responses, but such variants segregate only between distantly related populations. In contrast to gene expression, DNA methylation at genic regions is largely insensitive to the environment, and plastic methylation changes are not associated with differential gene expression. Besides genes, we detect environmental effects at transposable elements (TEs): TEs at the high-altitude field site have higher expression and methylation levels, suggestive of a broad-scale TE activation. Compared to the lowland population, plants native to the alpine environment harbor and excess of recent TE insertions, and we observe that specific TE families are enriched within environmentally responsive genes. In sum, although our results suggest that strong evolutionary constraint at environmentally responsive genes may limit the species' adaptive potential, plastic responses at TEs could rapidly create novel heritable variation that is primed towards environmental adaptation.

Introduction

To maintain viability in a changing environment, populations need to track shifting fitness optima through genetic adaptation or phenotypic plasticity (Alberto et al., 2013; Ghalambor et al., 2007; Parmesan, 2006). Although environmentally induced variation within a generation can efficiently facilitate population persistence in novel environments, selection on heritable variation is ultimately required for long-term adaptability (Wright, 1931). As such, strong phenotypic plasticity may constrain adaptive evolution by masking genetic variation from directional selection (Falconer, 1981; Grant, 1977; Wright, 1931). On the other hand, increased population persistence due to plasticity might provide more opportunities for selection to act on heritable variation, making adaptive plasticity an important first step in evolution (Baldwin, 1896; Schmalhausen, 1949; Waddington, 1942). Given the importance of phenotypic plasticity in adaptive evolution, understanding the selective forces that shape plastic and genetic variation is valuable for predicting how organisms respond to both natural and human-mediated selection.

In most cases, plasticity in morphological, developmental, and physiological traits is thought to result from changes in gene expression (West-Eberhard, 2003), making the study of expression plasticity a promising approach for uncovering the genetic basis of phenotypic plasticity. Although studies of environmentally responsive genes have discovered a wide range of expression responses (Hodgins-Davis and Townsend, 2009), many populations and species have reacted consistently to environmental stress (He et al., 2021; Lowry et al., 2013; Yeaman et al., 2014). As gene expression is primarily controlled by regulatory elements acting either in *cis* (affecting nearby genes) or *trans* (affecting distant genes), the conserved expression responses are indicative of conserved regulatory systems (Horvath et al., 2021; Lu et al., 2019; Rodgers-Melnick et al., 2016). Such plastically expressed genes have also shown signals of strong purifying selection at the coding regions (Hodgins et al., 2016; Hunt et al., 2013; Lasky et al., 2014; Lowry et al., 2013; Zhang et al., 2017), suggesting that the regulatory conservation is frequently coupled with the conservation of the gene product. Consistent with this expectation, genes exhibiting genetically variable responses to the environment have harbored signals of relaxed selective constraint (Hodgins et al., 2016; Koenig et al., 2013; Lasky et al., 2014; Zhang et al., 2017). Given these general observations, we may expect that strong evolutionary conservation at environmentally responsive genes limits the emergence of heritable variation, whereas such variation is more likely to arise in rapidly evolving genes.

Besides inducing phenotypic plasticity through changes in gene expression, environmental stress may invoke plastic responses at transposable elements (TEs) (Capy et al., 2000; Casacuberta and González, 2013; Ito et al., 2011; McClintock, 1984; Pietzenuk et al., 2016; Wos et al., 2021). TEs are commonly divided into two major classes depending on their mode of transposition:

retrotransposons (or class I) that move by "copy-and-paste" mechanism and DNA transposons (or class II) that move by "cut-and-paste" mechanism (Wicker et al., 2007). Both classes can be further separated into distinct orders and families, which often occupy different "niches" within the host genome (Stitzer et al., 2021). The activation of TEs leads both to proliferation of new TE copies and mobility of existing ones, which can have a considerable influence on the adaptive potential of a population (Bourgeois and Boissinot, 2019). On one hand, a large majority of new TEs and other structural variants are expected to be deleterious (Baduel et al., 2021; Bourgeois and Boissinot, 2019; Hämälä et al., 2021; Kou et al., 2020), and so TE activation likely increases the genetic load of a population. On the other hand, such activation can create novel functional variants, which may facilitate adaptation under new selective environments (Capy et al., 2000; Casacuberta and González, 2013; Ito et al., 2011; McClintock, 1984). Indeed, TEs have been associated with the emergence of several adaptive phenotypes, including industrial melanism in peppered moth (Hof et al., 2016), early flowering in *Arabidopsis thaliana* (Quadrana et al., 2016), and single-stalk branching pattern in maize (Studer et al., 2011).

Here, we conducted a reciprocal transplant experiment to study selective processes underlying environmental responses. To do so, we grew *Arabidopsis lyrata* plants from three populations at natural low- and high-altitude field sites, leading to sharp differences in exposure to abiotic (e.g., temperature and solar radiation) and biotic (e.g., herbivores and pathogens) factors. Besides examining variation in gene expression, we searched for differentiation in DNA methylation between our experimental plants. DNA methylation is a common epigenetic modification that modulates gene expression (Law and Jacobsen, 2010). The regulatory mechanisms underlying DNA methylation can be rapidly activated by the environment, thus modifying gene expression in response to changing environmental conditions (Liu and He, 2020; Thiebaut et al., 2019). In plants, DNA methylation is also stably inherited from parent to offspring (Law and Jacobsen, 2010), whereas other non-genetic modulators of gene expression are more likely to be reset during meiosis and embryonic development (Escobar et al., 2021; Lloyd and Lister, 2022). Furthermore, DNA methylation is commonly involved in epigenetic silencing of TEs (Law and Jacobsen, 2010), and thus environmentally induced changes in the DNA methylome could either promote or suppress TE activity. Therefore, characterizing patterns of DNA methylation in naturally contrasting conditions may help us to understand long-term environmental effects on gene expression and TE activity. Using these data, we address the following questions: How are gene expression and DNA methylation influenced by the growing environment and population history? Does differential methylation lead to differential gene expression? Are differentially expressed and methylated genes under strong or relaxed evolutionary constraint? Do

we find evidence of environmentally induced TE activation? Does TE activation create novel genetic variation capable of influencing environmental adaptation?

Results

To study patterns of short- and long-distance local adaptation in *A. lyrata*, we grew plants from three populations at two contrasting common garden sites in Norway (Fig. 1A). At both low- (300 meters above sea level [m.a.s.l.]) and high-altitude (1100 m.a.s.l.) field, we planted individuals from local Norwegian populations (J1, 300 m.a.s.l.; J3, 1100 m.a.s.l.) as well as individuals from a nonlocal population from Germany (GER) (Fig. S1). Although the field sites were closely situated (Fig. 1A), the altitudinal difference resulted in considerable environmental differences between the fields (Fig. 1B and Fig. S2), thus leading to distinct responses in ecologically important traits (Fig. 1C). Using multi-year fitness estimated during the experiment, we found evidence of home-site advantage between the J1 and J3 populations, while individuals from the GER population fared poorly at both field sites (Hämälä et al., 2018). Here, we combine this reciprocal transplant experiment with transcriptome and whole-genome bisulfite sequencing to examine the adaptive roles of gene expression and DNA methylation. Leaf samples from the plants were collected one year after planting, thus allowing long-term exposure to the environment. By comparing the field sites and populations, we can distinguish between three types of expression and methylation responses: 1) plastic (difference between the fields, but no difference between populations), 2) genetic (difference between populations, but no difference between the fields), and 3) plastic × genetic (population-specific difference between the fields).

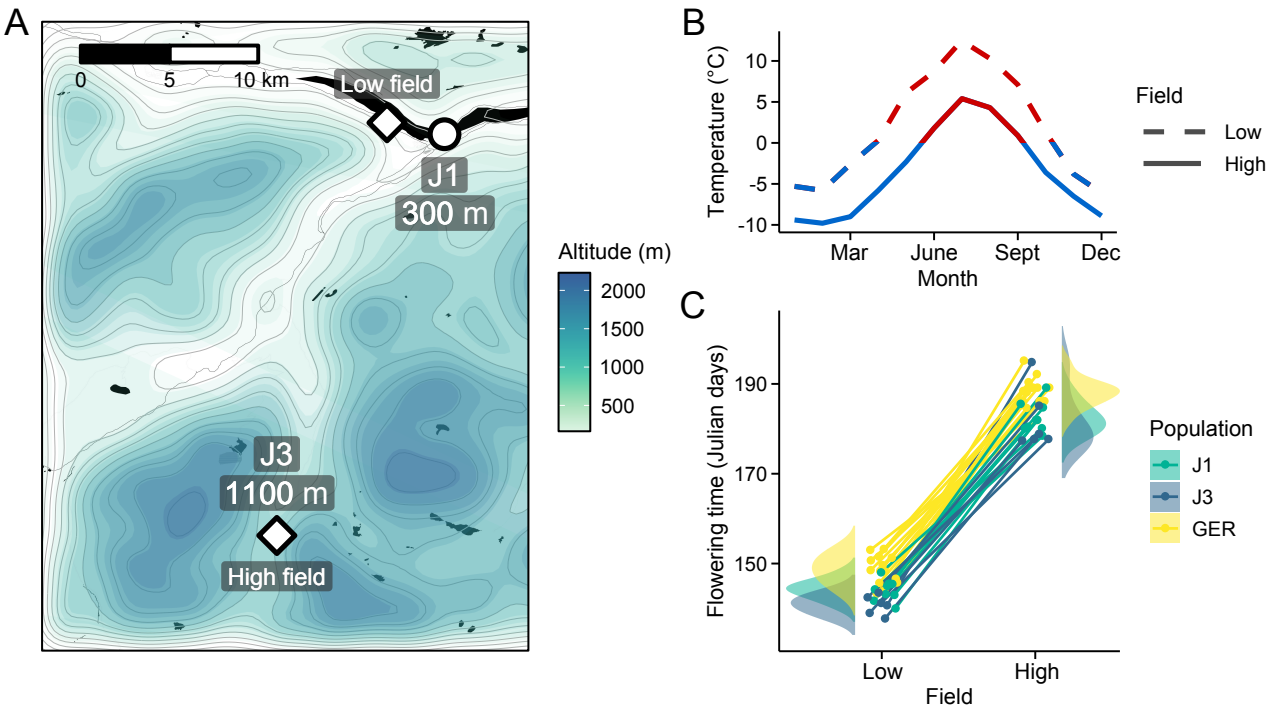

Figure 1. Reciprocal transplant experiment to study environmental adaptation in *A. lyrata*. **A**: Locations and altitudes of the Norwegian populations and field sites. Map tiles by Stamen Design, under CC BY 3.0. Map data by OpenStreetMap, under ODbL. Altitude data from SRTM. **B**: Average monthly temperature at areas around the field sites (red line: ≥ 0 °C, blue line: < 0 °C). Data from MET Norway. **C**: Average flowering time of full-sib families grown at the field sites. Data from Hämälä et al. (2018).

