

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

Tuomas Härmälä<sup>\*,†</sup>, Weixuan Ning<sup>§</sup>, Helmi Kuittinen, Nader Aryamanesh<sup>†,‡</sup>, Outi Savolainen<sup>†</sup>

Department of Ecology and Genetics, University of Oulu, Oulu, Finland

**\*For correspondence:**  
tuomas.hamala@gmail.com

<sup>†</sup>These authors contributed equally to this work

**Present address:** <sup>‡</sup>School of Life Sciences, University of Nottingham, Nottingham, United Kingdom; <sup>§</sup>School of Natural Sciences, Massey University, Palmerston North, New Zealand; <sup>‡</sup>Embryology Research Unit, Bioinformatics Group, Children's Medical Research Institute, University of Sydney, Westmead, Australia

**Competing interest:** The authors declare that no competing interests exist.

**Funding:** See page 18

**Received:** 31 August 2022  
**Accepted:** 12 October 2022

**Reviewing Editor:** Jeffrey Ross-Ibarra, University of California, Davis, United States

© Copyright Härmälä et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

**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 environmentally 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. These 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 the maintenance of genetically variable environmental 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 an excess of recent TE insertions, and we observe that specific TE families are enriched within environmentally responsive genes. Our findings provide insight into selective forces shaping plastic and genetic variation. We also highlight how plastic responses at TEs can rapidly create novel heritable variation in stressful conditions.

## Editor's evaluation

This work provides a thorough look at changes in expression, methylation, and nucleotide and transposable element diversity among three populations of *Arabidopsis lyrata* in two different environments. It is a rich dataset, and the authors present a number of nice findings with relevance for our understanding of local adaptation and the process of – and potential constraints to – adaptation to rapid climate change.

## Introduction

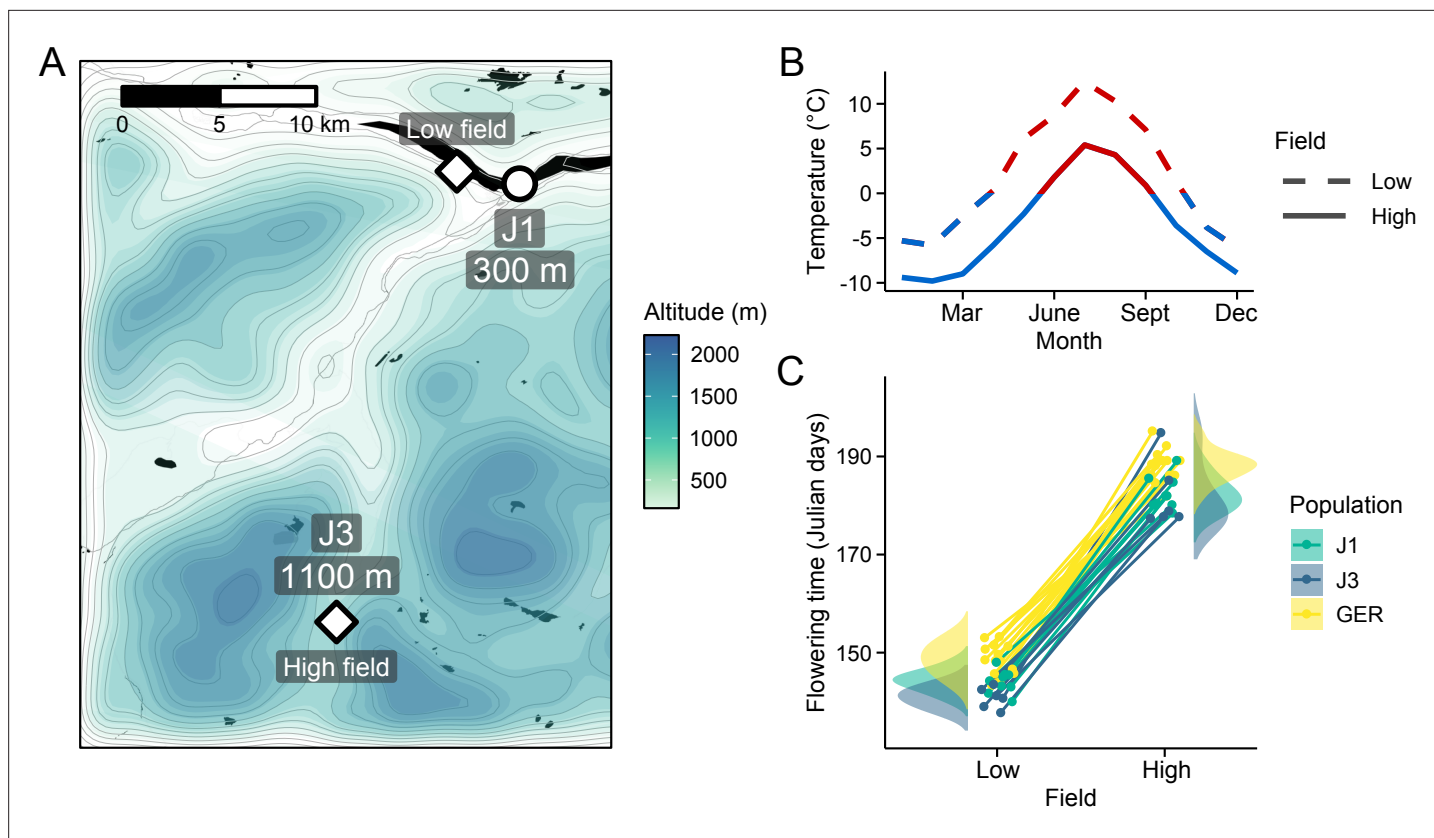
To maintain viability in a changing environment, populations need to track shifting fitness optima through genetic adaptation or phenotypic plasticity (Alberto et al., 2013; Ghalambor et al., 2007; Parmesan, 2006). Although environmentally induced variation within a generation can efficiently facilitate 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 pressure.

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; Wos et al., 2021a; 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 consistently 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; Pietzenek et al., 2016; Wos et al., 2021b**). 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 superfamilies, 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 (**Van’t 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 may also be transmitted from parent to offspring (**Law and Jacobsen, 2010; Lloyd and Lister, 2022**), although most stably inherited methylation variants have a genetic (and not plastic) basis (**Kawakatsu et al., 2016; Lloyd and Lister, 2022; Muyle et al., 2022**). DNA methylation is further involved in epigenetic silencing of TEs (**Law and Jacobsen, 2010**), and thus environmentally induced changes in



**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 Shuttle Radar Topography Mission (SRTM) (Farr and Kobrick, 2000). (B) Average monthly temperature at areas around the field sites (red line:  $\geq 0^{\circ}\text{C}$ , blue line:  $< 0^{\circ}\text{C}$ ). Data from MET Norway. (C) Average flowering time of full-sib families grown at the field sites. Data from Hämälä et al., 2018.

The online version of this article includes the following figure supplement(s) for figure 1:

**Figure supplement 1.** Locations of the *A. lyrata* populations.

**Figure supplement 2.** Photos of the low- and high-altitude field sites, taken during June 6 and 7, 2015, respectively.

the DNA methylome could either promote or suppress TE activity. Therefore, characterizing patterns of DNA methylation in naturally contrasting conditions may help us to better understand environmental effects in 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? Is differential methylation associated with 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 (Figure 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; Figure 1—figure supplement 1). Although the field sites were closely situated (Figure 1A), the altitudinal difference resulted in considerable environmental differences between the fields (Figure 1B and Figure 1—figure supplement 2), thus leading to distinct responses in ecologically important traits (Figure 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 (WGBS) to examine variation in gene expression and DNA methylation. Leaf samples from the plants were collected 1 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  $\times$  genetic (difference between fields and populations).

## 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 differentiated from GER along the second PC axis (Figure 2A). By contrast, DNA methylation was mainly influenced by population structure, with first two PCs differentiating the three populations (Figure 2A). Separating DNA methylation into three sequence contexts, CG, CHG, and CHH (where H is A, T, or C), revealed that variation in CG methylation was congruent with the total methylation data (Figure 2—figure supplement 1). Methylation in the CHG and CHH contexts was primarily affected by differences between the Norwegian populations and GER, but the second PC in both PCAs revealed a slight influence of the field site, particularly in the Norwegian populations (Figure 2—figure supplement 1).

