Epigenetic inheritance mediates phenotypic diversity in natural populations

Zaigham Shahzad[1,2,*], Jonathan D. Moore[1], Jaemyung Choi[1], Gaëlle Cassin-Ross[4,5], Hatem Rouached[4,5], and Daniel Zilberman[1,3,*]

[1]Department of Cell & Developmental Biology, John Innes Centre, Norwich, NR4 7UH, UK

[2]Department of Life Sciences, Syed Babar Ali School of Science and Engineering, Lahore, Pakistan

[3]Institute of Science and Technology, 3400 Klosterneuburg, Austria

[4]Plant Resilience Institute, Michigan State University, East Lansing, MI 48824, USA

[5]Department of Plant, Soil, and Microbial Sciences, Michigan State University, USA

*Correspondence: zaigham.shahzad@lums.edu.pk; daniel.zilberman@ist.ac.at

Abstract

Genetic variation is regarded as a prerequisite for evolution. In principle, epigenetic information inherited independently of DNA sequence can also enable evolution, but whether this occurs in natural populations is unknown. Here we show that natural epigenetic DNA methylation polymorphism in *Arabidopsis thaliana* gene bodies regulates gene expression and influences the variation of complex traits: fitness under heat and drought, flowering time, and accumulation of diverse minerals. The phenotypic effects of DNA methylation and local DNA sequence polymorphism are comparable, but they operate through largely distinct gene sets. Our epigenetic association analyses directly identify the relevant gene more frequently than genetic associations, likely due to reduced linkage disequilibrium. We identify numerous associations between methylation epialleles and diverse environmental conditions in native habitats, suggesting that intragenic methylation facilitates adaptation to fluctuating environments. Overall, our results demonstrate that epigenetic methylation variation fundamentally shapes phenotypic diversity in natural populations.

Introduction

The neo-Darwinian or Modern synthesis is at the heart of evolutionary biology[1]. This framework posits that DNA sequence changes generate genetic variation, and mechanisms such as natural selection and genetic drift shape this variation to adapt populations to specific environments[2,3]. Epigenetic information, which can be encoded independently of the DNA sequence, is essential for cell fate determination, development, and environmental responses in eukaryotes[4–7]. In theory, if stably heritable and independent of DNA sequence polymorphism, epigenetic variation could influence inherited traits and contribute to adaptation[8–10]. Epiallelic variation in several angiosperm genes, including *Linaria vulgaris Cyc*, tomato *CNR* and *VTE3*, maize *Spm*, rice *D1*, oil palm *MANTLED*, and *Arabidopsis thaliana FWA*, *PAI2* and *IAA7*, influences phenotypic outcomes[11,12]. However, these epialleles are either too unstable for the epigenetic states to influence a response to selection[8–10] (such as *Cyc*[13], *D1*[14] and *MANTLED*[15]), are confounded by *trans* genetic variation (such as *PAI2*[16] and *IAA7*[12]), are artificially induced (such as *FWA*[17] and *MANTLED*[15]), or evidence is lacking that heritable epiallelic variation at the gene occurs in nature (such as *CNR*[18], *VTE3*[19], *Spm*[20] and *D1*[14]), in part because the relevant studies were conducted in domesticated crops. Due to these limitations, there is currently little evidence that epigenetic inheritance mediates phenotypic diversity or influences evolutionary outcomes within natural populations[11,21].

DNA methylation can be epigenetically inherited over many generations[11,22] and occurs in transposable elements (TEs) and bodies of transcribed genes[23–26]. Plant TEs are generally methylated in all sequence contexts – CG, CHG, and CHH (H being A, T, or C)[7,23,24,27]. TE methylation induces silencing[27], confers genome stability[26,28], can influence the expression of neighboring genes[12,29–33], and its variation has been associated with all known epialleles[11,21]. Gene body methylation (gbM) occurs only in the CG context[23,24,34], although genes can also feature TE-like methylation in all contexts (teM)[35,36]. TeM is associated with silencing[25,35,36], but the function of gbM has been extensively debated. GbM is nearly ubiquitous in flowering plants[37,38] and is common in animals[23,24,39]. In both groups, gbM preferentially resides in nucleosome-wrapped DNA within the exons of conserved, constitutively transcribed genes[25,40–43]. Conservation and phenomenological coherence suggest important functions[39]. Indeed, gbM is associated with (small) gene expression differences within and between plant species[36,44–48], represses aberrant intragenic transcripts[49], and appears to be under natural selection[46,47,50]. Moreover, loss of methyltransferase function causes developmental

abnormalities in honeybees[51], animals in which methylation is principally restricted to gene bodies[52]. However, loss of gbM has not been causatively linked to changes in gene expression in plants or animals[25,53,54], leading to the proposals that gbM is a non-functional and somewhat deleterious by-product of TE methylation (in plants)[25,54,55] or has functions unrelated to gene expression (in animals)[53]. Thus, the functional and evolutionary significance of gbM has been mysterious and controversial.