Patterns of variation in gene expression and DNA methylation

We first used principal components analysis (PCA) to detect the main sources of variation in the gene expression and methylation data. In the gene expression data, the first principal component (PC) primarily corresponded to differences between the two field sites, whereas the Norwegian populations (J1 and J3) were separated from GER along the second PC axis (Fig. 2A). By contrast, methylation in the CG context was mainly influenced by population structure, with first two PCs separating the three populations (Fig. 2A). Methylation in the CHG and CHH contexts (where H is A, T, or C) was primarily affected by differences between the Norwegian populations and GER, but second PC in both PCAs revealed a slight influence of the field site, particularly in the Norwegian populations (Fig. S3).

All three populations contained genes that were differentially expressed between the fields (Fig. 2B), but the largest number of DEGs (differentially expressed genes) was found in J1. Expression differences between populations were mainly due to DEGs between the Norwegian populations and GER (Fig. 2C), with a total of 1675 DEGs between GER and either J1 or J3. By contrast, we only detected 31 DEGs between J1 and J3 (Fig. 2C). Therefore, environmental effects

were evident in all populations, whereas population effects mainly arose from differences between Norway and Germany.

Levels of DNA methylation were highly variable between the three contexts, with approximately 30%, 10%, and 3.5% of cytosines methylated in the CG, CHG, and CHH contexts, respectively (Fig. 2D). CG methylation levels showed only a subtle difference between the field sites, but a clear difference between the populations (Fig. 2D): both J1 and J3 had considerably higher methylation levels than GER ($P < 2 \times 10^{-16}$, likelihood-ratio test [LRT]). For both CHG and CHH contexts, the Norwegian populations had higher methylation levels at the high-altitude field site ($P < 2 \times 10^{-16}$, LRT), whereas GER had higher methylation levels at the low-altitude field site ($P < 2 \times 10^{-16}$, LRT).

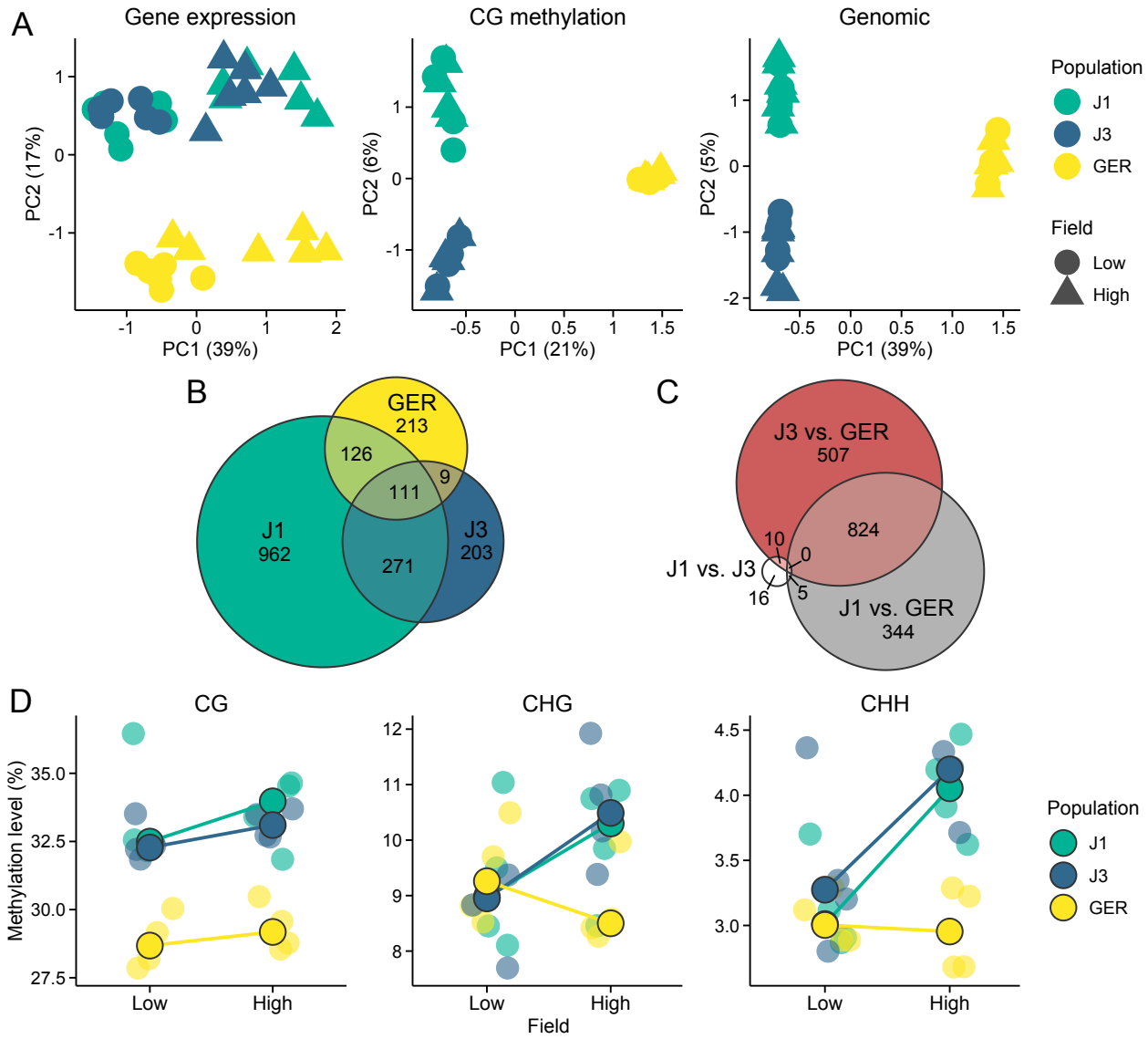

Figure 2. Gene expression and methylation variation across the field sites and populations. **A:** Gene expression, CG methylation, and genomic (based on SNPs called from the transcriptome data) variation along the first two eigenvectors of a PCA. The proportion of variance explained by the PCs is shown in parentheses. **B**: The number of DEGs between field sites. **C**: The number of DEGs between populations (across both fields). **D:** Average methylation levels at the two field sites, shown for CG, CHG, and CHH contexts. Black lined circles show median estimates for the populations, while individual estimates are shown with transparent colors in the background. Note the difference in *y*-axis scales between the panels.

Plastic and genetic responses at gene expression and gene body methylation

We found ample variation in gene expression and DNA methylation among our experimental plants. To more clearly distinguish the source of the variation, we used LRTs in DESeq2 (Love et al., 2014) to detect DEGs due to field site (DEG ~ Field), population (DEG ~ Population), and their interaction (DEG ~ Field:Population). We also used logistic regression and LRTs to conduct a similar analysis among differentially methylated genes (DMGs). As gene body methylation (GBM) is almost

exclusively found in the CG context (Law and Jacobsen, 2010), we searched for DMGs using CG methylation only.

We detected between 477 and 3456 genes that had their expression or methylation levels affected by the field site and/or population history (Fig. 3A). Consistent with the PCA results, most DEGs were found between the field sites (DEG ~ Field) and most DMGs were found between populations (DMG ~ Population). Field × population interactions (DEG ~ Field:Population) were rare in gene expression, and using the same criteria for outlier detection as in DEG ~ Field and DEG ~ Population, only 28 genes passed the DEG threshold. As this low number did not allow us to examine selective signals at the field × population genes, we used a more lenient threshold for multiple correction (see Methods), while acknowledging the potentially higher false-positive rate among these genes. Although levels of GBM were positively correlated with gene expression levels (Spearman's $\rho = 0.21$, $P < 2 \times 10^{-16}$), DEG ~ Field shared fewer than expected genes with the DMG groups (Fig. 3A). Besides GBM, changes in gene expression may result from methylation of the promoter regions (Law and Jacobsen, 2010). To explore this, we compiled methylation data from 1 kb upstream of each gene and searched for an overlap between differentially methylated promoter regions and DEGs. For DEG ~ Field and DEG ~ Field:Population, these results did not deviate from random expectation (Fig. S4). DEG ~ Population, by contrast, was associated with differential methylation at both gene bodies (DMG ~ Population) and promoter regions (Fig. S4), suggesting that genetic differentiation in DNA methylation results in genetic expression responses.

Based on gene ontology (GO) terms, all DEG and DMG sets were enriched for genes involved in specific biological processes (Fig. 3B). DEG ~ Field, a group of genes with plastic expression responses, had multiple enriched GO terms involved in photosynthesis ("photosynthesis", "chlorophyll biosynthetic process", and "reductive pentose-phosphate cycle" as the top three terms). Among the top five terms were also "response to cold" and "response to light intensity", which were previously found as enriched terms among local adaptation candidates in J3 (Hämälä and Savolainen, 2019). Genes showing genetic expression responses (DEG ~ Population) were enriched for only three GO terms, two of which were related to defense against pathogens. DEG ~ Field:Population showed enrichment for multiple different processes, including "response to cold". Enriched GO terms among the DMGs were less varied, with terms involving calcium ion transport (e.g., "calcium-mediated signaling") strongly represented among all three groups (Fig. 3B). We note, however, that the consistent results among the DMGs are likely influenced by the larger than expected overlap between the three groups (Fig. 3A).