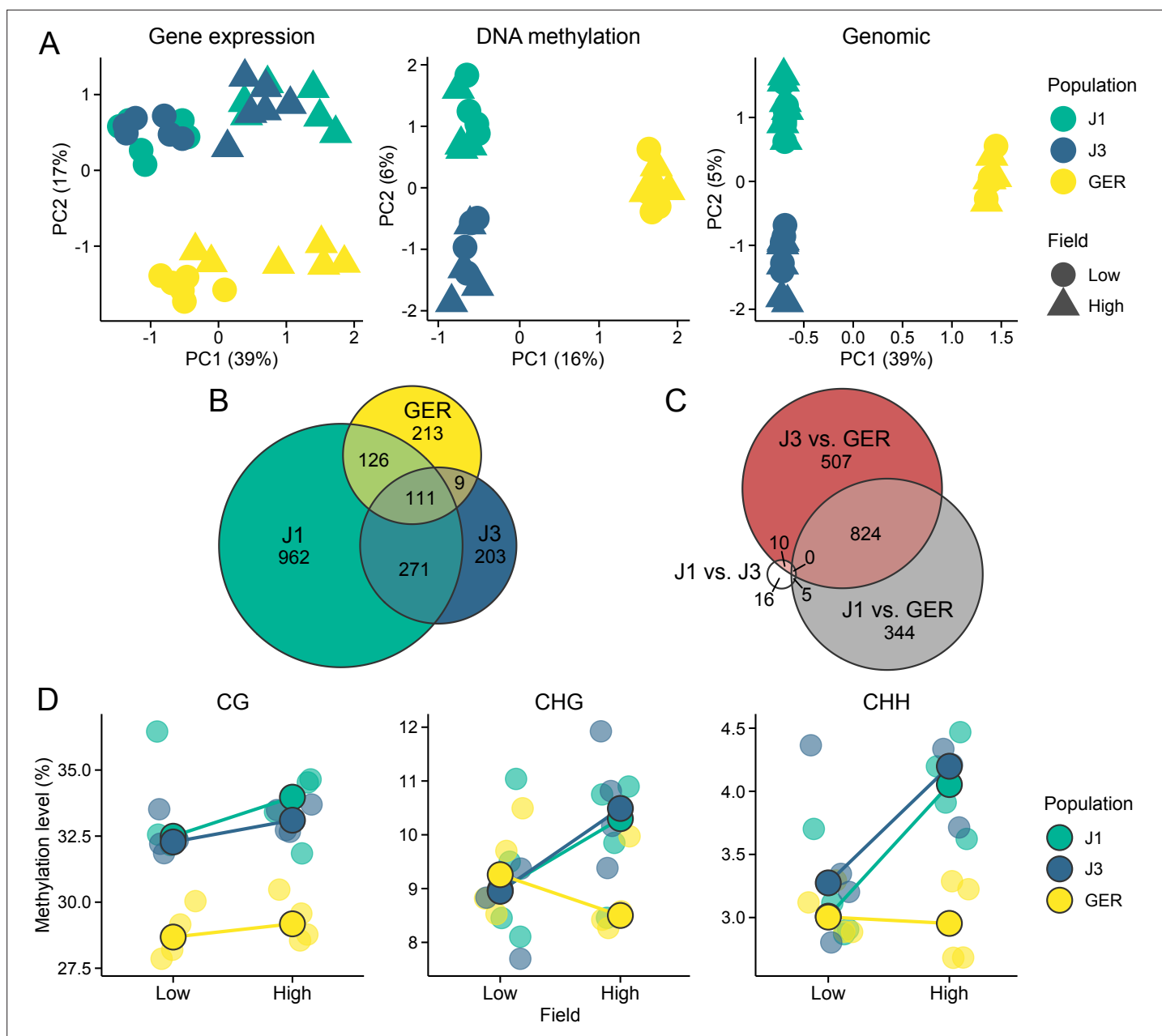
All three populations contained genes that were differentially expressed between the fields (Figure 2B), but the largest number of DEGs (differentially expressed genes) was found in J1. Expression differences between populations were mainly due to DEGs between the Norwegian populations and GER (Figure 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 (Figure 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 (Figure 2D). CG methylation levels showed only a subtle difference between the field sites but a clear difference between the populations (Figure 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).

## Plastic and genetic responses at gene expression and DNA 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  $\sim$  field), population (DEG  $\sim$  population), and their interaction (DEG  $\sim$  field:population). We also used logistic regression and LRTs to conduct a similar analysis among differentially methylated genes (DMGs). As genic methylation can have distinct effects on gene expression depending on the sequence context (Muyle *et al.*, 2022), we searched for DMGs using CG and non-CG methylation separately.

We detected between 112 and 3456 genes that had their expression or methylation levels affected by the field site and/or population history (Table 1). Consistent with the PCA results, most DEGs were found between the field sites (DEG  $\sim$  field), and most DMGs were found between populations (DMG  $\sim$  population). Despite the relative lack of methylation plasticity, differential methylation between the field sites was more common at non-CG than CG sites (Table 1). Field  $\times$  population interactions (DEG  $\sim$  field:population) were rare in gene expression, and using the same criteria for outlier detection as with DEG  $\sim$  field and DEG  $\sim$  population, only 28 genes passed the DEG threshold. As this small number did not allow us to examine selective signals at the field  $\times$  population genes, we used a more lenient threshold for multiple correction (see Materials and methods), while acknowledging the potentially higher false-positive rate among these genes. Of the 21,969 genes expressed in our experimental plants, we defined 3933 (18%) as environmentally responsive. Although methylation



**Figure 2.** Gene expression and methylation variation across the field sites and populations. **(A)** Gene expression, DNA methylation, and genomic (based on SNPs called from the transcriptome data) variation along the first two eigenvectors of a principal components analysis (PCA). The proportion of variance explained by the principal components (PCs) is shown in parentheses. **(B)** The number of differentially expressed genes (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.

The online version of this article includes the following figure supplement(s) for figure 2:

**Figure supplement 1.** Methylation variation along the first two eigenvectors of a principal components analysis (PCA), shown for the three methylation contexts.

**Figure supplement 2.** Overlap between candidate gene groups.

**Figure supplement 3.** The  $\log_2$  OR of association between differentially expressed genes (DEGs) and differentially methylated regions 1 kb upstream of each gene, shown for each methylation context.

**Figure supplement 4.** Top five enriched ( $Q < 0.05$ , hypergeometric test) gene ontology (GO) terms among each candidate gene set.



**Table 1.** Models for likelihood-ratio tests (LRTs) and the number of identified differentially expressed genes (DEGs) and differentially methylated genes (DMGs).

Gene set	Full and reduced model	Number
DEG ~ field	Expression ~ population + field Expression ~ population	3456
DEG ~ population	Expression ~ field + population Expression ~ field	1476
DEG ~ field:population	Expression ~ field + population + field:population Expression ~ field + population	477
DMG-CG ~ field	CG methylation ~ CH% + population + field CG methylation ~ CH% + population	112
DMG-CG ~ population	CG methylation ~ CH% + field + population CG methylation ~ CH% + field	641
DMG-CG ~ field:population	CG methylation ~ CH% + field + population + field:population CG methylation ~ CH% + field + population	260
DMG-CH ~ field	CH methylation ~ population + field CH methylation ~ population	1036
DMG-CH ~ population	CH methylation ~ field + population CH methylation ~ field	3580
DMG-CH ~ field:population	CH methylation ~ field + population + field:population CH methylation ~ field + population	680

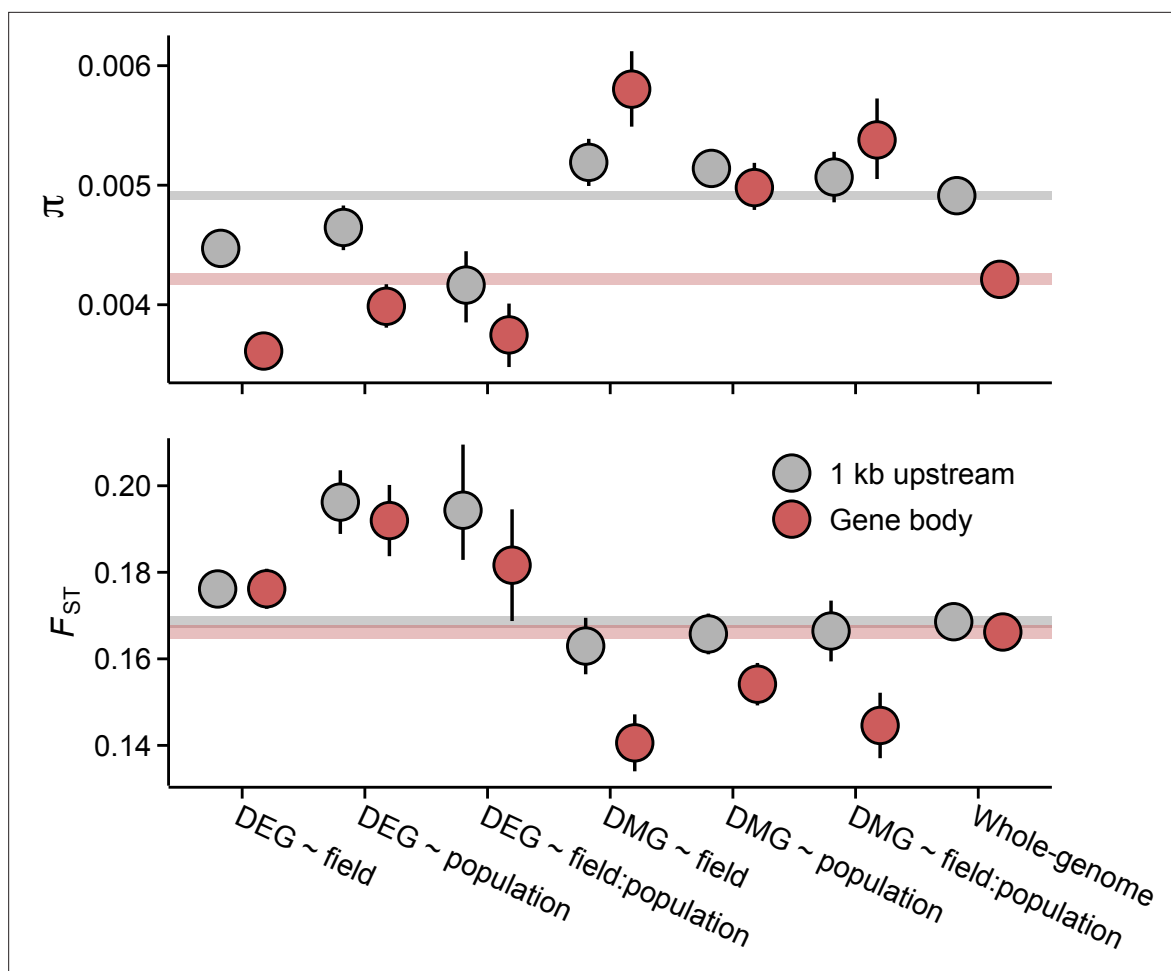
CH = CHG and CHH.  
CH% = methylation rate at CHG and CHH contexts.

levels were correlated with gene expression levels (Spearman’s  $\rho_{CG} = 0.29$ ,  $\rho_{CHG} = -0.23$ ,  $\rho_{CHH} = -0.21$ ;  $p < 2 \times 10^{-16}$ ), DEG ~ field shared fewer than expected genes with the DMG groups (**Figure 2—figure supplement 2**). Besides methylation of gene bodies, 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 expectations. DEG ~ population, by contrast, was associated with differential methylation at both gene bodies and promoter regions (**Figure 2—figure supplement 3**).