Recent studies have reported extensive natural variation in TE methylation, gbM and teM in *Arabidopsis*[35,36]. Methylation levels of natural accessions are associated with climate[35], suggesting that methylation variation could be of adaptive significance. Furthermore, genetically induced methylation polymorphism can account for the inheritance of complex *Arabidopsis* traits[56–58], and methylation changes have been linked to adaptation under artificial selection[59,60]. Variation in TE methylation and teM has been repeatedly linked to genetic variation[21], but natural gbM variation is primarily epigenetic, at least in *Arabidopsis*[36,61], and hence gbM appears particularly suited to mediate natural phenotypic variation. Nonetheless, whether variation of gbM or any other type of methylation underlies phenotypic diversity or drives the evolution of complex traits in natural populations is unknown[11].

Results

GbM and teM are independent phenomena

To elucidate the biological significance of intragenic DNA methylation, we categorized genes of 948 *Arabidopsis* natural accessions into three epigenetic states: unmethylated (UM), gbM and teM using published data[35]. Consistent with published results[46], we find that gbM conservation is bimodal, with genes most commonly methylated either in over 90% of accessions, or in ≤10% (Figure 1A). The high frequency of gbM conservation supports the hypothesis that gbM is maintained by natural selection[39,46,50]. In contrast, the vast majority of genes exhibit teM in ≤10% of accessions (Figure 1B), suggesting that teM is disfavored in most genes, likely due to its deleterious effects on expression[35]. A small fraction (7.3%) of teM genes is methylated in >50% of accessions (Figure 1B), suggesting that teM effects, potentially including transcriptional silencing[32,62], might be advantageous in some genes and/or environments.

Numbers of teM and gbM genes are very weakly (negatively) correlated across accessions (Figure 1C) and are similarly weakly (positively) correlated under more restrictive published

definitions[55] of gbM and teM (Figure S1A). Additionally, intragenic teM prevalence across accessions tends to increase with decreasing gbM conservation under either gbM/teM definition (Figures 1D, and S1B- S1F), indicating that gbM and teM tend to occupy different genes. These results do not support the hypotheses that gbM originates as a by-product of teM[54] or that gbM promotes the transition to teM[55]. Instead, intragenic gbM and teM appear largely independent, which is consistent with many linages having only TE methylation (fungi and some land plants) or only gbM (many invertebrates)[23–25].

Intragenic DNA methylation is associated with transcription in natural populations

To investigate the effects of *cis* gbM and teM variation on gene expression, we analyzed associations between CG methylation (mCG) and mRNA levels of individual genes among 625 accessions for which methylation and expression data are available[35] using a linear model. We identified 614 +eQTL^{gbM} genes (eQTL stands for expression quantitative trait locus) that show a positive association between gbM and gene expression and 151 -eQTL^{gbM} genes that exhibit a negative association (Figures 1E and S2, and Tables S1 and S2) at a conservative significance threshold (Bonferroni α=0.05). Using less stringent thresholds, we identified 1,543 +eQTL^{gbM} and 715 -eQTL^{gbM} genes at 0.05 false discovery rate (FDR), and 1,887 +eQTL^{gbM} and 966 -eQTL^{gbM} genes at 0.1 FDR (Tables S1 and S2). Notably, +eQTL^{gbM} genes are 4-fold enriched under the stringent Bonferroni threshold, and only 2-fold enriched under FDR thresholds, suggesting that positive associations between gbM and expression are more likely to be real. Furthermore, +eQTL^{gbM} genes are more likely to have had gbM prior to the speciation of *A. thaliana* (Figure S3)[46], suggesting that they are under selection to retain gbM.