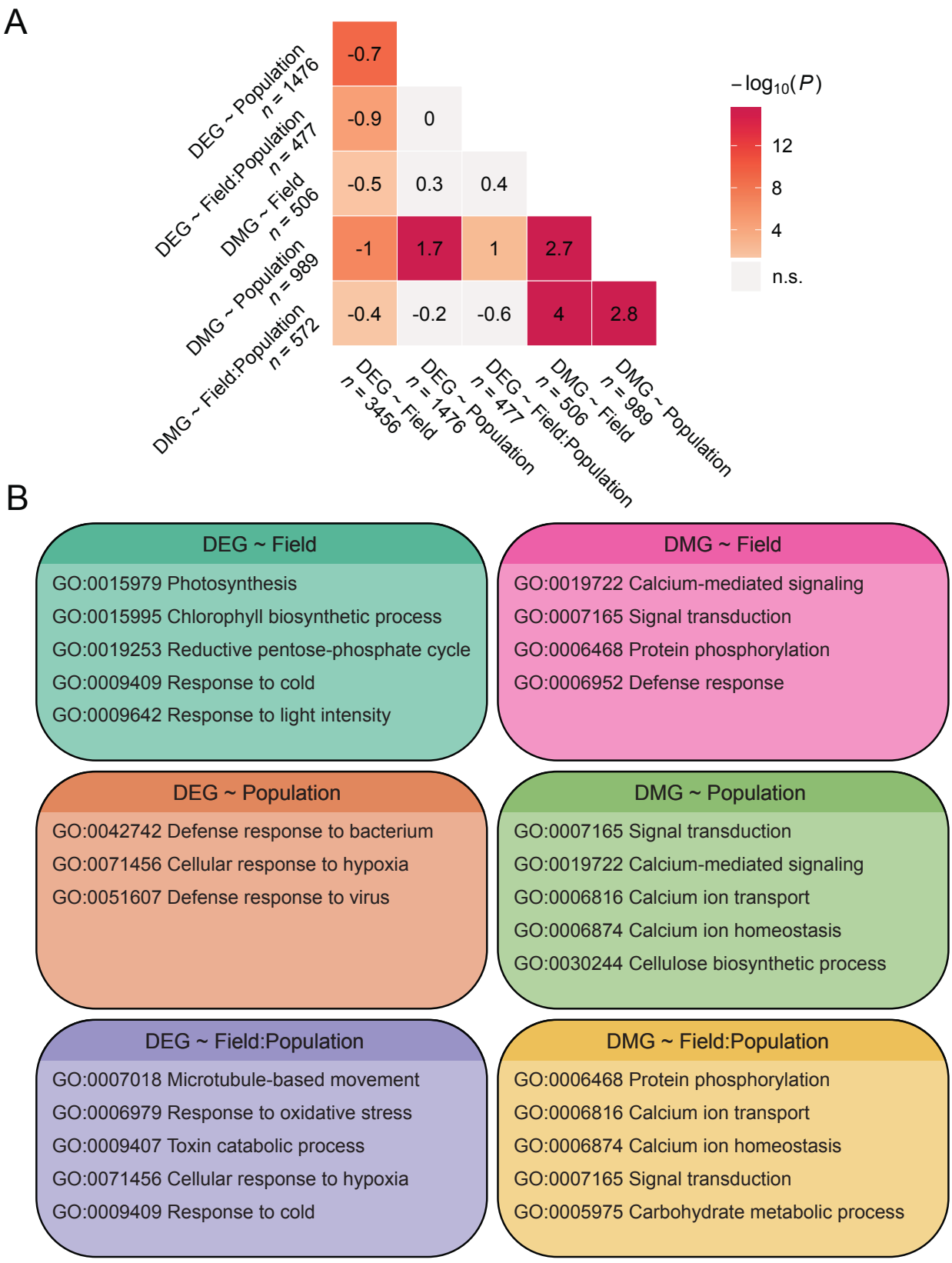

Figure 3. Overlap and functional enrichment among candidate gene sets. **A:** Overlap among candidate gene sets. Shown are $\log_2$ odds ratios (numbers) and *P* values (color gradient) from Fisher's exact tests. **B:** Top five enriched ($Q < 0.05$, hypergeometric test) GO terms among each candidate gene set.

Footprints of selection at candidate gene sets

To examine selective signals at DEGs and DMGs, we used whole-genome data from independent J1, J3, and GER individuals (Hämälä et al., 2018; Mattila et al., 2017; Takou et al., 2021). Compared to the genome-wide average, genes belonging to the DEG groups harbored lower than expected nucleotide diversity within populations as well as higher than expected nucleotide differentiation between populations (Figs. 4 and S5), indicative of purifying selection or selective sweeps. In particular, the promoter regions at DEG ~ Field:Population genes exhibited a combination low genetic diversity and high differentiation, suggesting that selection on *cis*-regulatory variants at these genes may have contributed to local adaptation. The signals of expression and sequence differentiation were, however, almost exclusively found between the Norwegian populations and GER (Figs. S6 and S7).

The estimates of genetic diversity indicate that selective processes have shaped the nucleotide composition of the candidate gene sets. To study this in more detail, we inferred the strength of selective constraint on the DEGs and DMGs by modelling the distribution of fitness effects (DFE) of new nonsynonymous mutations (Kim et al., 2017) and by estimating the rate of adaptive evolution (α) (Messer and Petrov, 2013). In all three populations, genes belonging to the DEG ~ Field and DEG ~ Field:Population groups appeared under strong purifying selection: fewer than expected mutations were predicted to be nearly neutral ($2N_es < 1$) and more predicted to be deleterious ($2N_es > 10$) (Figs. 5A and S8A; $P = 0.004$). If the strong selective constraint is due to localized increase of N_e, we might expect a similar increase in the efficacy of positive selection. However, α estimates indicated that fewer than expected nonsynonymous mutations have been fixed by positive selection in the DEG ~ Field and DEG ~ Field:Population groups (Figs. 5B and S8B; $P < 0.004$, LRT). In contrast to the plastic DEG groups, the DFE and α estimates indicated relaxed evolutionary constraint at other candidate gene groups (Fig. 5A and S8A). In particular, the two groups with genetic methylation responses, DMG ~ Population and DMG ~ Field:Population, had higher than expected α in all three populations (Figs. 5B and S8B; $P < 0.02$, LRT), suggesting long-term adaptive evolution. Together, these results are indicative of reduced evolutionary rates at the plastic DEG groups as well as rapid molecular evolution at the DMGs. Both results were further supported by an analysis of nucleotide conservation among 26 eudicot species, which identified higher than expected sequence conservation at the two DEG groups and lower than expected conservation at the DMGs (Fig. 5C; $P < 0.0005$, Wilcoxon rank-sum test).

To explore how the differing evolutionary rates influence the short-term adaptive potential of these gene groups, we used RAiSD (Alachiotis and Pavlidis, 2018) to search for footprints of recent positive selection at each gene (gene body ± 2 kb). We found that DEG ~ Field exhibited a clear deficit of selective sweeps (Figs. 5D and S8C; $P < 0.006$, Fisher's exact test), consistent with the reduced signals of adaptive evolution. By contrast, DEG ~ Field:Population showed an enrichment of selective sweeps in all three populations (Figs. 5D and S8C; $P < 0.002$, Fisher's exact test). Between 13 and 14 genes overlapped the sweep regions in the three populations, with 11 genes shared between the populations. At each of these genes, windows producing the strongest selection signals contained the putative promoter regions (within 2 kb upstream of transcription start site), suggesting that *cis*-regulatory elements may have been subject to recent positive selection.

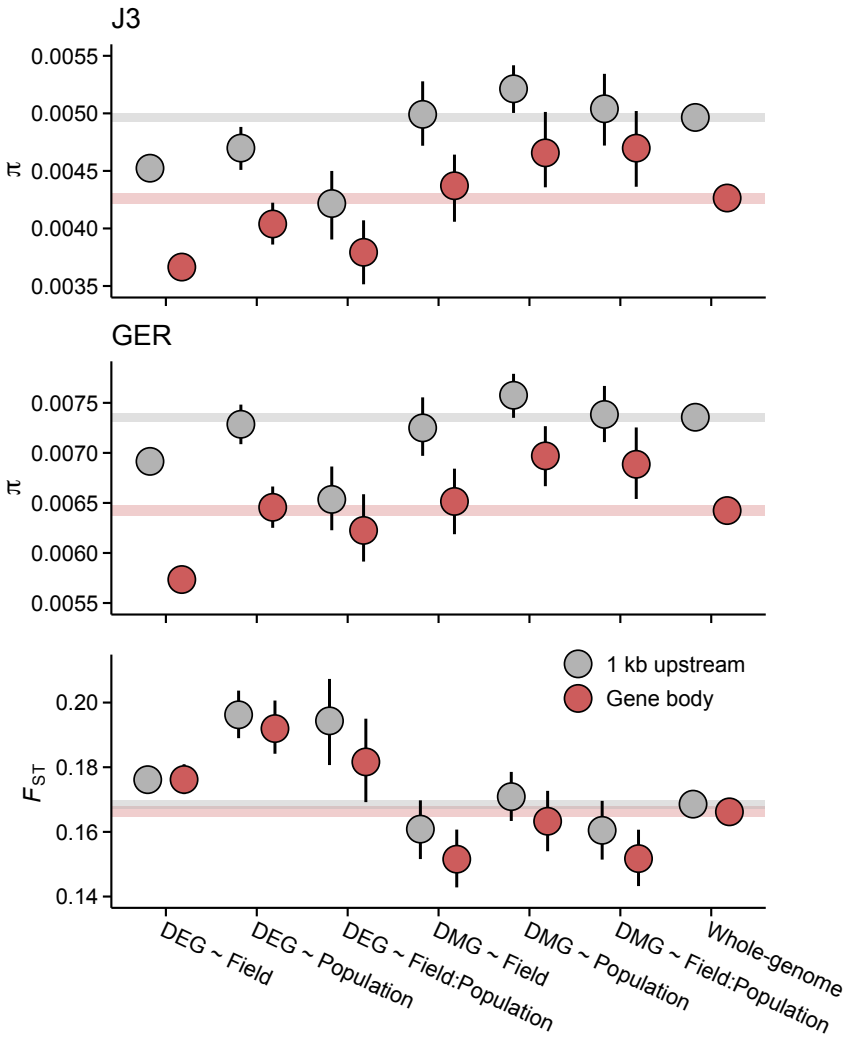

Figure 4. Pairwise nucleotide diversity (π) and F_{ST} at the candidate gene sets. π estimates are shown for J3 ($n = 22$) and GER ($n = 17$) populations. Estimates for J1 ($n = 9$) are shown in Fig. S5. F_{ST} was estimated across the three populations. See Fig. S7 for pairwise F_{ST} estimates. Error bars show 95% confidence intervals (CIs). Shaded areas mark the 95% CIs across all genes.