Based on gene ontology (GO) terms, most DEG and DMG sets were enriched for genes involved in specific biological processes (**Figure 2—figure supplement 4**). 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 (‘defense response to bacterium and virus’). DEG ~ field:population showed enrichment for multiple different processes, including ‘response to cold.’ Although we found fewer enriched GO terms among the DMGs, terms involved in cellular signaling and transport (e.g. ‘signal transduction’ and ‘cation transport’) were strongly represented among the non-CG DMGs (**Figure 2—figure supplement 4**).

Footprints of selection at candidate gene sets

To examine selective signals at DEGs and DMGs, we used whole-genome sequence data from independent J1, J3, and GER individuals (**Hämälä et al., 2018; Mattila et al., 2017; Takou et al., 2021**). As environmental effects in CG methylation were rare, we primarily characterize selection at DEGs and non-CG DMGs. 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 (**Figure 3** and **Figure 3—figure supplement 1**),



**Figure 3.** Pairwise nucleotide diversity ( $\pi$ ) and  $F_{ST}$  at the candidate gene sets.  $\pi$  estimates are shown for J3 ( $n=22$ ) population. Estimates for J1 ( $n=9$ ) and GER ( $n=17$ ) are shown in **Figure 3—figure supplements 1 and 3**.  $F_{ST}$  was estimated across the three populations. See **Figure 3—figure supplement 3** for pairwise  $F_{ST}$  estimates. Error bars show 95% bootstrap-based CIs. Shaded areas mark the 95% CIs across all genes.

The online version of this article includes the following figure supplement(s) for figure 3:

**Figure supplement 1.** Pairwise nucleotide diversity ( $\pi$ ) at the candidate gene sets.

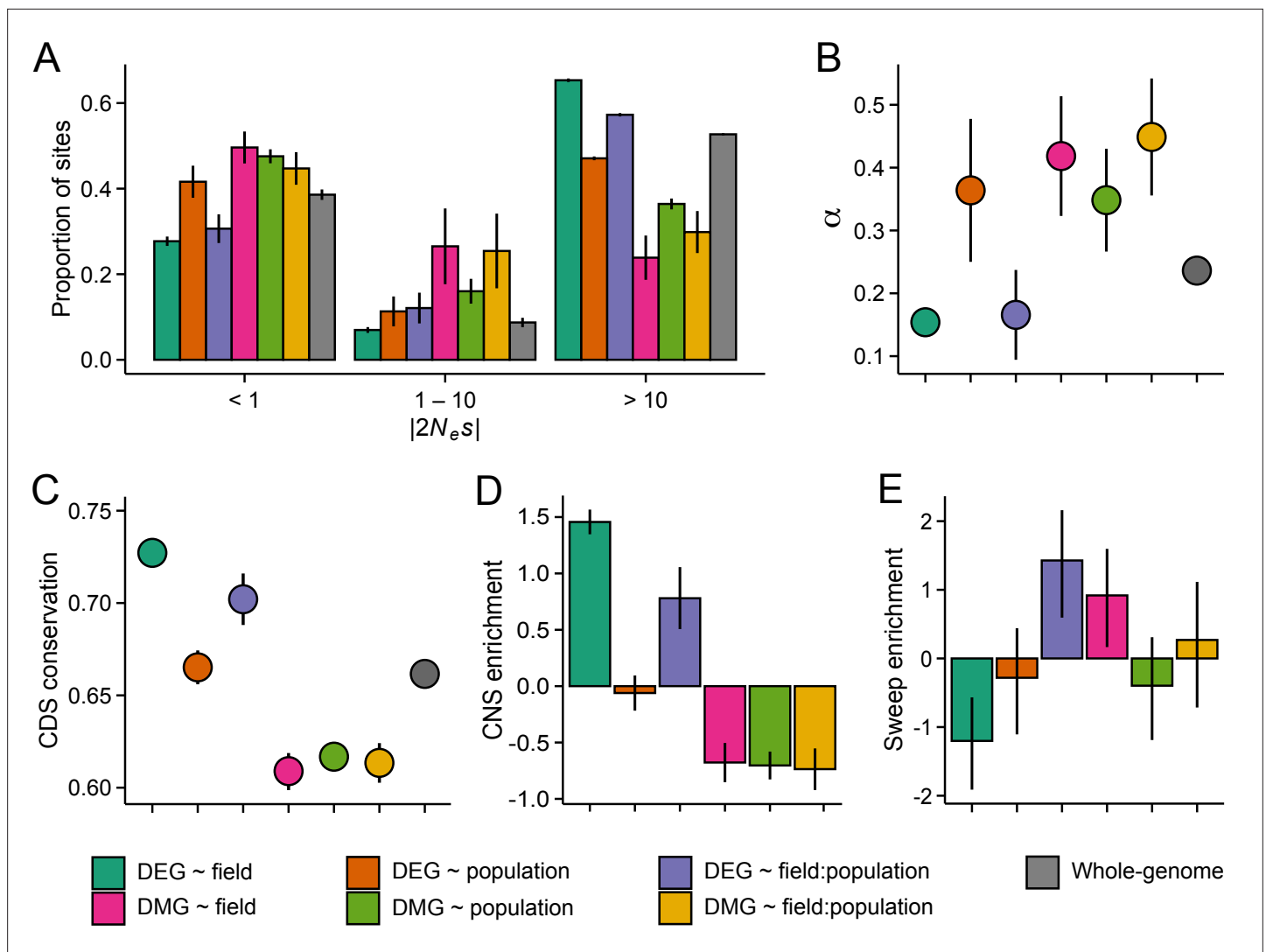
**Figure supplement 2.** Population effects at differentially expressed gene (DEG) ~ field:population genes.

**Figure supplement 3.** Pairwise  $F_{ST}$  estimates for the candidate gene sets.

**Figure supplement 4.** Pairwise nucleotide diversity ( $\pi$ ) at CG and non-CG differentially methylated genes (DMGs).

indicative of purifying selection or selective sweeps. In particular, the promoter regions at DEG ~ field:population genes exhibited a combination of low genetic diversity and high differentiation, suggesting that parts of the transcriptional machinery, such as cis-regulatory elements, may have been subject to differential selection. The signals of expression and sequence differentiation were, however, almost exclusively found between the Norwegian populations and GER (**Figure 3—figure supplements 2 and 3**). In contrast to DEGs, non-CG DMGs harbored higher than average levels of genetic diversity (diversity levels at CG DMGs were lower and deviated less from the whole-genome average; **Figure 3—figure supplement 4; Figure 3**), potentially reflecting relaxed selective constraint. In the rest of the paper, the abbreviation DMG refers to non-CG DMGs.

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 coding sequence of DEGs and DMGs by modelling the distribution of fitness effects (DFEs) of new nonsynonymous mutations (**Kim et al., 2017**) and by estimating the rate of adaptive evolution ( $\alpha$ ) (**Messer and Petrov, 2013**). In all three populations, genes belonging



**Figure 4.** The efficacy of negative and positive selection at the candidate gene sets. **(A)** The distribution of fitness effects of new nonsynonymous variants. The mutations were divided into three bins based on the strength of purifying selection ( $2N_e s$ ): nearly neutral, intermediate, and deleterious, respectively. **(B)** The proportion of sites fixed by positive selection ( $\alpha$ ). **(C)** The conservation of coding sequence (CDS). Shown are average GERP scores estimated for each gene group. The scores were rescaled from 0 to 1 using the range of possible values at each site. **(D)** The enrichment of conserved noncoding sequence (CNS) 1 kb upstream of candidate genes. Shown are the  $\log_2$  OR of association between CNSs and the candidate gene sets. **(E)** The enrichment of selective sweeps at the candidate genes. Shown are the  $\log_2$  ORs of association between selective sweeps and the candidate gene sets. Panels A, B, and E show results for the J3 population ( $n=22$ ). See **Figures 1 and 2** for results on J1 ( $n=9$ ) and GER ( $n=17$ ), respectively. For all panels, error bars show 95% bootstrap-based CIs.

The online version of this article includes the following figure supplement(s) for figure 4:

**Figure supplement 1.** The efficacy of negative and positive selection at candidate gene sets in the J1 population ( $n=9$ ).