Average effects of gbM on expression variance are around 10%, and gbM generally accounts for ≤10% of expression variance (85.3% of eQTL^{gbM} Bonferroni genes), but in exceptional cases gbM explains around 70% of expression variance (Figures 1F and S4A). Non-associated gbM genes (NA^{gbM}) and eQTL^{gbM} genes have similar CG dinucleotide composition and length (Figures S5A and S5B) – key distinguishing characteristics of gbM genes[55] – indicating that eQTL^{gbM} genes are fairly typical. However, +eQTL^{gbM} genes show lower expression than NA^{gbM} genes (Figure S5C), suggesting that gbM may have a stronger effect when transcription is weaker. Overall, our results suggest that gbM tends to have a modest positive effect on gene expression, which is consistent with published correlations between gbM and expression[36,44,45].

In contrast to gbM, teM associations with expression are overwhelmingly negative. At Bonferroni α=0.05, we identified 202 genes in which teM is negatively associated with gene expression (-eQTL^{teM}) and only 15 positively associated genes (+eQTL^{teM}) (Table S1). As with gbM, the less stringent FDR thresholds produced more associations, but with a reduced ratio of -eQTL^{teM} to +eQTL^{teM} (Tables S1 and S2). TeM effects on gene expression are stronger than for gbM, as teM explains >20% of expression variance in 46% of eQTL^{teM} Bonferroni genes (Figures 1F and S4B). The primarily negative associations between teM and expression are consistent with existing knowledge[35,36] and indicate that teM and gbM have fundamentally different effects on gene expression.

GbM is associated with transcription in the absence of local SNP polymorphism

Because genetic and epigenetic variation can be linked in populations, we investigated whether epigenetic variants influence expression independently of *cis*-acting DNA sequence changes. To identify *cis* single nucleotide polymorphisms (SNPs) associated with expression of the 982 eQTL$^{gbM/teM}$ Bonferroni genes we performed genome-wide association (GWA) analyses. For genes where *cis* SNP eQTLs (eQTL^{SNP}) colocalized with eQTL^{gbM} or eQTL^{teM}, we defined nested populations fixed for GWA SNPs and analyzed associations between mCG and expression within each population (Figure S6). Using this approach, we found that 82.4% of +eQTL^{gbM} were retained, either because an eQTL^{SNP} is not detected (28.7%), or because a +eQTL^{gbM} was detected in at least one nested population (53.7%) (Figure 1G and Table S3). Similarly, nearly all (98.5%) -eQTL^{teM} were retained (Figure 1H and Table S3). However, 39.7% of -eQTL^{gbM} and 33.3% of +eQTL^{teM} were lost after accounting for SNP variation (Figures 1G and 1H, and Table S3). These results support the conclusion that most +eQTL^{gbM} and -eQTL^{teM} associations are real, whereas many -eQTL^{gbM} and +eQTL^{teM} associations are probably spurious and detected due to linkage between epigenetic and local SNP variation.

Allelic heterogeneity is common in *Arabidopsis*[63–67]. To account for confounding of epigenetic eQTLs by residual *cis* SNPs (other than GWA SNPs), we defined haplogroups so that accessions in a haplogroup are invariant for SNPs within retained epigenetic eQTL genes and 4 kb upstream and downstream. This analysis created numerous small groups (Table S4), and consequently many associations are expected to be lost due to lack of statistical power (Figure S7A). Nonetheless, we detected significant associations between mCG and gene expression within at least one haplogroup for a clear majority of gbM eQTLs (70.8% +eQTL^{gbM}, 64.4% -eQTL^{gbM}), for a similar fraction of -eQTL^{teM} (72.2%) and for a smaller but substantial

proportion of +eQTL^{teM} (37.5%; Figures S7B and S7C, and Table S5). Therefore, our results show that intragenic methylation is quantitatively associated with gene expression independently of local SNPs.