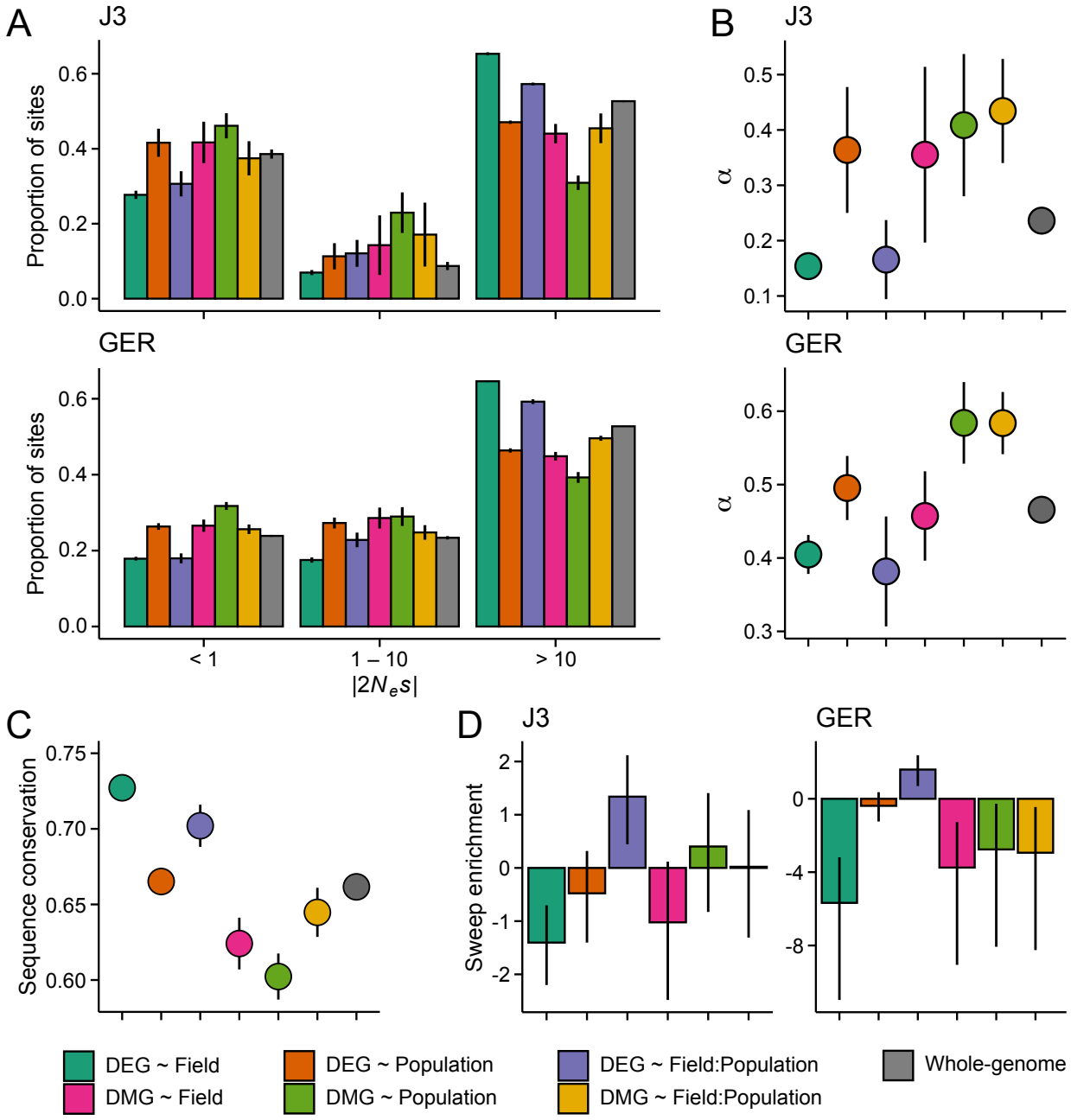

Figure 5. The efficacy of negative and positive selection at the candidate gene sets. Estimates are shown for J3 (n = 22) and GER (n = 17) populations. Estimates for J1 (n = 9) are shown in Fig. S8. **A:** The distribution of fitness effects (DFE) of new nonsynonymous variants. The mutations were divided into three bins based on the strength of purifying selection ($2N_es$): nearly neutral, intermediate, and deleterious, respectively. **B:** The proportion of sites fixed by positive selection (α). **C:** Average GERP (RS) scores estimated for each gene group. The scores were rescaled from 0 to 1 using the range of possible values for each site. **D**: The $\log_2$ odds ratio of association between selective sweeps and the candidate gene sets. For all panels, error bars show 95% CIs.

Environmental response at transposable elements

Besides genes, stressful environments may invoke plastic responses at TEs (Capy et al., 2000). To examine this in our data, we quantified DNA methylation and expression of individual TEs. In contrast to genes, we found that methylation levels across TEs were clearly influenced by the field sites (Fig. 6A). Compared to the low-altitude field, TEs at the high-altitude field site had higher methylation levels in all three cytosine contexts, but CHH methylation responded most strongly to the environment (Fig. 6A; logistic regression $\beta_{CG} = 0.07$, $\beta_{CHG} = 0.12$, $\beta_{CHH} = 0.17$; $P < 2 \times 10^{-16}$, LRT). Population effects were also evident in TE methylation: J1 and J3 exhibited clearer environmental responses compared to GER (Fig. 6A). As methylation is commonly involved in epigenetic silencing of TEs (Lisch, 2009), we explored a connection between TE methylation and TE expression. Consistent with the notion of TE silencing, expression levels were negatively correlated with methylation levels (Spearman's $\rho_{CG} = -0.28$, $\rho_{CHG} = -0.16$, $\rho_{CHH} = -0.14$; $P < 6 \times 10^{-10}$). However, increased methylation at the high-altitude field site did not lead to a large-scale silencing of TEs, as expression levels of both retrotransposons and DNA transposons were lower in the low-altitude field site (Figs. 6B and S9; $P < 2 \times 10^{-16}$, Wilcoxon rank-sum test). Rather than proactively suppressing TEs in the more stressful environment, this pattern suggests that hypermethylation at the high-altitude field site is a response to increased TE activation (Lisch, 2009; Lloyd and Lister, 2022). Furthermore, increased TE activation can directly influence gene expression, as genes whose promoters lie close to TEs may also get silenced by the DNA methylation (Horvath and Slotte, 2017; Wyler et al., 2020). Indeed, by examining the association between gene's expression level and its distance from a methylated TE, we found evidence that epigenetic regulation directed at TEs suppresses the expression of nearby genes (Fig. S10). TE methylation did not, however, appear to underlie the expression differences observed between the field sites (or *vice versa*, Secco et al. 2015), as we found that DEG ~ Field genes were, on average, further away from differentially methylated TEs than other genes (683 bp vs. 214 bp; $P = 3.9 \times 10^{-8}$, Wilcoxon rank-sum test).

Taken together, the methylation and expression results are consistent with broad-scale activation of TEs at the high-altitude field site. To explore how such activation shapes genome evolution, we identified TE insertion polymorphisms from the whole-genome data. Based on allele frequency spectra summarized with Tajima's *D* (Tajima, 1989), TEs segregate at lower frequencies in J3 than in J1 (Fig. 6C). Given that TEs generally have negative fitness consequences (Bourgeois and Boissinot, 2019), the lower frequency could be due to more efficient purging of TEs in J3. However, based on SNP data, the strength of selection in J3 appears equal (or weaker) to J1 (Figs. 6C and S11), pointing towards an alternative explanation: a recent burst of TE insertions can result

in an excess of rare variants, as not enough time has passed for neutral or slightly deleterious TEs to increase in frequency (Bergman and Bensasson, 2007). Furthermore, if TE activation has resulted in an excess of recent TE insertions, we might expect different effects on retrotransposons and DNA transposons: the probability of TEs impairing gene function is increased along with their activation, but as DNA transposons move by "cut-and-paste" mechanism and retrotransposons move by "copy-and-paste" mechanism (Wicker et al., 2007), the number of active DNA transposons may be more effectively reduced by purifying selection. Consistent with this expectation, we found an enrichment of retrotransposons and a deficit of DNA transposons in J3 (Figs. 6D and S12). Therefore, our results suggest that stress-induced activation of TEs has resulted in recent proliferation of retrotransposons and/or purging of DNA transposons in the alpine population. Finally, by examining the locations of the TE insertions, we discovered that many genes belonging to the six candidate groups harbored greater than expected numbers of TEs (Figs. 6E and S13). In particular, the DEG ~ Field:Population group was highly enriched for *Gypsy* (and to a lesser extent *Copia*) retrotransposons in the J3 population ($P = 1.3 \times 10^{-6}$, Fisher's exact test).

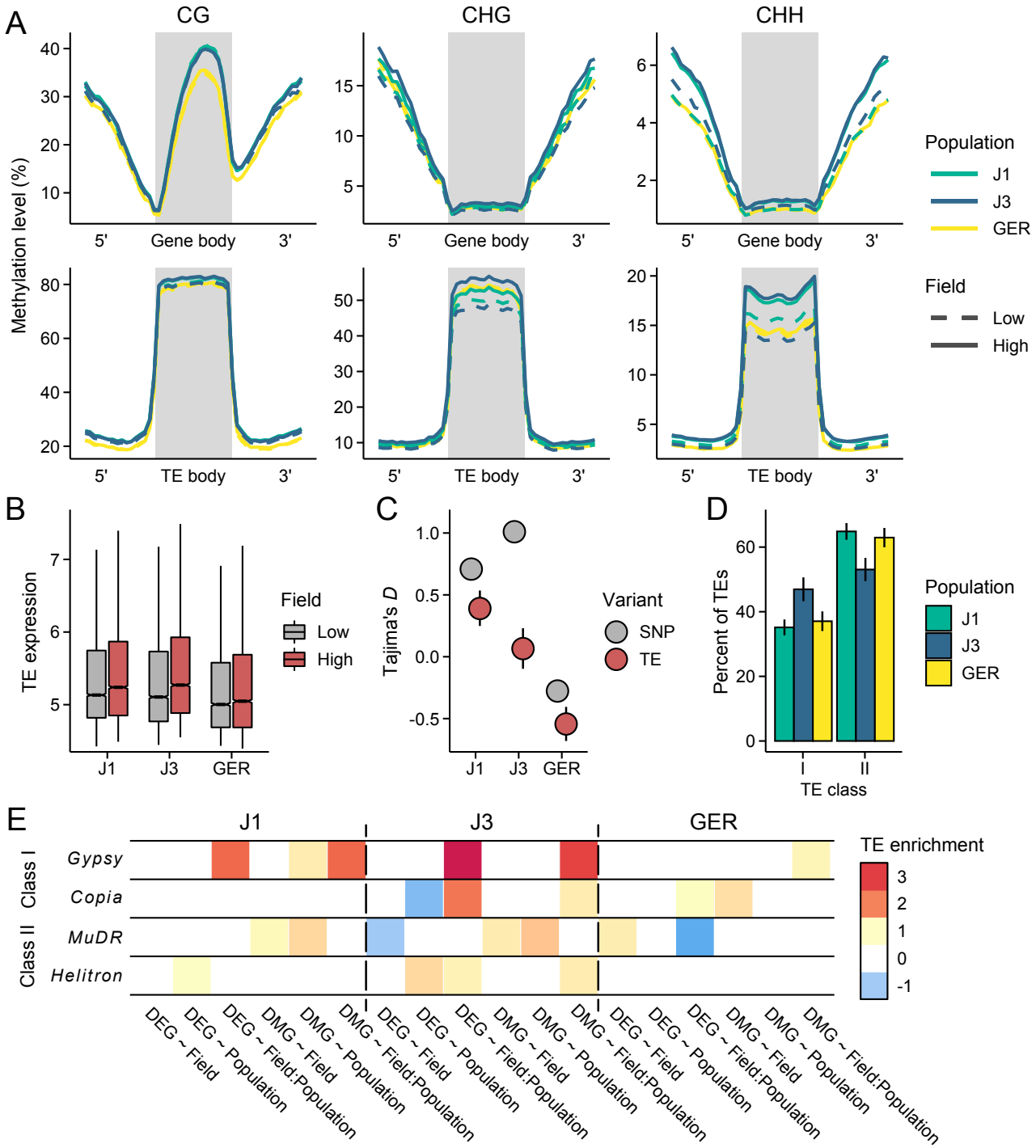

Figure 6. Environmental response at transposable elements (TEs). **A**: Methylation levels across meta-genes and meta-TEs, shown for CG, CHG, and CHH contexts. Note the difference in *y*-axis scales between the panels. **B**: TE expression at low- and high-altitude field sites. Read counts were normalized using variance stabilization transformation in DESeq2 (see Fig. S9 for results on different TE families and Fig. S14 for results on gene expression). **C**: Tajima's *D* for TEs and nonsynonymous SNPs in the three populations. **D:** The proportion of retrotransposons (class I) and DNA transposons (class II) in each population (see Fig. S12 for results on different TE families). **E**: The $\log_2$ odds ratio of association between TEs and the candidate gene sets. Results are shown for the four largest TE families found in gene bodies (see Fig. S13 for results on all TEs found in genes, including 1 kb up- and downstream regions). For C and D panels, error bars show 95% CIs.

Discussion

Gene expression is strongly influenced by the environment

We conducted a reciprocal transplant experiment to study environmentally induced variation in gene expression and DNA methylation in three populations of *A. lyrata*. Our study had the benefit of exposing the plants to full extent of environmental factors for an entire year, likely proving a more realistic view of transcriptome and DNA methylome responses in the face of environmental change than laboratory experiments (Savolainen et al., 2013). We found that gene expression variation was strongly plastic, as we detected many more DEGs between the field sites than between populations. In particular, the two Norwegian populations had highly similar transcriptional responses, likely owing to their close geographical and genetic similarity (diverged ~1700 year ago, Hämälä et al., 2018). However, we previously found evidence of local adaptation between these populations (Hämälä et al., 2018), indicating that some ecologically important traits are genetically differentiated despite the strong plasticity (e.g., flowering time, Fig 1C). Compared to the Norwegian populations, GER exhibited distinct expression responses, consistent with its greater genetic dissimilarity (diverged from the Norwegians ~ 150,000 years ago, Takou et al., 2021) and different growing environment (Leinonen et al., 2009). These transcriptional differences may have been amplified by the timing of our sampling (late August), as the local and nonlocal plants likely differed in their physiological preparation for winter.