**Figure supplement 2.** The efficacy of negative and positive selection at candidate gene sets in the GER population ( $n=17$ ).

to the DEG ~ field and DEG ~ field:population groups appeared under strong purifying selection: fewer than expected mutations were predicted to be nearly neutral ( $2N_e s < 1$ ) and more predicted to be deleterious ( $2N_e s > 10$ ) (**Figure 4A** and **Figure 4—figure supplements 1A and 2A**;  $p=0.004$ ). 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,  $\alpha$  estimates indicated that fewer than expected nonsynonymous mutations have been fixed by positive selection in the DEG ~ field and DEG ~ field:population groups (**Figure 4B**, **Figure 4—figure supplements 1B and 2B**;  $p<0.004$ , LRT). In contrast to the plastic DEG groups, the DFE and  $\alpha$  estimates indicated relaxed evolutionary constraint at other



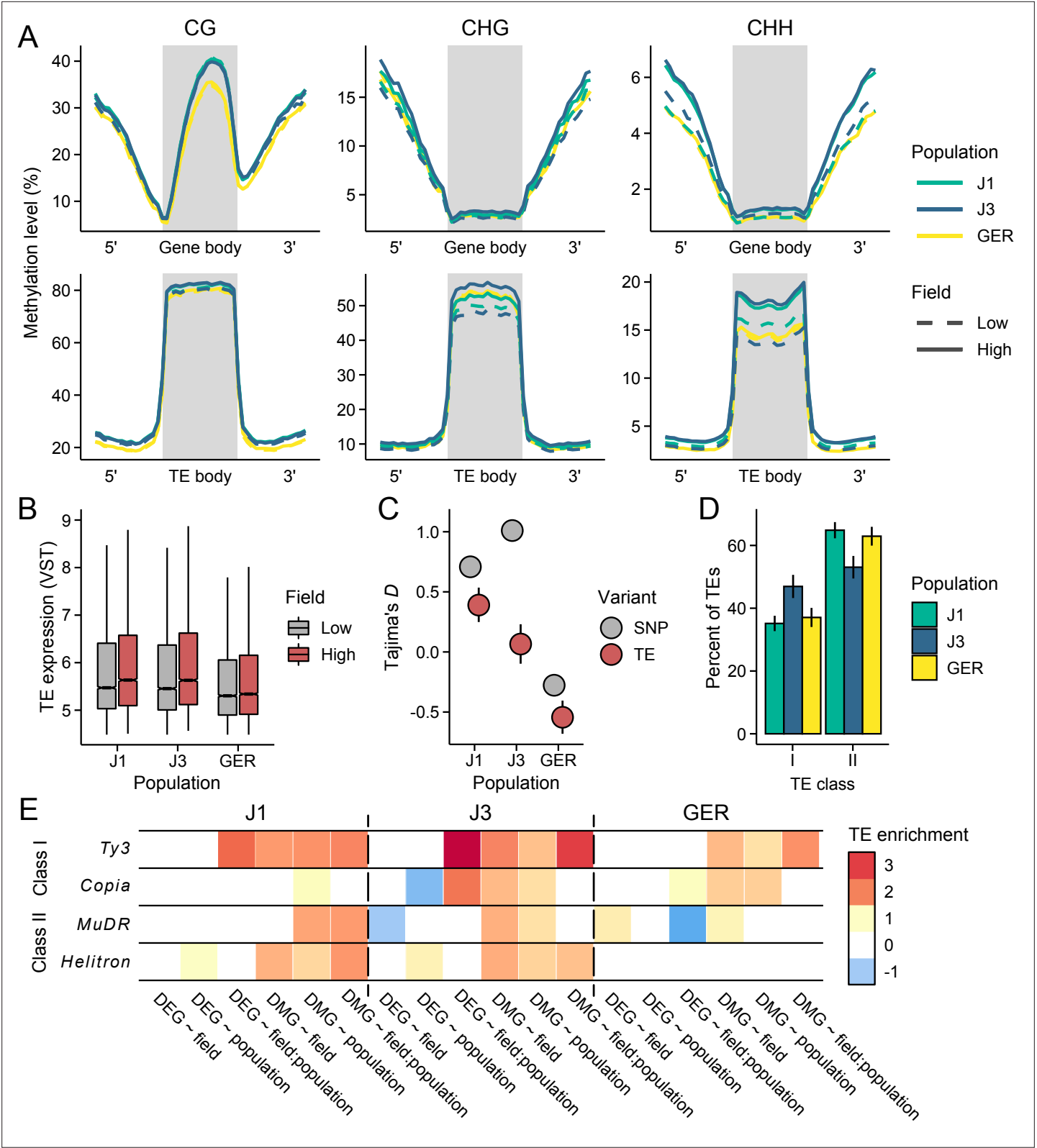
candidate gene groups. In particular, the DMGs had higher than expected  $\alpha$  in all three populations (**Figure 4B**, **Figure 4—figure supplements 1B and 2B**;  $p < 0.0002$ , LRT), suggesting long-term adaptive evolution. Together, these results are indicative of reduced evolutionary rates at the plastic DEG groups as well as rapid molecular evolution at the DMGs. Both results were further supported by an analysis of coding sequence conservation among 26 eudicot species, which identified higher than expected conservation at the two DEG groups and lower than expected conservation at the DMGs (**Figure 4C**;  $p < 0.0005$ , Wilcoxon rank-sum test). Furthermore, by searching for conserved noncoding sequences (CNSs) 1 kb upstream of each gene (**Haudry et al., 2013**), we found that CNSs were over-represented at the DEG ~ field and DEG ~ field:population genes and underrepresented at the DMGs (**Figure 4D**), indicating that the promotor regions are similarly (un)conserved.

To explore how the differing evolutionary rates influence the short-term adaptive potential of these gene groups, we used RAI<sub>SD</sub> (**Alachiotis and Pavlidis, 2018**) to search for footprints of recent positive selection at each gene (gene body  $\pm 2$  kb). We found that DEG ~ field exhibited a clear deficit of selective sweeps (**Figure 4E** and **Figure 4—figure supplements 1C and 2C**;  $p < 0.006$ , 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 (**Figure 4E** and **Figure 4—figure supplements 1C and 2C**;  $p < 0.002$ , Fisher's exact test). Between 10 and 16 genes overlapped the sweep regions in the three populations, with 8 genes shared between the populations. At each of these genes, windows producing the strongest selection signals were found upstream of the transcription start site (TSS), suggesting that the promotor regions may have been subject to recent positive selection.

## Environmental response at TEs

Besides genes, stressful environments may invoke plastic responses at TEs (**Capy et al., 2000**). To examine this in our data, we quantified TE methylation and TE expression. In contrast to genes, we found that methylation levels across TEs were clearly influenced by the field sites (**Figure 5A**). 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 (**Figure 5A** and **Figure 5—figure supplements 1–3**; logistic regression  $\beta_{CG} = 0.07$ ,  $\beta_{CHG} = 0.12$ , and  $\beta_{CHH} = 0.17$ ;  $p < 2 \times 10^{-16}$ , LRT). Population effects were also evident in TE methylation: J1 and J3 exhibited clearer environmental responses compared to GER (**Figure 5A** and **Figure 5—figure supplements 1–3**). 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  $\rho_{CG} = -0.28$ ,  $\rho_{CHG} = -0.16$ , and  $\rho_{CHH} = -0.14$ ;  $p < 6 \times 10^{-10}$ ). 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 (**Figure 5B** and **Figure 5—figure supplement 4**;  $p < 2 \times 10^{-16}$ , Wilcoxon rank-sum test). By comparing the TE results to randomly compiled gene sets of equal size, we found that the observed expression differences are unlikely caused by technical bias (**Figure 5—figure supplement 5**). Therefore, rather than DNA methylation proactively suppressing TEs in the more stressful environment, these patterns suggest 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 (**Hollister and Gaut, 2009**; **Horvath and Slotte, 2017**; **Wylers 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 (**Figure 5—figure supplement 6**). 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 whether such activation can shape 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



**Figure 5.** 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. See **Figure 5—figure supplements 1–3** for results on different TE superfamilies **(B)** The expression of TE families at low- and high-altitude field sites. See **Figure 5—figure supplement 4** for results on different TE superfamilies. **(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). **(E)** TE enrichment across different populations and field sites. **Figure 5 continued on next page**

Figure 5 continued

II) in each population. See **Figure 5—figure supplement 8** for results on different superfamilies. **(E)** Enrichment of TEs at gene bodies of differentially expressed genes (DEGs) and non-CG differentially methylated genes (DMGs). Shown are the  $\log_2$  ORs of association between the four largest TE superfamilies and the candidate gene sets ( $Q < 0.05$ ). See **Figure 5—figure supplement 10** for results on all TEs, including up- and downstream regions of genes. For C and D panels, error bars show 95% bootstrap-based CIs.

The online version of this article includes the following figure supplement(s) for figure 5:

**Figure supplement 1.** CG methylation levels of different transposable element (TE) superfamilies at low- and high-altitude field sites.

**Figure supplement 2.** CHG methylation levels of different transposable element (TE) superfamilies at low- and high-altitude field sites.

**Figure supplement 3.** CHH methylation levels of different transposable element (TE) superfamilies at low- and high-altitude field sites.

**Figure supplement 4.** Expression levels of different transposable element (TE) superfamilies at low- and high-altitude field sites.

**Figure supplement 5.** Observed transposable element (TE) expression difference between the field sites compared to 10,000 randomly compiled gene sets of equal size.

**Figure supplement 6.** The association between gene expression and distance from the closest transposal element (TE; upstream of transcription start site [TSS]).

**Figure supplement 7.** Efficacy of selection in the three populations.

**Figure supplement 8.** The proportion of different transposal element (TE) superfamilies in each population.

**Figure supplement 9.** The count of different transposal element (TE) superfamilies in each population.

**Figure supplement 10.** The  $\log_2$  OR of association between transposal elements (TEs) and the candidate gene sets.

(**Figure 5C**). 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 efficacy of selection in J3 appears equal (or weaker) to J1 (**Figure 5C** and **Figure 5—figure supplement 7**), pointing toward an alternative explanation: a recent influx 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; as retrotransposons move by ‘copy-and-paste’ mechanism and DNA transposons move by ‘cut-and-paste’ mechanism (**Wicker et al., 2007**), the number of new retroelements likely increases more rapidly as a result of such activation. As the probability of TEs impairing gene function is increased along with their activation, the number of active DNA transposons could also be more effectively reduced by purifying selection. Consistent with these expectations, we found an enrichment of LTR retrotransposons and a deficit of DNA transposons in J3 compared to J1 (**Figure 5D** and **Figure 5—figure supplements 8 and 9**). Therefore, our results suggest that stress-induced TE activation could have resulted in recent proliferation of retrotransposons (and potentially purging of DNA transposons) in the alpine population, although we cannot rule out the role of drift in shaping the composition of the TE landscape. Finally, by examining the locations of the TE insertions, we discovered that many genes belonging to the six candidate groups were more likely to harbor TE insertions than the average gene (**Figure 5E** and **Figure 5—figure supplement 10**). In particular, the DEG ~ field:population group was highly enriched for Ty3 retrotransposons in the J3 population ( $p = 1.3 \times 10^{-6}$ , Fisher’s exact test). We also found that DMGs were TE insertion hotspots (**Figure 5E**), suggesting that many of the non-CG methylation changes may have arisen as a result of TE methylation spreading to genic regions.