Local structural genetic variation does not confound eQTL^{gbM}

To evaluate whether eQTL$^{gbM/teM}$ are confounded by structural genetic variation (SV) such as insertions and deletions, we analyzed associations between methylation and expression in populations invariant for published transposable element insertion polymorphisms (TIPs)[62] within retained epigenetic eQTL genes and 4 kb upstream and downstream. We found that 98.2% of +eQTL^{gbM} are not confounded by these SVs, either because an eQTL^{SV} is not detected (85.4%) within the analyzed region, or because a +eQTL^{gbM} was detected in the population without SV (12.8%; Figure 1I). Similarly, all -eQTL^{gbM} are retained after accounting for SV (Figure 1I). In comparison to eQTL^{gbM}, a higher fraction of eQTL^{teM} (10% +eQTL^{teM} and 4% -eQTL^{teM}) are lost after accounting for SV (Figure IJ). More frequent confounding of eQTL^{teM} by SV is expected, because teM variation is known to be associated with TE SVs[32,33,36], which may promote teM and repress expression (or *vice versa*).

Undetected cryptic SV exists in *Arabidopsis*[62], but it is very unlikely to explain a substantial fraction of eQTL^{gbM}. First, fewer than 2% of +eQTL^{gbM} are confounded by known TIPs, so cryptic TIPs would have to be unrealistically abundant (>10-fold of known TIPs) to explain most +eQTL^{gbM}. Second, most SVs are linked to SNPs, and SNP-based GWA therefore identifies most SV-trait associations[68]. Systematic confounding by cryptic SV would thus manifest as systematic confounding by SNPs. Because >80% of +eQTL^{gbM} are not confounded by SNPs, to explain most +eQTL^{gbM}, cryptic SV would have to be fundamentally different from known SV by being linked with gbM variation, but unlinked with genetic variation (SNPs). This is neither parsimonious nor plausible. In contrast, the conclusion that eQTL^{gbM} are very rarely confounded by SV is consistent with the findings that gbM variation develops independently of SV in mutation accumulation (MA) lines[69,70], gbM variation in MA lines is highly correlated with that in natural accessions[71], gbM is generally neither associated with nor affected by SV in accessions[33], and local gbM variation in the *Arabidopsis* population is primarily caused by stochastic epigenetic fluctuations[61].

GbM patterns are weakly influenced by *trans* genetic polymorphism

Our results show that natural DNA methylation variation influences gene expression largely independently of local genetic polymorphism. However, methylation patterns could be governed by *trans* genetic polymorphism, such as mutations in DNA methyltransferases and other regulatory genes. Such polymorphism has indeed been identified for teM[35,36,44,72,73], and consistently we find 47.4% of retained eQTL^{teM} genes are significantly affected by *trans* QTLs (Figure 1K). Each *trans* QTL on average explains 28.5% of associated teM variation, with all *trans* QTLs accounting for 13.4% of teM variation in tested genes (Figures 1L and S8A). Thus, a substantial amount of teM variation at eQTL^{teM} genes is not epigenetic but governed by *trans* (presumably genetic) polymorphism.

Previous studies of *trans* effects on gbM have produced mixed results. Several analyses of broadly distributed *Arabidopsis* accessions did not identify strong *trans* influences on gbM[35,36], whereas a study of Swedish *Arabidopsis* did find strong *trans* effects for around 1,300 gbM genes[44]. Indeed, if we consider only the 133 Swedish accessions, we find strong *trans* effects, with on average 37.9% of gbM variance explained at 11.5% of eQTL^{gbM} (4.4% gbM variance explained overall; Figures S8B-S8D). However, when we consider all 625 worldwide accessions, these *trans* effects nearly disappear – 9.7% of genes have significant *trans* QTLs, which on average explain 10.5% of gbM variance, with *trans* genetic variation accounting for only 1% of gbM variance over all tested eQTL^{gbM} (Figures 1K, 1L, and S8A-S8D). Notably, a panel of 133 randomly chosen worldwide accessions (same size as the Swedish panel) produced results that are almost identical to those of the Swedish panel, and significantly different from the entire worldwide panel (Figures S8B-S8D). This indicates that estimates of *trans* effects on gbM variation are inflated in analyses of small populations, a phenomenon known as the Beavis effect[74,75]. Therefore, *trans* genetic polymorphism explains only a very minor fraction of overall gbM variation.

Recently, we identified several *trans* factors that influence global gbM[61]. However, in the same study we found that local gbM variation is primarily caused by epigenetic fluctuations, with *trans* factors affecting overall global gbM levels of accessions[61]. Consistently, global gbM levels of accessions are weakly associated with gbM levels of individual genes (R^2<0.1 for ~80% genes, linear model; Figure S8E), indicating that gbM at individual genes evolves independently of *trans* factors that cause global gbM differences, such as the high overall gbM in *Arabidopsis* accessions from northern Sweden[35,61]. Taken together, our results support

earlier conclusions that local gbM variation is not strongly affected by *trans* (or *cis*) genetic polymorphism, but is instead primarily epigenetic[36,61].