The DNA methylome is largely insensitive to the environment

In contrast to gene expression, variation in DNA methylation was mainly shaped by population history, and environmental effects were largely restricted to TEs. As such, our experiment revealed little evidence of environmentally induced DNA methylation influencing gene expression. Studies conducted in controlled conditions have discovered similar results in the sister species *A. thaliana*: patterns of DNA methylation have been weakly influenced by temperature (Dubin et al., 2015), phosphate starvation (Secco et al., 2015), drought stress (Ganguly et al., 2017), and light intensity (Ganguly et al., 2018). Taken together, these results suggest that the DNA methylome in *Arabidopsis* species is largely insensitive to environmental factors and thus unlikely to induce large-scale changes in gene expression within a single generation. We note, however, that here we only studied DNA methylation, whereas other epigenetic modifications might be more readily induced by the environment. For example, histone modifications are known to exert a major influence on gene expression (Lloyd and Lister, 2022), and evidence is accumulating that some of them are stably inherited from parent to offspring (Escobar et al., 2021). Therefore, future work studying environmental induction and transgenerational inheritance of histone modifications in natural

conditions holds promise to expand our understanding of the evolutionary consequences of epigenome plasticity (Husby, 2022). In any case, we discovered greater than expected overlap between genes that were differentially expressed and methylated between the populations, suggesting that changes in GBM, when they accumulate over generations, do influence patterns of gene expression (Takuno et al., 2017). In particular, genes involved in calcium ion transport were strongly enriched among the DMGs, demonstrating that the methylation status of these stress-reactive genes (Patra et al., 2021) contributes to population differentiation and potentially local adaptation in *A. lyrata*.

Environmentally responsive genes evolve under strong selective constraint

We found that environmentally induced expression responses were highly consistent across our experimental plants, indicating that regulatory elements are largely conserved between the three *A. lyrata* populations. Previous studies examining environmentally responsive genes have found that the regulatory conservation is frequently associated with the conservation of the gene product (He et al., 2021; Hodgins et al., 2016; Hunt et al., 2013; Lasky et al., 2014; Lowry et al., 2013; Zhang et al., 2017). Although our results qualitatively support these findings, as genes showing consistent expression responses to the environment (DEG ~ Field) were under stronger purifying selection than those showing genetically variable responses (DEG ~ Field:Population), we discovered that both gene sets exhibited stronger than expected signals of evolutionary constraint. By examining nucleotide diversity and differentiation at the promoter regions, we further found that population-specific selection at *cis*-regulatory elements has likely contributed to the maintenance of heritable variation at the DEG ~ Field:Population genes. We note, however, that our experimental design did not allow us to associate selective signals with *trans*-regulatory variants, which may have also played a role in the observed expression differentiation (Hämälä et al., 2020; Lopez-Arboleda et al., 2021).

Our results suggest that genes whose products are involved in conserved processes are prone to exhibit plastic expression responses to environmental stressors, but genetic variation in the environmental response still arises through changes at regulatory elements (*cis* and potentially *trans*). However, the expression differentiation was almost exclusively found between the Norwegian populations and GER, indicating that such rewiring of the regulatory network primarily happens over long evolutionary timescales. Therefore, the strong selective constraint and low genetic diversity at environmentally responsive genes could limit the adaptive potential of these *A. lyrata* populations, especially in cases where environmental perturbations exceed the buffering mechanism provided by phenotypic plasticity (Ghalambor et al., 2007; Price et al., 2003).

Plastic responses at TEs create novel genetic variation

Although DNA methylation at genic regions was largely insensitive to the environment, we discovered clear environmental effects in TE methylation. TEs at the high-altitude field site had increased methylation and expression levels, suggestive of stress-induced activation of TEs. Consistent with the effects of TE activation, plants native to the alpine environment harbored an excess of rare TEs. On average, such burst of TE insertions is likely to have detrimental effects, as TEs can impair gene function by either directly disrupting coding regions and regulatory elements (Bourgeois and Boissinot, 2019) or by influencing the expression of nearby genes through the spread of epigenetic silencing (Horvath and Slotte, 2017; Wyler et al., 2020). Indeed, we also discovered that genes whose promoters lie close to methylated TEs had reduced expression levels. On the other hand, TE mobility is expected to create novel genetic variants that may, in some cases, facilitate adaptation under new selective environments (Capy et al., 2000; Casacuberta and González, 2013; McClintock, 1984).

Here, we found that genes exhibiting genetically variable responses to the environment harbored an enrichment of retrotransposons in the alpine population. Although we lack functional validation for the detected TEs, such enrichment demonstrates that TEs have the potential to create new heritable variation that is relevant for environmental adaptation. The role of TEs in creating novel functional variants is also see in results by Quadrana et al. (2019), who found experimental evidence that retrotransposons of the *Copia* family preferentially integrate into the bodies of environmentally responsive genes in *A. thaliana*. This integration is likely mediated by the histone variant H2A.Z, which is commonly found within the gene bodies of temperature sensing genes (Talbert and Henikoff, 2014), providing a possible mechanism for the observed TE accumulation in the J3 population. However, although we also detected an enrichment of *Copia* elements in the environmentally responsive genes, the enrichment of *Gypsy* elements was far greater (odds ratio: 4.5 vs. 11.7). This pattern indicates that the *Gypsy* family exhibits a similar integration preference in *A. lyrata* and/or that *Copia* elements have been more readily purged by purifying selection. If similar dynamics exist in other Brassicaceae species, it could explain the apparent excess of *Gypsy* TEs in the alpine species *Arabis alpina* and *Draba nivalis* (Nowak et al., 2021). In any case, our results suggest that novel genetic polymorphisms created by stress activated TEs could be beneficial for *A. lyrata* under new selective environments, particularly as we discovered that most environmentally responsive genes have weak potential for evolutionary change.

Conclusions

By studying gene expression and DNA methylation in natural conditions, we have gained new insights into evolutionary processes shaping plastic and genetic variation. We observed contrasting results between the molecular phenotypes, as gene expression was primarily influenced by the environment and DNA methylation was primarily influenced by population history. Although we found that most environmentally responsive genes are slowly evolving, our results suggest that novel heritable variation may be rapidly created by TEs integrating into the gene bodies of transcriptionally variable genes. Overall, these findings have important implications for adaptation under climate change, as they suggest that strong evolutionary constraint at environmentally responsive genes and their regulatory elements may limit organisms' potential to adapt to future climates, and that environmentally induced changes in the DNA methylome are likely too modest to counteract the adaptation lag caused by rapidly changing climate. However, these effects may be alleviated by stress-induced activation of TEs, which could facilitate adaptation under new selective environments by creating novel genetic variation.

Material and methods

Reciprocal transplant experiment and sample collection

We conducted a reciprocal transplant experiment to study altitude adaptation among Norwegian populations of *Arabidopsis lyrata* ssp. *petraea*. *A. lyrata* is a predominantly outcrossing, perennial herb with a wide circumpolar distribution across the northern hemisphere (Jalas and Suominen, 1994). Detailed experimental designs are presented in Hämälä et al. (2018) and here we give a brief explanation of the relevant methodology. Seed material was collected from two locations around the Jotunheimen national park, Lom (61.84°N, 8.57°E; altitude 300 m.a.s.l.) and Spiterstulen (61.62°N, 8.40°E; altitude 1100 m.a.s.l.). We additionally used an *A. lyrata* population from Germany (49.65°N, 11.48°E; altitude 400 m.a.s.l.) as a comparison group. In Hämälä et al. (2018), the Jotunheimen alpine region was represented by four populations, which were abbreviated as J1 to J4. To keep the naming convention consistent with previous work (Hämälä et al., 2018; Hämälä and Savolainen, 2019), we refer to these populations as J1 (Lom), J3 (Spiterstulen), and GER (Germany).

Plants were initially grown in a growth chamber and crossed to produce full-sib families for each population. The crossing progeny were germinated and pre-grown in a greenhouse at University of Oulu, Finland for about two months. In August 2014, we established two experimental fields in Jotunheimen, Norway: a low-altitude field site in Lom (300 m.a.s.l.), close to the natural growing

environment of J1, and a high-altitude field site in Spiterstulen (1100 m.a.s.l.), close to the natural growing environment of J3 (Fig. 1A). Using multi-year fitness measurements, we found evidence of home-site advantage between the J1 and J3 populations, whereas GER had poor fitness at both field sites (Hämälä et al., 2018). After one year, we chose six individuals per population from both fields (a total of 36 individuals) for sample collection and sequencing. Individuals from both Norwegian populations represented six seed families, with individuals from the same families collected from both fields, whereas GER was represented by ten seed families (due to high mortality, many GER seed families were not available in both fields). Leaf samples were collected during two consecutive days (21st and 22nd of August 2015, between 18:30 and 19:30), immersed in RNAlater stabilizing solution (Thermo Fisher Scientific), and kept at –20°C until library preparation.

Transcriptome sequencing

Total RNA was extracted using a RNeasy Plant Mini kit (QIAGEN), following the manufacturer's instructions. RNA integrity number (RIN) and concentration were determined using an Agilent RNA 6000 Pico kit (Agilent). NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, NEB#E7490) from NEBNext Ultra Directional RNA Library Prep kit for Illumina was used to create stranded RNA-seq libraries according to the manufacturer's instructions. RNA-seq libraries were quantified using KAPA Library Quantification kits (Kapa Biosystems) in combination with an Agilent High Sensitivity DNA kit (Agilent Genomic). The RNA-seq libraries were sequenced on four lanes of Illumina HiSeq2000 (paired-end 100 bp) at the Institute for Molecular Medicine Finland (FIMM), University of Helsinki.

After removing low quality reads and sequencing adapters with Trimmomatic (Bolger et al., 2014), we used STAR (Dobin et al., 2013) to align reads to the *A. lyrata* reference genome (Hu et al., 2011) and HTSeq (Anders et al., 2015) to count reads overlapping each gene model. To quantify TE expression, we first identified TEs using RepeatMasker (Smit et al. 2015) and a library *A. lyrata* consensus TEs from RepetDB (Amselem et al., 2019), covering the main orders of both retrotransposons and DNA transposons (we ignored pseudogenes and TEs without classification). This resulted in the discovery of 80,624 individual TEs across the eight main chromosomes. We then used HTSeq to count reads mapping to TEs. To exclude reads mapping to genes, we supplied an annotation file to HTSeq that contained both TEs and genes (including UTRs, introns, and exons).