## 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 stressors 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*), demonstrating that some ecologically important traits are genetically differentiated despite the strong plasticity (e.g. flowering time, *Figure 1C*). In fact, if such differentiation is primarily manifested at earlier life history stages, our sampling after 1 year likely missed those effects. 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*). Given that GER plants fared poorly at both field sites during our multi-year experiment (*Hämälä et al., 2018*), we may assume that much of the expression differentiation in GER is maladaptive in the Norwegian environment. Nevertheless, as such differentiation can arise from evolved effects in the local populations (i.e. for some genes GER exhibits the ancestral expression), comparing the Norwegian populations to GER likely revealed adaptive expression responses. Such differences may have further 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*), salt stress (*Wibowo et al., 2016*), 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, as histone modifications are known to exert a major influence on gene expression (*Lloyd and Lister, 2022*), future work studying environmental induction 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 DNA methylation, when they accumulate over generations, may influence patterns of gene expression (*Muyle et al., 2022; Takuno et al., 2017*).

### Environmentally responsive genes evolve under strong selective constraint

We found that environmentally induced expression responses were highly consistent across our experimental plants, indicating that the underlying 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 upstream of the TSS, we further found that population-specific selection at promoter regions (potentially at cis-regulatory variants) has likely contributed to the maintenance of heritable expression variation at the DEG ~ field:population genes. We note, however, that our experimental design did not allow us to associate selective signals with trans-acting variants, which may have also played a role in the observed expression differentiation (*Hämälä et al., 2020; Lopez-Arboleda et al., 2021*). Overall, 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. However, the expression differentiation was almost exclusively found between the Norwegian

populations and GER, indicating that such rewiring of the regulatory network may primarily happen over long evolutionary timescales.

### 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 TE activation. Compared to the lowland population, plants native to the alpine environment harbored an excess of rare TE insertions, suggesting that certain TE families, particularly LTR retrotransposons, may have recently expanded their copy numbers. On average, such TE mobilization 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 (**Hollister and Gaut, 2009; Horvath and Slotte, 2017; Wyler et al., 2020**). Indeed, we also discovered that genes whose promoters lie close to the 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 LTR 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 seen in results by **Quadrana et al., 2019**, who found experimental evidence that retrotransposons of the *Copia* superfamily 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. Indeed, by searching the *A. lyrata* genome for H2A.Z binding sites defined in *A. thaliana* (**Chow et al., 2019**), we confirmed that the DEG ~ field:population genes were more likely to carry such sites than the average gene (OR = 3.10;  $p < 2 \times 10^{-16}$ , Fisher's exact test). Interestingly, the DEG ~ field genes were also enriched for H2A.Z binding sites (OR = 2.93;  $p < 2 \times 10^{-16}$ , Fisher's exact test), suggesting that these genes may be similarly targeted by TE insertions, but due to stronger selective constraint, TEs are purged more efficiently. However, although we also detected an enrichment of *Copia* elements in the environmentally responsive genes, the enrichment of *Ty3* elements was far greater (OR: 4.5 vs. 11.7). This pattern indicates that the *Ty3* superfamily 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 *Ty3* TEs in the alpine species *Arabis alpina* and *Draba nivalis* (**Nowak et al., 2021**).

### Ideas and speculation

By studying gene expression and DNA methylation in natural conditions, we have gained new insights into evolutionary forces 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. We found that the coding and promoter regions of most environmentally responsive genes are slowly evolving, which could influence the adaptive potential of these *A. lyrata* populations by keeping the pool of segregating variants small. In particular, as adaptation to rapidly shifting fitness optima can be limited by the lack of standing genetic variation (**Matuszewski et al., 2015**), the strong selective constraint and low genetic diversity at environmentally responsive genes could pose a risk in cases where environmental perturbations exceed the buffering mechanism provided by phenotypic plasticity (**Ghalambor et al., 2007; Price et al., 2003**). However, one caveat with this interpretation is that mutations in large-effect trans-loci could lead to wide-ranging expression effects despite the conservation of the cis-regulatory elements (**Josephs et al., 2020; Lopez-Arboleda et al., 2021**). Furthermore, if the genetic architecture of adaptation is much more polygenic than detected here, the selective constraint would likely offer less resistance to evolutionary responses (**Hayward and Sella, 2022; Stetter et al., 2018**). Although epigenetic modifications have been proposed as a mechanism of coping with rapid environmental



change (McGuigan *et al.*, 2021), we found little evidence that environmentally induced changes in the DNA methylome could influence adaptive phenotypes in *A. lyrata*. Our results further suggest that novel heritable variation may be rapidly created by TEs integrating into the gene bodies of environmentally responsive genes, but further functional validation would be required to determine the phenotypic and fitness effects of such variants.

## Materials and methods

### Reciprocal transplant experiment and sample collection

We conducted a reciprocal transplant experiment to study altitude adaptation among Norwegian populations of *A. lyrata* ssp. *petraea*. *A. lyrata* is a predominantly outcrossing, perennial herb with a wide circumpolar distribution across the northern hemisphere (Jalas and Suominen, 1994). It is closely related to the model species *A. thaliana*, but differences in key life history traits (*A. thaliana* is a highly selfing annual) make *A. lyrata* a useful study system in ecology and evolution (Savolainen and Kuittinen, 2011). 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 GER (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–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 was germinated and pre-grown in a greenhouse at University of Oulu, Finland for about 2 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 (Figure 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 1 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 10 seed families (due to high mortality, many GER seed families were not available in both fields). Leaf samples were collected during two consecutive days (August 21 and 22, 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 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.

The transcriptome sequencing yielded a total of 1.036 billion read pairs, with an average of 29 million read pairs per individual (Supplementary file 1). 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 to count reads mapping uniquely to each gene model. To quantify TE expression, we first identified TEs using RepeatMasker ([www.repeatmasker.org](http://www.repeatmasker.org)) and a library of *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. Following Anderson *et al.*, 2019, we characterized the expression of different TE



families using both uniquely and nonuniquely mapped reads. To do so, we used mmquant (Zytnicki, 2017) to count reads mapping to individual TEs and compiled count data for each TE family. We only included reads that mapped uniquely to a single element or to multiple elements belonging to the same TE family (we allowed up to 665 multi-mapping locations in STAR, which was the maximum number of TEs from a single family). Reads that mapped to multiple TE families, as well as reads mapping to genes, were excluded. The TE expression data were then normalized by applying variance stabilization transformation (Love et al., 2014) to combined TE and gene counts.

## Whole-genome bisulfite sequencing

To examine patterns of DNA methylation, we used 24 samples (four individuals per population from both fields) for 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.

The WGBS yielded a total of 768 million read pairs, with an average of 32 million read pairs per individual (Supplementary file 1). Low-quality reads and sequencing adapters were removed with Trimmomatic (Bolger et al., 2014), and the surviving reads were 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 (Supplementary file 2).

## Differential expression analysis

After confirming that sequencing batches are unlikely to bias the detection of DEGs (see Appendix 1 for details), we searched for DEGs between the field sites and populations using DESeq2 (Love et al., 2014). To identify population specific DEGs, we fit a single dispersion parameter to each gene with median read counts >1 and used the DESeq2's contrast function and Wald test to detect differences between the field sites. 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 (Table 1). To focus on top-ranking DEGs, we took a relatively stringent approach to account for multiple testing and required that outliers had 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) below 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 methylation at gene bodies. Genic methylation can be separated into two classes: CG methylation, which tends to be associated with increased gene expression levels (outside transcription start and stop sites), and non-CG methylation (i.e. CHG and CHH), which tends to be associated with decreased gene expression levels (Muyle et al., 2022). We therefore searched for DMGs using data from the two methylation classes separately. Focusing on

cytosines with <50% missing data, we calculated the methylated ( $2 \times$  the sum of methylation proportions) and unmethylated ( $2 \times$  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. To find DMGs associated with CG methylation, we included non-CG methylation proportions as a cofactor in the models (**Table 1**). The LRT-based p values were transformed to Q values (**Storey, 2002**) to account for multiple testing. To define DMGs, we required that the genes had  $\geq 10$  cytosines and  $Q < 0.05$ . Given that CG methylation was strongly influenced by population structure (**Figure 2A**), we corrected p values from the field + population to field LRTs (i.e. CG DMGs due to population) using genomic inflation factor (**Devlin and Roeder, 1999**).

## GO enrichment analysis

We conducted a 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 DNA 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  $\geq 30$ , and base quality  $\geq 20$ . 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  $E[G] = \sum_{g=0}^2 gP(G=g)$ , 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.

## 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 ( $\pi$ ) and  $F_{ST}$  for each candidate gene set.  $\pi$  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  $\pi$  and  $F_{ST}$  for each candidate gene group, defined CIs for the estimates by resampling with replacement across genes 1000 times, and compared the estimates against  $\pi$  and  $F_{ST}$  calculated across all genes.

## Distribution of fitness effects

We examined the strength of selective constraint on the candidate gene sets by modelling the DFE. To do so, we estimated folded site frequency spectra (SFS) for synonymous (fourfold) and nonsynonymous (zerofold) sites using ANGSD (**Korneliussen et al., 2014**). We then used  $\partial\text{a}\partial\text{i}$  (**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  $\text{fit}\partial\text{a}\partial\text{i}$  (**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  $\theta$  at nonsynonymous sites. This approach, as opposed to optimizing  $\theta$  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 followed a multi-nomial 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 ( $2N_e s < 1$ ), intermediate ( $1 \leq 2N_e s \leq 10$ ), and deleterious ( $2N_e s > 10$ ) 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 ( $\alpha$ ) (Smith and Eyre-Walker, 2002). Because  $\alpha$  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* – *A. 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  $\alpha$  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 fixed 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 ( $a$ ,  $b$ , and  $c$ ) 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 (Supplementary file 3), 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  $< 1 \times 10^{-5}$  for each gene. For sets of homologs with  $\geq 13$  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.