Whole-genome bisulfite sequencing

To examine patterns of DNA methylation, we used 24 samples (four individuals per population from both fields) for whole-genome bisulfite sequencing (WGBS). DNA was extracted using a DNeasy Plant Mini kit (QIAGEN), following the manufacturer's instructions. We prepared the WGBS libraries according to MethylC protocol (Urich et al., 2015) and quantified them using a combination of KAPA Library Quantification kits (Kapa Biosystems) and an Agilent High Sensitivity DNA kit (Agilent Genomic). The WGBS libraries were sequenced on six lanes of Illumina HiSeq2000 (paired-end 100 bp) at FIMM, University of Helsinki.

Low-quality reads and sequencing adapters were removed with Trimmomatic (Bolger et al., 2014) and the surviving reads aligned to the *A. lyrata* reference genome (Hu et al., 2011) using Bismark (Krueger and Andrews, 2011) and Bowtie2 (Langmead and Salzberg, 2012). We removed duplicated reads using deduplicate_bismark script from Bismark and estimated the number of methylated and unmethylated reads at each cytosine context (CG, CHG, CHH, where H is A, T, or C) using a Bismark script bismark_methylation_extractor. We then removed sites with known C to T and A to G (the reverse orientation of C to T) SNPs from the methylation calls, because such SNPs will be incorrectly called as unmethylated by Bismark. To do so, we first identified SNPs using the RNA-seq data, which contained the same individuals as used for the WGBS (see below for details on SNP calling). For regions not covered by the RNA-seq in at least 50% of individuals, we used whole-genome sequencing data from independent J1 (n = 9), J3 (n = 22), and GER (n = 17) individuals. By examining the methylation patterns of chloroplast DNA, which is expected to be naturally unmethylated, we estimated an overall conversion efficacy of 99.1% for CG context, 99.4% for CHG context, and 99.7% for CHH context.

Differential expression analysis

Analyses of expression data were conducted using the R package DESeq2 (Love et al., 2014). Because our samples were sequenced on four lanes, we first determined whether sequencing batches may confound the detection of differentially expressed genes (DEGs). To do so, we examined the results of stably expressed reference genes, which, compared to other genes, are less likely to be influenced by treatment (here field site and population) but equally likely influenced by batch effect. We used the direct *A. thaliana* orthologs of 20 genes identified by Czechowski et al. (2005) and 13 genes identified by Kudo et al. (2016). We normalized the read counts with variance stabilization transformation in DESeq2 and used principal components analysis (PCA) to identify the main sources

of variation. Although these reference genes have shown stable expression across treatments in *A. thaliana*, four first principal components (PCs) from both gene sets were primarily impacted by the field site and population rather than sequencing batch (Fig. S15). We also found no DEGs (see below for details) between batches B1 and B2 (containing samples from the high-altitude field) or between batches B3 and B4 (containing samples from the low-altitude field), whereas four of the 33 genes were differentially expressed between the field sites. Therefore, the sequencing batches likely have only a minor influence on our results, and by focusing on top-ranking DEGs, we are more likely to capture effects resulting from the growing environment and population history of our experimental plants.

Using genes with median read counts > 1, we searched for DEGs between field sites and populations with DESeq2. To identify population-specific DEGs, we fit a single dispersion parameter to each gene, and used the DESeq2's contrast function and Wald test to detect differences between the field sites. Likelihood ratio tests (LRTs) were used to identify global DEGs due to field site or population: we compared the fit of a full model, containing field, population, and their interaction as predictors, to a reduced model with one of the predictors removed. To focus on top-ranking DEGs, we took a relatively stringent approach to account for multiple testing and required that outliers have a Bonferroni corrected P value < 0.05. However, as DEGs showing an interaction between field site and population (DEG ~ Field:Population) only had 28 genes with Bonferroni corrected $P < 0.05$, we also used a more lenient approach and required that outliers from the interaction tests had a false discovery rate-based Q value (Storey, 2002) < 0.05, leading to 477 DEGs. By being more lenient with multiple comparison, we reach a sample size needed to examine selective signals at these genes, although at the same time we risk accepting more false positives due to confounding factors.

Differential methylation analysis

We removed sites with ≤ 3 reads and defined methylation levels as the proportion of unconverted cytosines. To associate the methylation patterns more closely with gene expression levels and measures of selective constraint, we primarily focused on CG methylation at gene bodies (i.e., gene body methylation, GBM). Using cytosines with < 50% missing data, we calculated the methylated (2 × the sum of methylation proportions) and unmethylated (2 × the sum of 1 – methylation proportions) allelic dosage for each individual at a given gene and tested the effects of the field site and population using logistic regression and LRTs. The LRT-based P values were transformed to Q values (Storey, 2002) to account for multiple testing. To define differentially methylated genes (DMGs), we required

that the genes had ≥ 10 cytosines and $Q < 0.05$. Given that CG methylation was strongly influenced by population structure (Fig 2A), we corrected *P* values from the field + population vs. field LRTs (i.e., DMGs due to population) using genomic inflation factor (Devlin and Roeder, 1999).

Gene ontology enrichment analysis

We conducted a gene ontology (GO) enrichment analysis to identify biological processes associated with DEGs and DMGs. To do so, we used the direct *A. thaliana* orthologs of 21,784 genes to define GO terms and tested for an enrichment of biological processes (molecular functions and cellular components were ignored) using hypergeometric tests and *Q*-values (Storey, 2002). We then used REVIGO (Supek et al., 2011) to combine redundant GO terms.

Whole-genome sequence data

To study the influence of selection on patterns of gene expression and methylation, we used previously published whole-genome sequence data from J1 ($n = 9$), J3 ($n = 22$), and GER ($n = 17$) individuals (Hämälä et al., 2018; Mattila et al., 2017; Takou et al., 2021). We removed low-quality reads and sequencing adapters with Trimmomatic (Bolger et al., 2014) and aligned the reads to the *A. lyrata* reference genome (Hu et al., 2011) with BWA-MEM (Li, 2013). We removed duplicated reads with Picard tools (https://broadinstitute.github.io/picard/) and realigned indels with GATK (McKenna et al., 2010).

To incorporate genotype uncertainty directly into our analyses, we used ANGSD (Korneliussen et al., 2014) to estimate genotype likelihoods and probabilities at each mono- and biallelic site. We used the GATK likelihood model and required reads to map uniquely, have mapping quality ≥ 30, and base quality ≥ 20. For each population, we had ANGSD estimate posterior genotype probabilities using allele frequency as a prior. To account for genetic structure in multi-population data, we estimated genotype probabilities using PCAngsd (Meisner and Albrechtsen, 2018), which employs a model that incorporates the effects of population structure (in the form of PCs) in the prior. We then estimated the allelic dosage, or the expected genotype, from the posterior probabilities as $\mathrm{E}[G] = \sum_{g=0}^{2} gP(G = g)$, where G is the genotype. As the identification of selective sweeps (see below) required VCF files as input, we further used ANGSD to call genotypes from the posterior probabilities. All analyses described in the subsequent sections were conducted using the whole-genome datasets.

Genetic diversity and differentiation

To assess how selection has acted on genes determined by their DEG and DMG status, we first estimated pairwise nucleotide diversity (π) and F_{ST} for each candidate gene set. π was estimated across all callable (variant and invariant) sites to avoid biasing the estimates by missing sites. We used the method by Weir and Cockerham (1984) to estimate F_{ST} for both three- and two-population comparisons. We estimated π and F_{ST} for each candidate gene group and defined confidence intervals (CIs) for the estimates by resampling with replacement across genes 1000 times.

Distribution of fitness effects

We examined the strength of selective constraint on the candidate gene sets by modelling the distribution of fitness effects (DFE). To do so, we estimated folded SFS for synonymous (4-fold) and nonsynonymous (0-fold) sites using ANGSD (Korneliussen et al., 2014). We then used ∂a∂i (Gutenkunst et al., 2009) to fit three-epoch demographic models to synonymous SFS and inferred a deleterious gamma DFE for nonsynonymous sites, conditional on the demography, using fit∂a∂i (Kim et al., 2017). Population mutation rates ($\theta = 4N_e\mu$) were estimated from the synonymous data and multiplied by 2.76 (Takou et al., 2021) to approximate θ at nonsynonymous sites. This approach, as opposed to optimizing θ along with the DFE parameters, can take into account variants expected to be missing due to strong purifying selection (Kim et al., 2017). Demographic parameters were estimated from the whole-genome data and fixed for the analysis of candidate gene sets. We estimated CIs for the DFE parameters by fitting the models to 500 parametric bootstrap SFS. We assumed that counts in the bootstrap replicates follow a multinomial distribution with number of trials corresponding to total number of sites in the SFS and the probability of success corresponding to proportion of sites in a given allele frequency bin. We then discretized the DFE into three bins, nearly neutral ($2N_es < 1$), intermediate ($1 \leq 2N_es \leq 10$), and deleterious ($2N_es > 10$), and compared bins in the candidate gene sets to the genome-wide average. *P* values were defined as twice the proportion of overlapping bootstrap replicates in each bin.

The rate of adaptive evolution

We examined the efficacy of positive selection by estimating the proportion of sites fixed by positive selection (α) (Smith and Eyre-Walker, 2002). Because α estimation requires outgroup information,

we inferred unfolded SFS for synonymous and nonsynonymous sites using a method that accounts for uncertainty in the assignment of ancestral vs. derived alleles (Keightley and Jackson, 2018). We used *A. lyrata – A. thaliana – Capsella rubella – Arabis alpina* whole-genome alignments from Hämälä and Tiffin (2020), requiring each site to have outgroup information in at least two of the three species. Following Messer and Petrov (2013), we estimated α for each polymorphic allele frequency bin of the unfolded SFS as:

$$\alpha(x) = 1 - \frac{p_N(x)/p_S(x)}{d_N/d_S},$$

where $p_N(x)$ and $p_S(x)$ are the number of nonsynonymous and synonymous polymorphic sites at frequency x, and d_N and d_S are the number of nonsynonymous and synonymous fixes sites. We then used the R package nls2 (Grothendieck, 2013) to fit an asymptotic function of the form: $\alpha(x) = a + be^{-cx}$ to the data. As in Haller and Messer (2017), we used the "brute-force" algorithm to find suitable starting values for the free parameters (a, b, c) and refined the values using a second step of standard nonlinear least-squares regression. We used LRTs to determine whether the fitted functions differ between the candidate gene sets and the genome-wide average, and estimated CIs by fitting the models to 500 parametric bootstrap SFS.

Sequence conservation

To examine the conservation of selective processes, we used GERP++ (Davydov et al., 2010) to estimate nucleotide conservation at DEGs and DMGs. We chose 25 eudicot species belonging to the clade Superrosidae (Table S1), whose divergence times in relation to *A. lyrata* ranged from 1.6 million years (*A. halleri*) to 123 million years (*Vitis vinifera*) (Hohmann et al., 2015). We first identified homologs by conducting BLAST queries against protein databases constructed for the species, keeping only the best match with alignment *e*-value < 1×10^{-5} for each gene. For sets of homologs with ≥ 13 species, we aligned the coding sequences with MAFFT (Nakamura et al., 2018). As GERP++ requires an evolutionary tree, we used the R package phangorn (Schliep, 2011) to estimate a maximum likelihood tree across 1000 randomly selected genes with no missing species. Using the species tree and multiple alignments, we had GERP++ estimate the rejected substitutions (RS) score for sites in the *A. lyrata* coding sequence, quantifying the level of nucleotide conservation in relation to neutral substitution rate. Last, using the range of possible values at each site, we rescaled the RS scores from 0 to 1, where 0 is the weakest possible conservation and 1 is the strongest.

Scan for selective sweeps

To identify genes that have undergone recent selective sweeps, we used RAiSD (Alachiotis and Pavlidis, 2018) to scan the genomes for patterns of segregating sites, linkage-disequilibrium, and nucleotide diversity indicative of positive selection. We estimated the composite statistic μ in 50 SNP sliding windows and characterized selective signals at each gene using the maximum μ value of windows within 2 kb. We then ran RAiSD on simulated neutral data (Hämälä and Savolainen, 2019) to find outliers. We considered observed estimated exceeding 99% of the simulated values as reflecting selective sweeps.

Identification of TE insertion polymorphisms

We identified TE insertions polymorphisms using PoPoolationTE2 (Kofler et al., 2016). Following the recommended workflow, we masked the *A. lyrata* consensus TEs (Amselem et al., 2019) from the reference genome using an iterative mapping approach: simulated TE reads were aligned to the genome with BWA-MEM (Li, 2013), aligned regions masked from the reference with BEDtools (Quinlan and Hall, 2010), and the process repeated until no new unmasked regions were found. We then merged the consensus TEs with the masked genome, aligned quality-trimmed DNA-seq reads to the TE merged reference using BWA-MEM, and removed duplicated reads using Picard tools (https://broadinstitute.github.io/picard/). We only used samples with an average coverage ≥ 10× across the eight main chromosomes (this excluded five individuals from J3 and two from GER) and required each site to have a minimum coverage of 6×. We then had PoPoolationTE2 estimate the proportion of reads supporting TEs in each individual and filtered the list to remove overlapping TEs. Last, we discretized the read proportions into genotypes: > 0.85 TE homozygote, 0.85 – 0.15 TE heterozygote, and < 0.15 non-TE homozygote.

Data availability

The transcriptome and whole-genome bisulfite sequencing data are available at NCBI SRA: PRJNA459481. Scripts for conducting the analyses are available at: https://github.com/thamala/lyrataRnaMet.

Author contributions

T.H., H.K., N.A., and O.S. designed the study. T.H. conducted the field experiments. W.N. and N.A. conducted the laboratory work. T.H., W.N., and N.A. analyzed the data. T.H. wrote the manuscript with contributions from all authors.

Acknowledgments

We thank Marit and Lars Bakkom for help with the field experiments, Soile Alatalo for help with sequencing library preparation, and Margarita Takou and members of the Pyhäjärvi lab for comments on the manuscript. Computational resources were provided by CSC – Finnish IT Center for Science. This work was supported by the Academy of Finland's Research Council for Biosciences, Health and the Environment (decision 339702 to T.H. and 132611 to O.S.).

References

Alachiotis N, Pavlidis P. 2018. RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Communications Biology* **1**. doi:10.1038/s42003-018-0085-8

Alberto FJ, Aitken SN, Alía R, González-Martínez SC, Hänninen H, Kremer A, Lefèvre F, Lenormand T, Yeaman S, Whetten R, Savolainen O. 2013. Potential for evolutionary responses to climate change - evidence from tree populations. *Global Change Biology* **19**:1645–1661. doi:10.1111/gcb.12181

Amselem J, Cornut G, Choisne N, Alaux M, Alfama-Depauw F, Jamilloux V, Maumus F, Letellier T, Luyten I, Pommier C, Adam-Blondon AF, Quesneville H. 2019. RepetDB: A unified resource for transposable element references. *Mobile DNA* **10**:1–8. doi:10.1186/s13100-019-0150-y

Anders S, Pyl PT, Huber W. 2015. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**:166–169. doi:10.1093/bioinformatics/btu638

Baduel P, Leduque B, Ignace A, Gy I, Gil J, Loudet O, Vincent C, Quadrana L, Baduel P, Leduque B, Ignace A, Gy I, Gil J. 2021. Genetic and environmental modulation of transposition shapes the evolutionary potential of *Arabidopsis thaliana*. *Genome Biology* **22**:1–26.

Baldwin JM. 1896. A new factor in evolution. *The American Naturalist* **30**:441–451. doi:10.1086/276408

Bergman CM, Bensasson D. 2007. Recent LTR retrotransposon insertion contrasts with waves of non-LTR insertion since speciation in Drosophila melanogaster. *PNAS* **104**:11340–11345. doi:10.1073/pnas.0702552104

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120. doi:10.1093/bioinformatics/btu170

Bourgeois Y, Boissinot S. 2019. On the population dynamics of junk: A review on the population genomics of transposable elements. *Genes* **10**. doi:10.3390/genes10060419

Capy P, Gasperi G, Biémont C, Bazin C. 2000. Stress and transposable elements: co-evolution or useful parasites? *Heredity* **85**:101–106. doi:10.1046/j.1365-2540.2000.00751.x

Casacuberta E, González J. 2013. The impact of transposable elements in environmental adaptation. *Molecular Ecology* **22**:1503–1517. doi:10.1111/mec.12170

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible W-R. 2005. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiology* **139**:5–17. doi:10.1104/pp.105.063743

Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLOS Computational Biology* **6**:e1001025. doi:10.1371/journal.pcbi.1001025

Devlin B, Roeder K. 1999. Genomic control for association studies. *Biometrics* **55**:997–1004. doi:10.1111/j.0006-341X.1999.00997.x

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. doi:10.1093/bioinformatics/bts635

Dubin MJ, Zhang P, Meng D, Remigereau MS, Osborne EJ, Casale FP, Drewe P, Kahles A, Jean G, Vilhjálmsson B, Jagoda J, Irez S, Voronin V, Song Q, Long Q, Rätsch G, Stegle O, Clark RM, Nordborg M. 2015. DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. *eLife* **4**:1–23. doi:10.7554/eLife.05255

Escobar TM, Loyola A, Reinberg D. 2021. Parental nucleosome segregation and the inheritance of cellular identity. *Nat Rev Genet* **22**:379–392. doi:10.1038/s41576-020-00312-w

Falconer DS. 1981. Introduction to quantitative genetics. New York: Longman,.

Ganguly DR, Crisp PA, Eichten SR, Pogson BJ. 2018. Maintenance of pre-existing DNA methylation states through recurring excess-light stress. *Plant, Cell & Environment* **41**:1657–1672. doi:10.1111/pce.13324

Ganguly DR, Crisp PA, Eichten SR, Pogson BJ. 2017. The *Arabidopsis* DNA methylome is stable under transgenerational drought stress. *Plant Physiology* **175**:1893–1912. doi:10.1104/pp.17.00744

Ghalambor CK, McKAY JK, Carroll SP, Reznick DN. 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology* **21**:394–407. doi:10.1111/j.1365-2435.2007.01283.x

Grant V. 1977. Organismic evolution. San Francisco: Freeman.

Grothendieck G. 2013. Non-linear regression with brute force. *R package version 02*.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics* **5**:e1000695. doi:10.1371/journal.pgen.1000695

Haller BC, Messer PW. 2017. AsymptoticMK: A web-based tool for the asymptotic McDonald-Kreitman test. *G3: Genes, Genomes, Genetics* **7**:1569–1575. doi:10.1534/g3.117.039693

Hämälä T, Gorton AJ, Moeller DA, Tiffin P. 2020. Pleiotropy facilitates local adaptation to distant optima in common ragweed (*Ambrosia artemisiifolia*). *PLOS Genetics* **16**:e1008707. doi:10.1371/journal.pgen.1008707

Hämälä T, Mattila TM, Savolainen O. 2018. Local adaptation and ecological differentiation under selection, migration and drift in *Arabidopsis lyrata*. *Evolution* **72**:1373–1386. doi:10.1111/evo.13502

Hämälä T, Savolainen O. 2019. Genomic patterns of local adaptation under gene flow in *Arabidopsis lyrata*. *Molecular Biology and Evolution* **36**:2557–2571. doi:10.1093/molbev/msz149

Hämälä T, Tiffin P. 2020. Biased gene conversion constraints adaptation in *Arabidopsis thaliana*. *Genetics* **215**:831–846. doi:10.1534/genetics.120.303335

Hämälä T, Wafula EK, Guiltinan MJ, Ralph PE, DePamphilis CW, Tiffin P. 2021. Genomic structural variants constrain and facilitate adaptation in natural populations of *Theobroma*

cacao, the chocolate tree. *Proceedings of the National Academy of Sciences* **118**:e2102914118. doi:10.1073/pnas.2102914118
He F, Steige KA, Kovacova V, Göbel U, Bouzid M, Keightley PD, Beyer A, Meaux J de. 2021. Cis-regulatory evolution spotlights species differences in the adaptive potential of gene expression plasticity. *Nature Communications* **12**:1–13. doi:10.1038/s41467-021-23558-2
Hodgins KA, Yeaman S, Nurkowski KA, Rieseberg LH, Aitken SN. 2016. Expression Divergence Is Correlated with Sequence Evolution but Not Positive Selection in Conifers. *Molecular Biology and Evolution* **33**:1502–1516. doi:10.1093/molbev/msw032
Hodgins-Davis A, Townsend JP. 2009. Evolving gene expression: From G to E to G x E. *Trends Ecol Evol* **24**:649–658. doi:10.1016/j.tree.2009.06.011
Hof AE van't, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. 2016. The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534**:102–105. doi:10.1038/nature17951
Hohmann N, Wolf EM, Lysak MA, Koch MA. 2015. A time-calibrated road map of brassicaceae species radiation and evolutionary history. *Plant Cell* **27**:2770–2784. doi:10.1105/tpc.15.00482
Horvath R, Josephs EB, Pesquet E, Stinchcombe JR, Wright SI, Scofield D, Slotte T. 2021. Selection on accessible chromatin regions in *Capsella grandiflora*. *Molecular Biology and Evolution* **38**:5563–5575. doi:10.1093/molbev/msab270
Horvath R, Slotte T. 2017. The role of small RNA-based epigenetic silencing for purifying selection on transposable elements in *Capsella grandiflora*. *Genome Biology and Evolution* **9**:2911–2920. doi:10.1093/gbe/evx206
Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J-F, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, Haberer G, Hollister JD, Ossowski S, Ottilar RP, Salamov AA, Schneeberger K, Spannagl M, Wang X, Yang L, Nasrallah ME, Bergelson J, Carrington JC, Gaut BS, Schmutz J, Mayer KFX, Van De Peer Y, Grigoriev I V, Nordborg M, Weigel D, Guo Y-L. 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics* **43**:476–483. doi:10.1038/ng.807
Hunt BG, Ometto L, Keller L, Goodisman MAD. 2013. Evolution at two levels in fire ants: the relationship between patterns of gene expression and protein sequence evolution. *Mol Biol Evol* **30**:263–271. doi:10.1093/molbev/mss234
Husby A. 2022. Wild epigenetics: insights from epigenetic studies on natural populations. *Proceedings of the Royal Society B: Biological Sciences* **289**:20211633. doi:10.1098/rspb.2021.1633
Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J. 2011. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* **472**:115–119. doi:10.1038/nature09861
Jalas J, Suominen J. 1994. Atlas florae europaea. Distribution of vascular plants in Europe. 10: Cruciferae (Sisymbrium to Aubrieta), Feddes Repertorium. The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo. doi:10.1002/fedr.19951060115
Keightley PD, Jackson BC. 2018. Inferring the probability of the derived vs. the ancestral allelic state at a polymorphic site. *Genetics* **209**:897–906. doi:10.1534/genetics.118.301120
Kim BY, Huber CD, Lohmueller KE. 2017. Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. *Genetics* **206**:345–361. doi:10.1534/genetics.116.197145
Koenig D, Jiménez-Gómez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, Kumar R, Covington MF, Devisetty UK, Tat AV, Tohge T, Bolger A, Schneeberger K, Ossowski S, Lanz C, Xiong G, Taylor-Teeples M, Brady SM, Pauly M, Weigel D, Usadel B, Fernie AR, Peng J, Sinha NR, Maloof JN. 2013. Comparative transcriptomics reveals patterns of

selection in domesticated and wild tomato. *Proc Natl Acad Sci U S A* **110**:E2655-2662. doi:10.1073/pnas.1309606110