As our RS scores were estimated only for the coding sequence, we further used publicly available data on CNSs (based on the comparison of nine Brassicaceae species, Haudry et al., 2013) to assess nucleotide conservation at the promoter regions. To do so, we searched for the presence of CNSs 1 kb upstream of each gene and tested whether they are over- or underrepresented at our candidate gene groups using the Fisher’s exact test.

## 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 excluded regions that contained no

sequencing data, estimated the composite statistic  $\mu$  in 50 SNP sliding windows, and characterized selective signals at each gene using the maximum  $\mu$  value of windows within 2 kb. We then ran RAiSD on simulated neutral data (Hämälä and Savolainen, 2019) to find outliers. We considered observed estimates 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  $\geq 10\times$  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  $6\times$ . Furthermore, to compare how many TE insertions have accumulated in the three populations without biasing our estimates with different sample sizes and sequencing depths, we chose nine individuals from each population (i.e. the sample size of J1) and randomly sampled an equal number of aligned read pairs (20 million) from each individual. 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,  $0.85\text{--}0.15$  TE heterozygote, and  $<0.15$  non-TE homozygote.

## Acknowledgements

We thank M and L Bakkom for help with the field experiments, S Alatalo for help with sequencing library preparation, and M Takou, members of the Pyhäjärvi lab, the reviewing editor J Ross-Ibarra, and two anonymous reviewers for their comments on improving 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 132611 to OS and 339702 to TH).

## Additional information

### Funding

Funder	Grant reference number	Author
Academy of Finland	132611	Outi Savolainen
Academy of Finland	339702	Tuomas Hämälä

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Tuomas Hämälä, Conceptualization, Resources, Data curation, Software, Formal analysis, Funding acquisition, Validation, Investigation, Visualization, Methodology, Writing – original draft; Weixuan Ning, Data curation, Formal analysis, Investigation, Writing – review and editing; Helmi Kuittinen, Conceptualization, Supervision; Nader Aryamanesh, Conceptualization, Resources, Data curation, Formal analysis, Supervision, Investigation, Methodology, Project administration, Writing – review and editing; Outi Savolainen, Conceptualization, Resources, Supervision, Funding acquisition, Methodology, Project administration, Writing – review and editing

### Author ORCIDs

Tuomas Hämälä  <http://orcid.org/0000-0001-8306-3397>

## Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.83115.sa1>Author response <https://doi.org/10.7554/eLife.83115.sa2>

## Additional files

## Supplementary files

- Supplementary file 1. Number of read pairs and mapping rates for each sample.
- Supplementary file 2. Bisulfite conversion efficacy for each sample.
- Supplementary file 3. Species used in estimating nucleotide conservation with GERP++.
- MDAR checklist

## 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/lyra-taRnaMet>, (copy archived at URL to be added).

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Hämälä T, Ning W, Kuittinen H, Aryamanesh N, Savolainen O	2022	Environmental response in gene expression and DNA methylation reveals factors influencing the adaptive potential of <i>Arabidopsis lyrata</i>	<a href="https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA459481">https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA459481</a>	NCBI Sequence Read Archive, PRJNA459481

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Mattila TM, Tyrmi J, Pyhäjärvi T, Savolainen O	2017	Genome-wide analysis of colonization history and concomitant selection in <i>Arabidopsis lyrata</i>	<a href="https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA357372">https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA357372</a>	NCBI Sequence Read Archive, PRJNA357372
Hämälä T, Mattila TM, Savolainen O	2018	Local adaptation and ecological differentiation under selection, migration, and drift in <i>Arabidopsis lyrata</i>	<a href="https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA459481">https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA459481</a>	NCBI Sequence Read Archive, PRJNA459481
Takou M, Hämälä T, Koch EM, Steige KA, Dittberner H, Yant L, Genete M, Sunyaev S, Castric V, Vekemans X, Savolainen O, MJ De	2021	Maintenance of adaptive dynamics and no detectable load in a range-edge outcrossing plant population	<a href="https://www.ebi.ac.uk/ena/browser/view/PRJEB33206?show=reads">https://www.ebi.ac.uk/ena/browser/view/PRJEB33206?show=reads</a>	ENA, PRJEB33206

## References

- Alachiotis N**, Pavlidis P. 2018. RAI<sub>SD</sub> detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Communications Biology* 1:79. DOI: <https://doi.org/10.1038/s42003-018-0085-8>, PMID: 30271960
- 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: <https://doi.org/10.1111/gcb.12181>, PMID: 23505261
- Amselem J**, Cornut G, Choise 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: <https://doi.org/10.1186/s13100-019-0150-y>, PMID: 30719103



- Anderson SN, Stitzer MC, Zhou P, Ross-Ibarra J, Hirsch CD, Springer NM. 2019. Dynamic patterns of transcript abundance of transposable element families in maize. *G3: Genes, Genomes, Genetics* 9:3673–3682. DOI: <https://doi.org/10.1534/g3.119.400431>, PMID: 31506319
- Baduel P, Leduque B, Ignace A, Gy I, Gil J Jr, Loudet O, Colot V, Quadana L. 2021. Genetic and environmental modulation of transposition shapes the evolutionary potential of *Arabidopsis thaliana*. *Genome Biology* 22:1–26. DOI: <https://doi.org/10.1186/s13059-021-02348-5>, PMID: 33957946
- Baldwin JM. 1896. A new factor in evolution. *The American Naturalist* 30:441–451. DOI: <https://doi.org/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: <https://doi.org/10.1073/pnas.0702552104>, PMID: 17592135
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30:2114–2120. DOI: <https://doi.org/10.1093/bioinformatics/btu170>, PMID: 24695404
- Bourgeois Y, Boissinot S. 2019. On the population dynamics of junk: a review on the population genomics of transposable elements. *Genes* 10:E419. DOI: <https://doi.org/10.3390/genes10060419>, PMID: 31151307
- Capy P, Gasperi G, Biémont C, Bazin C. 2000. Stress and transposable elements: co-evolution or useful parasites? *Heredity* 85 (Pt 2):101–106. DOI: <https://doi.org/10.1046/j.1365-2540.2000.00751.x>, PMID: 11012710
- Casacuberta E, González J. 2013. The impact of transposable elements in environmental adaptation. *Molecular Ecology* 22:1503–1517. DOI: <https://doi.org/10.1111/mec.12170>, PMID: 23293987
- Chow CN, Lee TY, Hung YC, Li GZ, Tseng KC, Liu YH, Kuo PL, Zheng HQ, Chang WC. 2019. PlantPAN3.0: a new and updated resource for reconstructing transcriptional regulatory networks from ChIP-Seq experiments in plants. *Nucleic Acids Research* 47:D1155–D1163. DOI: <https://doi.org/10.1093/nar/gky1081>, PMID: 30395277
- Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR. 2005. Genome-Wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiology* 139:5–17. DOI: <https://doi.org/10.1104/pp.105.063743>, PMID: 16166256
- 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: <https://doi.org/10.1371/journal.pcbi.1001025>, PMID: 21152010
- Devlin B, Roeder K. 1999. Genomic control for association studies. *Biometrics* 55:997–1004. DOI: <https://doi.org/10.1111/j.0006-341x.1999.00997.x>, PMID: 11315092
- 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: <https://doi.org/10.1093/bioinformatics/bts635>, PMID: 23104886
- Dubin MJ, Zhang P, Meng D, Remigereau MS, Osborne EJ, Paolo Casale F, Drewe P, Kahles A, Jean G, Vilhjálmsson B, Jagoda J, Irez S, Voronin V, Song Q, Long Q, Räscher G, Stegle O, Clark RM, Nordborg M. 2015. Dna methylation in arabidopsis has a genetic basis and shows evidence of local adaptation. *eLife* 4:e05255. DOI: <https://doi.org/10.7554/eLife.05255>, PMID: 25939354
- Falconer DS. 1981. Introduction to Quantitative Genetics. Longman.
- Farr TG, Kobrick M. 2000. Shuttle radar topography mission produces a wealth of data. *Eos, Transactions American Geophysical Union* 81:583. DOI: <https://doi.org/10.1029/EO081i048p00583>
- 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: <https://doi.org/10.1104/pp.17.00744>, PMID: 28986422
- 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: <https://doi.org/10.1111/pce.13324>, PMID: 29707792
- 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: <https://doi.org/10.1111/j.1365-2435.2007.01283.x>
- Grant V. 1977. Organismic Evolution. 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: <https://doi.org/10.1371/journal.pgen.1000695>, PMID: 19851460
- Haller BC, Messer PW. 2017. AsymptoticMK: a web-based tool for the asymptotic mcdonald-kreitman test. *G3: Genes, Genomes, Genetics* 7:1569–1575. DOI: <https://doi.org/10.1534/g3.117.039693>, PMID: 28341700
- Hämälä T, Mattila TM, Savolainen O. 2018. Local adaptation and ecological differentiation under selection, migration, and drift in *Arabidopsis lyrata*. *Evolution; International Journal of Organic Evolution* 72:1373–1386. DOI: <https://doi.org/10.1111/evo.13502>, PMID: 29741234
- Hämälä T, Savolainen O. 2019. Genomic patterns of local adaptation under gene flow in *Arabidopsis lyrata*. *Molecular Biology and Evolution* 36:2557–2571. DOI: <https://doi.org/10.1093/molbev/msz149>
- 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: <https://doi.org/10.1371/journal.pgen.1008707>, PMID: 32210431
- Hämälä T, Tiffin P. 2020. Biased gene conversion constrains adaptation in *Arabidopsis thaliana* *Genetics* 215:831–846. DOI: <https://doi.org/10.1534/genetics.120.303335>, PMID: 32414868