Kofler R, Gómez-Sánchez D, Schlötterer C. 2016. PoPoolationTE2: Comparative Population Genomics of Transposable Elements Using Pool-Seq. *Molecular Biology and Evolution* **33**:2759–2764. doi:10.1093/molbev/msw137

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15**:1471–2105. doi:10.1186/s12859-014-0356-4

Kou Y, Liao Y, Toivainen T, Lv Y, Tian X, Emerson JJ, Gaut BS, Zhou Y. 2020. Evolutionary Genomics of Structural Variation in Asian Rice (*Oryza sativa*) Domestication. *Molecular Biology and Evolution*. doi:10.1093/molbev/msaa185

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**:1571–1572. doi:10.1093/bioinformatics/btr167

Kudo T, Sasaki Y, Terashima S, Matsuda-Imai N, Takano T, Saito M, Kanno M, Ozaki S, Suwabe K, Suzuki G, Watanabe M, Matsuoka M, Takayama S, Yano K. 2016. Identification of reference genes for quantitative expression analysis using large-scale RNA-seq data of *Arabidopsis thaliana* and model crop plants. *Genes and Genetic Systems* **91**:111–125. doi:10.1266/ggs.15-00065

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**:357–359. doi:10.1038/nmeth.1923

Lasky JR, Marais DLD, Lowry DB, Povolotskaya I, McKay JK, Richards JH, Keitt TH, Juenger TE. 2014. Natural variation in abiotic stress responsive gene expression and local adaptation to climate in Arabidopsis thaliana. *Molecular Biology and Evolution* **31**:2283–2296. doi:10.1093/molbev/msu170

Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**:204–220. doi:10.1038/nrg2719

Leinonen PH, Sandring S, Quilot B, Clauss MJ, Thomas MO, Ågren J, Savolainen O. 2009. Local adaptation in European populations of *Arabidopsis lyrata* (Brassicaceae). *American Journal of Botany* **96**:1129–1137. doi:10.3732/ajb.0800080

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint* **arXiv**:1303.3997.

Lisch D. 2009. Epigenetic regulation of transposable elements in plants. *Annu Rev Plant Biol* **60**:43–66. doi:10.1146/annurev.arplant.59.032607.092744

Liu J, He Z. 2020. Small DNA methylation, big player in plant abiotic stress responses and memory. *Frontiers in Plant Science* **11**.

Lloyd JPB, Lister R. 2022. Epigenome plasticity in plants. *Nat Rev Genet* **23**:55–68. doi:10.1038/s41576-021-00407-y

Lopez-Arboleda WA, Reinert S, Nordborg M, Korte A. 2021. Global genetic heterogeneity in adaptive traits. *Molecular Biology and Evolution* **38**:4822–4831. doi:10.1093/molbev/msab208

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:550. doi:10.1186/s13059-014-0550-8

Lowry DB, Logan TL, Santuari L, Hardtke CS, Richards JH, DeRose-Wilson LJ, McKay JK, Sen S, Juenger TE. 2013. Expression quantitative trait locus mapping across water availability environments reveals contrasting associations with genomic features in *Arabidopsis*. *The Plant Cell* **25**:3266–3279. doi:10.1105/tpc.113.115352

Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ. 2019. The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nature Plants* **5**:1250–1259. doi:10.1038/s41477-019-0548-z

Mattila TM, Tyrmi J, Pyhäjärvi T, Savolainen O. 2017. Genome-wide analysis of colonization history and concomitant selection in *Arabidopsis lyrata*. *Molecular Biology and Evolution* **34**:2665–2677. doi:10.5061/dryad.m1v40

McClintock B. 1984. The significance of responses of the genome to challenge. *Science* **226**:792–801. doi:10.1126/science.15739260

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**:1297–303. doi:10.1101/gr.107524.110

Meisner J, Albrechtsen A. 2018. Inferring population structure and admixture proportions in low depth NGS data. *Genetics* **210**:719–731. doi:10.1534/genetics.118.301336

Messer PW, Petrov DA. 2013. Frequent adaptation and the McDonald-Kreitman test. *Proceedings of the National Academy of Sciences* **110**:8615–8620. doi:10.1073/pnas.1220835110

Nakamura T, Yamada KD, Tomii K, Katoh K. 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **34**:2490–2492. doi:10.1093/bioinformatics/bty121

Nowak MD, Birkeland S, Mandáková T, Roy Choudhury R, Guo X, Gustafsson ALS, Gizaw A, Schrøder-Nielsen A, Fracassetti M, Brysting AK, Rieseberg L, Slotte T, Parisod C, Lysak MA, Brochmann C. 2021. The genome of *Draba nivalis* shows signatures of adaptation to the extreme environmental stresses of the Arctic. *Molecular Ecology Resources* **21**:661–676. doi:10.1111/1755-0998.13280

Parmesan C. 2006. Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics* **37**:637–669.

Patra N, Hariharan S, Gain H, Maiti MK, Das A, Banerjee J. 2021. TypiCal but DeliCate Ca++re: dissecting the essence of calcium signaling network as a robust response coordinator of versatile abiotic and biotic stimuli in plants. *Frontiers in Plant Science* **12**.

Pietzenuk B, Markus C, Gaubert H, Bagwan N, Merotto A, Bucher E, Pecinka A. 2016. Recurrent evolution of heat-responsiveness in Brassicaceae *COPIA* elements. *Genome Biology* **17**:209. doi:10.1186/s13059-016-1072-3

Price TD, Qvarnström A, Irwin DE. 2003. The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society of London Series B: Biological Sciences* **270**:1433–1440. doi:10.1098/rspb.2003.2372

Quadrana L, Bortolini Silveira A, Mayhew GF, LeBlanc C, Martienssen RA, Jeddeloh JA, Colot V. 2016. The *Arabidopsis thaliana* mobilome and its impact at the species level. *eLife* **5**:e15716. doi:10.7554/eLife.15716

Quadrana L, Etcheverry M, Gilly A, Caillieux E, Madoui M-A, Guy J, Bortolini Silveira A, Engelen S, Baillet V, Wincker P, Aury J-M, Colot V. 2019. Transposition favors the generation of large effect mutations that may facilitate rapid adaption. *Nat Commun* **10**:3421. doi:10.1038/s41467-019-11385-5

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26**:841–2. doi:10.1093/bioinformatics/btq033

Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES. 2016. Open chromatin reveals the functional maize genome. *Proceedings of the National Academy of Sciences of the United States of America* **113**:3177–3184. doi:10.1073/pnas.1525244113

Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. *Nature Reviews Genetics* **14**:807–820. doi:10.1038/nrg3522

Schliep KP. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**:592–593. doi:10.1093/bioinformatics/btq706

Schmalhausen II. 1949. Factors of evolution: the theory of stabilizing selection. Philadelphia: Blakiston.

Secco D, Wang C, Shou H, Schultz MD, Chiarenza S, Nussaume L, Ecker JR, Whelan J, Lister R. 2015. Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. *eLife* **4**:1–26. doi:10.7554/elife.09343

Smith NGC, Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. *Nature* **415**:1022–1024. doi:10.1038/4151022a

Stitzer MC, Anderson SN, Springer NM, Ross-Ibarra J. 2021. The genomic ecosystem of transposable elements in maize. *PLOS Genetics* **17**:e1009768. doi:10.1371/journal.pgen.1009768

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

Studer A, Zhao Q, Ross-Ibarra J, Doebley J. 2011. Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nature Genetics* **43**:1160–1163. doi:10.1038/ng.942

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO Summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**:e21800. doi:10.1371/journal.pone.0021800

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585–595.

Takou M, Hämälä T, Koch EM, Steige KA, Dittberner H, Yant L, Genete M, Sunyaev S, Castric V, Vekemans X, Savolainen O, Meaux J De. 2021. Maintenance of adaptive dynamics and no detectable load in a range-edge outcrossing plant population. *Molecular Biology and Evolution* **38**:1820–1836. doi:10.1093/molbev/msaa322

Takuno S, Seymour DK, Gaut BS. 2017. The evolutionary dynamics of orthologs that shift in gene body methylation between *Arabidopsis* species. *Molecular Biology and Evolution* **34**:1479–1491. doi:10.1093/molbev/msx099

Talbert PB, Henikoff S. 2014. Environmental responses mediated by histone variants. *Trends Cell Biol* **24**:642–650. doi:10.1016/j.tcb.2014.07.006

Thiebaut F, Hemerly AS, Ferreira PCG. 2019. A role for epigenetic regulation in the adaptation and stress responses of non-model plants. *Frontiers in Plant Science* **10**.

Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR. 2015. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nature Protocols* **10**:475–483. doi:10.1038/nprot.2014.114

Waddington CH. 1942. Canalization of development and the inheritance of acquired characters. *Nature* **150**:563–565.

Weir BS, Cockerham CC. 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**:1358–1370. doi:10.2307/2408641

West-Eberhard MJ. 2003. Developmental plasticity and evolution. Oxford University Press.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH. 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* **8**:973–982.

Wos G, Choudhury RR, Kolář F, Parisod C. 2021. Transcriptional activity of transposable elements along an elevational gradient in Arabidopsis arenosa. *Mobile DNA* **12**:1–12. doi:10.1186/s13100-021-00236-0

Wright S. 1931. Evolution in Mendelian populations. *Genetics* **16**:97–159.

Wyler M, Stritt C, Walser J-C, Baroux C, Roulin AC. 2020. Impact of transposable elements on methylation and gene expression across natural accessions of *Brachypodium distachyon*. *Genome Biology and Evolution* **12**:1994–2001. doi:10.1093/gbe/evaa180

Yeaman S, Hodgins KA, Suren H, Nurkowski KA, Rieseberg LH, Holliday JA, Aitken SN. 2014. Conservation and divergence of gene expression plasticity following c. 140 million years of evolution in lodgepole pine (*Pinus contorta*) and interior spruce (*Picea glauca* × *Picea engelmannii*). *New Phytologist* **203**:578–591. doi:10.1111/nph.12819

Zhang Y, Ngu DW, Carvalho D, Liang Z, Qiu Y, Roston RL, Schnable JC. 2017. Differentially regulated orthologs in sorghum and the subgenomes of maize. *The Plant Cell* **29**:1938–1951. doi:10.1105/tpc.17.00354