- 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. *PNAS* **118**:e2102914118. DOI: <https://doi.org/10.1073/pnas.2102914118>, PMID: 34408075
- Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-Lopez Z, Steffen JG, Hazzouri KM, Dewar K, Stinchcombe JR, Schoen DJ, Wang X, Schmutz J, Town CD, Edger PP, Pires JC, Schumaker KS, Jarvis DE, et al. 2013. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nature Genetics* **45**:891–898. DOI: <https://doi.org/10.1038/ng.2684>, PMID: 23817568
- Hayward LK, Sella G. 2022. Polygenic adaptation after a sudden change in environment. *eLife* **11**:e66697. DOI: <https://doi.org/10.7554/eLife.66697>, PMID: 36155653
- He F, Steige KA, Kovacova V, Göbel U, Bouzid M, Keightley PD, Beyer A, de Meaux J. 2021. Cis-Regulatory evolution spotlights species differences in the adaptive potential of gene expression plasticity. *Nature Communications* **12**:1–13. DOI: <https://doi.org/10.1038/s41467-021-23558-2>, PMID: 34099660
- 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: <https://doi.org/10.1093/molbev/msw032>, PMID: 26873578
- Hodgins-Davis A, Townsend JP. 2009. Evolving gene expression: from G to E to gxe. *Trends in Ecology & Evolution* **24**:649–658. DOI: <https://doi.org/10.1016/j.tree.2009.06.011>, PMID: 19699549
- Hohmann N, Wolf EM, Lysak MA, Koch MA. 2015. A time-calibrated road map of brassicaceae species radiation and evolutionary history. *The Plant Cell* **27**:2770–2784. DOI: <https://doi.org/10.1105/tpc.15.00482>, PMID: 26410304
- Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research* **19**:1419–1428. DOI: <https://doi.org/10.1101/gr.091678.109>, PMID: 19478138
- 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: <https://doi.org/10.1093/gbe/evx206>, PMID: 29036316
- 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: <https://doi.org/10.1093/molbev/msab270>, PMID: 34498072
- 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, Ottillar RP, Salamov AA, Schneeberger K, Spannagl M, Wang X, Yang L, Nasrallah ME, et al. 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics* **43**:476–481. DOI: <https://doi.org/10.1038/ng.807>, PMID: 21478890
- 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. *Molecular Biology and Evolution* **30**:263–271. DOI: <https://doi.org/10.1093/molbev/mss234>, PMID: 23051842
- Husby A. 2022. Wild epigenetics: insights from epigenetic studies on natural populations. *Proceedings. Biological Sciences* **289**:20211633. DOI: <https://doi.org/10.1098/rspb.2021.1633>, PMID: 35135348
- 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: <https://doi.org/10.1038/nature09861>, PMID: 21399627
- 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: <https://doi.org/10.1002/fedr.19951060115>
- Josephs EB, Lee YW, Wood CW, Schoen DJ, Wright SI, Stinchcombe JR. 2020. The evolutionary forces shaping cis- and trans-regulation of gene expression within a population of outcrossing plants. *Molecular Biology and Evolution* **37**:2386–2393. DOI: <https://doi.org/10.1093/molbev/msaa102>, PMID: 32321158
- Kawakatsu T, Huang S-SC, Jupe F, Sasaki E, Schmitz RJ, Urlich MA, Castanon R, Nery JR, Barragan C, He Y, Chen H, Dubin M, Lee C-R, Wang C, Bemm F, Becker C, O’Neil R, O’Malley RC, Quarless DX, 1001 Genomes Consortium, et al. 2016. Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* **166**:492–505. DOI: <https://doi.org/10.1016/j.cell.2016.06.044>, PMID: 27419873
- 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: <https://doi.org/10.1534/genetics.118.301120>, PMID: 29769282
- 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: <https://doi.org/10.1534/genetics.116.197145>, PMID: 28249985
- 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-Teeple M, Brady SM, Pauly M, Weigel D, et al. 2013. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *PNAS* **110**:E2655–E2662. DOI: <https://doi.org/10.1073/pnas.1309606110>, PMID: 23803858
- 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: <https://doi.org/10.1093/molbev/msw137>, PMID: 27486221
- Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**:1471–2105. DOI: <https://doi.org/10.1186/s12859-014-0356-4>, PMID: 25420514

- 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* **37**:3507–3524. DOI: <https://doi.org/10.1093/molbev/msaa185>, PMID: 32681796
- Krueger F**, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics* **27**:1571–1572. DOI: <https://doi.org/10.1093/bioinformatics/btr167>, PMID: 21493656
- 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 & Genetic Systems* **91**:111–125. DOI: <https://doi.org/10.1266/ggs.15-00065>, PMID: 27040147
- Langmead B**, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. *Nature Methods* **9**:357–359. DOI: <https://doi.org/10.1038/nmeth.1923>, PMID: 22388286
- Lasky JR**, Des Marais DL, 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: <https://doi.org/10.1093/molbev/msu170>, PMID: 24850899
- Law JA**, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews. Genetics* **11**:204–220. DOI: <https://doi.org/10.1038/nrg2719>, PMID: 20142834
- Leinonen PH**, Sandring S, Quilot B, Clauss MJ, Mitchell-Olds T, Agren J, Savolainen O. 2009. Local adaptation in European populations of *Arabidopsis lyrata* (Brassicaceae). *American Journal of Botany* **96**:1129–1137. DOI: <https://doi.org/10.3732/ajb.0800080>, PMID: 21628263
- Li H**. 2013. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. *arXiv*. <https://arxiv.org/abs/1303.3997>
- Lisch D**. 2009. Epigenetic regulation of transposable elements in plants. *Annual Review of Plant Biology* **60**:43–66. DOI: <https://doi.org/10.1146/annurev.arplant.59.032607.092744>, PMID: 19007329
- Liu J**, He Z. 2020. Small DNA methylation, big player in plant abiotic stress responses and memory. *Frontiers in Plant Science* **11**:595603. DOI: <https://doi.org/10.3389/fpls.2020.595603>, PMID: 33362826
- Lloyd JPB**, Lister R. 2022. Epigenome plasticity in plants. *Nature Reviews. Genetics* **23**:55–68. DOI: <https://doi.org/10.1038/s41576-021-00407-y>, PMID: 34526697
- Lopez-Arboleda WA**, Reinert S, Nordborg M, Korte A. 2021. Global genetic heterogeneity in adaptive traits. *Molecular Biology and Evolution* **38**:4822–4831. DOI: <https://doi.org/10.1093/molbev/msab208>, PMID: 34240182
- 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: <https://doi.org/10.1186/s13059-014-0550-8>, PMID: 25516281
- 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: <https://doi.org/10.1105/tpc.113.115352>, PMID: 24045022
- 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: <https://doi.org/10.1038/s41477-019-0548-z>, PMID: 31740772
- 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: <https://doi.org/10.1093/molbev/msx193>, PMID: 28957505
- Matuszewski S**, Hermisson J, Kopp M. 2015. Catch me if you can: adaptation from standing genetic variation to a moving phenotypic optimum. *Genetics* **200**:1255–1274. DOI: <https://doi.org/10.1534/genetics.115.178574>, PMID: 26038348
- McClintock B**. 1984. The significance of responses of the genome to challenge. *Science* **226**:792–801. DOI: <https://doi.org/10.1126/science.15739260>, PMID: 15739260
- McGuigan K**, Hoffmann AA, Sgrò CM. 2021. How is epigenetics predicted to contribute to climate change adaptation? what evidence do we need? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **376**:20200119. DOI: <https://doi.org/10.1098/rstb.2020.0119>, PMID: 33866811
- McKenna A**, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky 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**:1297–1303. DOI: <https://doi.org/10.1101/gr.107524.110>, PMID: 20644199
- Meisner J**, Albrechtsen A. 2018. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**:719–731. DOI: <https://doi.org/10.1534/genetics.118.301336>, PMID: 30131346
- Messer PW**, Petrov DA. 2013. Frequent adaptation and the mcdonald-kreitman test. *PNAS* **110**:8615–8620. DOI: <https://doi.org/10.1073/pnas.1220835110>, PMID: 23650353
- Muyle AM**, Seymour DK, Lv Y, Huettel B, Gaut BS. 2022. Gene body methylation in plants: mechanisms, functions, and important implications for understanding evolutionary processes. *Genome Biology and Evolution* **14**:evac038. DOI: <https://doi.org/10.1093/gbe/evac038>, PMID: 35298639
- Nakamura T**, Yamada KD, Tomii K, Katoh K. 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **34**:2490–2492. DOI: <https://doi.org/10.1093/bioinformatics/bty121>, PMID: 29506019
- Nowak MD**, Birkeland S, Mandáková T, Roy Choudhury R, Guo X, Gustafsson ALS, Gizaw A, Schröder-Nielsen A, Fracassetti M, Brysting AK, Rieseberg L, Slotte T, Parisod C, Lysak MA, Brochmann C. 2021. The genome of

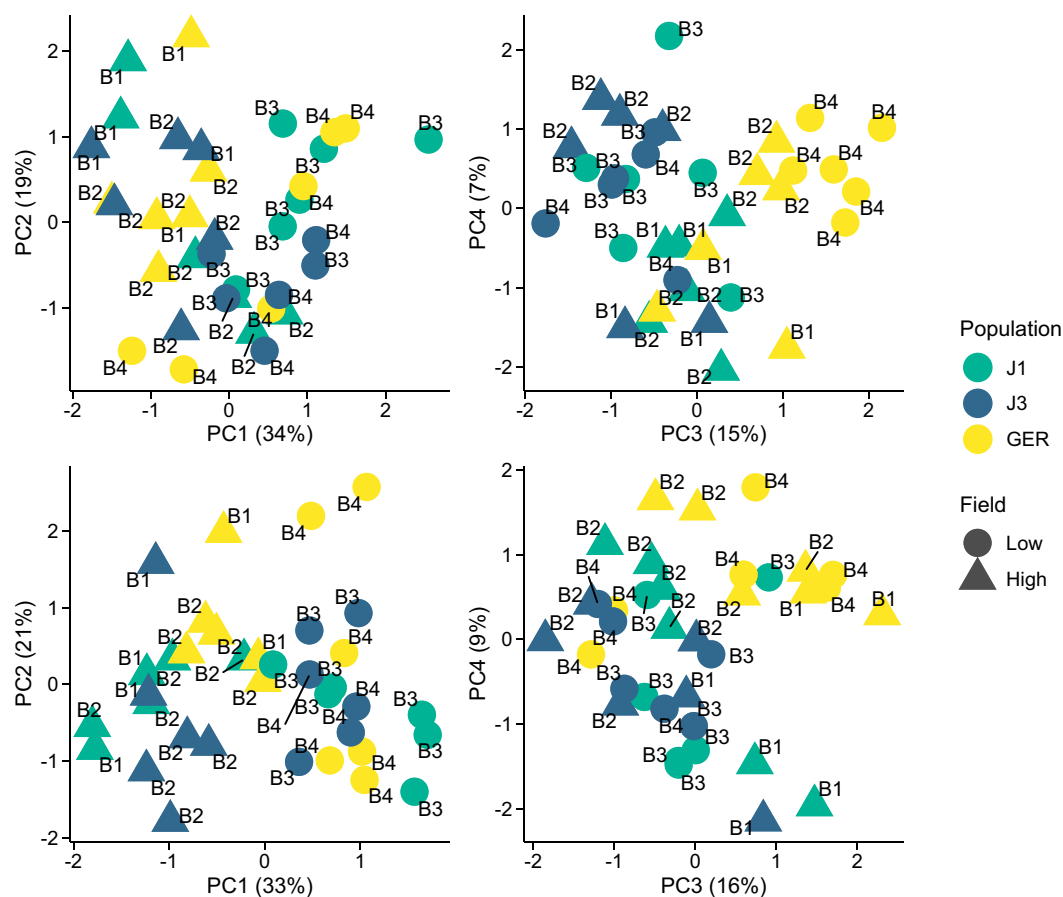
- draba nivalis* shows signatures of adaptation to the extreme environmental stresses of the Arctic. *Molecular Ecology Resources* **21**:661–676. DOI: <https://doi.org/10.1111/1755-0998.13280>, PMID: 33058468
- Parnesan C.** 2006. Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics* **37**:637–669. DOI: <https://doi.org/10.1146/annurev.ecolsys.37.091305.110100>
- 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: <https://doi.org/10.1186/s13059-016-1072-3>, PMID: 27729060
- Price TD, Qvarnström A, Irwin DE.** 2003. The role of phenotypic plasticity in driving genetic evolution. *Proceedings. Biological Sciences* **270**:1433–1440. DOI: <https://doi.org/10.1098/rspb.2003.2372>, PMID: 12965006
- 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: <https://doi.org/10.7554/eLife.15716>, PMID: 27258693
- Quadrana L, Etcheverry M, Gilly A, Caillieux E, Madoui MA, Guy J, Bortolini Silveira A, Engelen S, Baillet V, Wincker P, Aury JM, Colot V.** 2019. Transposition favors the generation of large effect mutations that may facilitate rapid adaption. *Nature Communications* **10**:3421. DOI: <https://doi.org/10.1038/s41467-019-11385-5>, PMID: 31366887
- Quinlan AR, Hall IM.** 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842. DOI: <https://doi.org/10.1093/bioinformatics/btq033>, PMID: 20110278
- Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES.** 2016. Open chromatin reveals the functional maize genome. *PNAS* **113**:E3177–E3184. DOI: <https://doi.org/10.1073/pnas.1525244113>, PMID: 27185945
- Savolainen O, Kuittinen H.** 2011. *Arabidopsis Lyrata Genetics* Genetics and Genomics of the Brassicaceae., Plant Genetics and Genomics: Crops and Models. Springer.
- Savolainen O, Lascoux M, Merilä J.** 2013. Ecological genomics of local adaptation. *Nature Reviews. Genetics* **14**:807–820. DOI: <https://doi.org/10.1038/nrg3522>, PMID: 24136507
- Schliep KP.** 2011. Phangorn: phylogenetic analysis in R. *Bioinformatics* **27**:592–593. DOI: <https://doi.org/10.1093/bioinformatics/btq706>, PMID: 21169378
- Schmalhausen II.** 1949. *Factors of Evolution: The Theory of Stabilizing Selection*. 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**:e09343. DOI: <https://doi.org/10.7554/eLife.09343>, PMID: 26196146
- Smith NGC, Eyre-Walker A.** 2002. Adaptive protein evolution in *Drosophila*. *Nature* **415**:1022–1024. DOI: <https://doi.org/10.1038/4151022a>, PMID: 11875568
- Stetter MG, Thornton K, Ross-Ibarra J.** 2018. Polygenic adaptation to distant trait optima. *PLOS Genetics* **14**:e1007794. DOI: <https://doi.org/10.1371/journal.pgen.1007794>
- Stitzer MC, Anderson SN, Springer NM, Ross-Ibarra J.** 2021. The genomic ecosystem of transposable elements in maize. *PLOS Genetics* **17**:e1009768. DOI: <https://doi.org/10.1371/journal.pgen.1009768>, PMID: 34648488
- Storey JD.** 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society* **64**:479–498. DOI: <https://doi.org/10.1111/1467-9868.00346>
- 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: <https://doi.org/10.1038/ng.942>, PMID: 21946354
- Supek F, Bošnjak M, Škunca N, Šmuc T.** 2011. REVIGO summarizes and visualizes long Lists of gene ontology terms. *PLOS ONE* **6**:e21800. DOI: <https://doi.org/10.1371/journal.pone.0021800>, PMID: 21789182
- Tajima F.** 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585–595. DOI: <https://doi.org/10.1093/genetics/123.3.585>, PMID: 2513255
- Takou M, Hämlä 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: <https://doi.org/10.1093/molbev/msaa322>, PMID: 33480994
- 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: <https://doi.org/10.1093/molbev/msx099>, PMID: 28333248
- Talbert PB, Henikoff S.** 2014. Environmental responses mediated by histone variants. *Trends in Cell Biology* **24**:642–650. DOI: <https://doi.org/10.1016/j.tcb.2014.07.006>, PMID: 25150594
- 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**:246. DOI: <https://doi.org/10.3389/fpls.2019.00246>, PMID: 30881369
- 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: <https://doi.org/10.1038/nprot.2014.114>, PMID: 25692984
- Van't Hof AE, 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: <https://doi.org/10.1038/nature17951>, PMID: 27251284
- Waddington CH.** 1942. Canalization of development and the inheritance of acquired characters. *Nature* **150**:563–565. DOI: <https://doi.org/10.1038/150563a0>

- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution; International Journal of Organic Evolution* **38**:1358–1370. DOI: <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>, PMID: 28563791
- West-Eberhard MJ. 2003. Developmental Plasticity and Evolution. Oxford University Press. DOI: <https://doi.org/10.1093/oso/9780195122343.001.0001>
- Wibowo A, Becker C, Marconi G, Durr J, Price J, Hagmann J, Papareddy R, Putra H, Kageyama J, Becker J, Weigel D, Gutierrez-Marcos J. 2016. Hyperosmotic stress memory in *Arabidopsis* is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity. *eLife* **5**:e13546. DOI: <https://doi.org/10.7554/eLife.13546>, PMID: 27242129
- 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. DOI: <https://doi.org/10.1038/nrg2165>, PMID: 17984973
- Wos G, Bohutínská M, Nosková J, Mandáková T, Kolář F. 2021a. Parallelism in gene expression between foothill and alpine ecotypes in *Arabidopsis arenosa*. *The Plant Journal* **105**:1211–1224. DOI: <https://doi.org/10.1111/tpj.15105>, PMID: 33258160
- Wos G, Choudhury RR, Kolář F, Parisod C. 2021b. Transcriptional activity of transposable elements along an elevational gradient in *Arabidopsis arenosa*. *Mobile DNA* **12**:1–12. DOI: <https://doi.org/10.1186/s13100-021-00236-0>, PMID: 33639991
- Wright S. 1931. Evolution in Mendelian populations. *Genetics* **16**:97–159. DOI: <https://doi.org/10.1093/genetics/16.2.97>, PMID: 17246615
- Wyler M, Stritt C, Walser JC, 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: <https://doi.org/10.1093/gbe/evaa180>, PMID: 32853352
- 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*). *The New Phytologist* **203**:578–591. DOI: <https://doi.org/10.1111/nph.12819>, PMID: 24750196
- 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: <https://doi.org/10.1105/tpc.17.00354>, PMID: 28733421
- Zytnicki M. 2017. Mmquant: how to count multi-mapping reads? *BMC Bioinformatics* **18**:411. DOI: <https://doi.org/10.1186/s12859-017-1816-4>, PMID: 28915787

## Appendix 1

### Quantification of batch effect on detection of differentially expressed genes

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 *Arabidopsis 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 to identify the main sources of variation. Although these reference genes have shown stable expression across treatments in *A. thaliana*, four first principal components from both gene sets were primarily impacted by the field site and population rather than sequencing batch (**Appendix 1—figure 1**). We also found no DEGs (see main text 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 4 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.



**Appendix 1—figure 1.** Expression variation along first four eigenvectors of a principal components analysis (PCA) conducted using stably expressed reference genes. The sequencing batch (B1, B2, B3, and B4) is marked next to the symbols. Top panels: 20 genes from [Czechowski et al., 2005](#). Bottom panels: 13 genes from [Kudo et al., 2016](#).