GbM and *cis* SNPs are associated with comparable effects on expression variance

Our results indicate that gbM is associated with expression variation independently of local and *trans* genetic polymorphism in natural populations. A key question is how much expression variance is accounted by gbM relative to local genetic variation. A published attempt to employ a statistical model to partition variance components was confounded by the linkage of methylation and SNPs, so that the model could find that most of the expression variance is accounted by SNPs, but also that SNP effects are marginal after accounting for associations between methylation and expression[45]. An alternative approach is to calculate the effects of gbM on expression variation after accounting for genetic variation and *vice versa*. Because the confounding effects of SV and *trans* genetic polymorphism on $eQTL^{gbM}$ are effectively negligible (Figures 1I-1L), we focused on *cis* SNPs. We found a strong decrease in the percent expression variance explained by gbM for lost positive and negative $eQTL^{gbM}$ after accounting for SNPs (Figure 2A). However, we did not find a significant change for gbM effects on expression of retained positive and negative $eQTL^{gbM}$ genes after accounting for SNPs (Figure 2A), with gbM on average explaining about 10% of expression variance. We obtained analogous results for effects of teM on expression variance, though these are not statistically significant (Figure S9A).

To evaluate how methylation in gene bodies confounds *cis* $eQTL^{SNP}$, we performed GWA analysis only in UM accessions for the 185 retained $eQTL^{gbM}$ and 35 retained $eQTL^{teM}$ genes that are UM in at least 100 accessions (185 NA^{gbM} genes that are UM in at least 100 accessions were randomly selected for a control group). For the control group, 65.3% of *cis* $eQTL^{SNP}$ detected in the entire population at the Bonferroni threshold remain significant after accounting for methylation effects, and 82.7% are significant at a relaxed P<1E-06 threshold (Figure 2B). Most of the $eQTL^{SNP}$ detected in the entire population are also detected in UM populations for lost $eQTL^{gbM}$, and retained as well as lost $eQTL^{teM}$ genes (Figure 2B). However, very few (1.3% at the Bonferroni threshold and 7.8% at P<1E-06) $eQTL^{SNP}$ remain in retained $eQTL^{gbM}$ genes after accounting for methylation (Figure 2B). This indicates that most *cis* $eQTL^{SNP}$ are detected in retained $eQTL^{gbM}$ genes due to linkage with gbM, as the $eQTL^{SNP}$ effectively disappear in analyses of unmethylated accessions.

Local SNPs on average explain 12.5% of expression variance in control group genes having *cis* eQTL^{SNP} at the Bonferroni threshold in the entire population (Figure 2C), which is comparable to the expression variance explained by gbM at retained eQTL^{gbM} (10.8% for +eQTL^{gbM}; 8.7% for -eQTL^{gbM}; Figure 2A). Analysis of control genes in UM accessions produces a significant increase in expression variance explained by *cis* SNPs (Figure 2C), which is expected due to the smaller population sizes[74,75]. An analogous (non-significant) increase in expression variance explained by *cis* SNPs is observed in UM populations for retained eQTL^{teM} (Figure S9B). In contrast, we find a 4 to 5-fold decrease in the effects of *cis* SNPs on expression variance of retained positive and negative eQTL^{gbM} genes in UM populations (Figure 2C), indicating that these eQTLs^{SNP} are overwhelmingly detected due to linkage of SNPs with gbM. Importantly, this analysis does not evaluate the influence of *trans* genetic polymorphism on gene expression, but seeks to compare the effects of local gbM and SNP polymorphism. Our results show that, in comparison to *cis* SNPs, gbM principally accounts for expression variance of retained eQTL^{gbM} genes and explains about the same extent of expression variance as *cis* SNPs do in control genes.

Loss of gbM specifically downregulates +eQTL^{gbM} genes

Our results show that gbM explains a significant amount of natural gene expression variation. To determine whether intragenic DNA methylation directly affects gene expression, we analyzed published RNA sequencing data from *met1* mutant seedlings and wild type Col-0 controls[49]. Mutation of the *MET1* methyltransferase causes complete loss of gbM and nearly complete loss of mCG throughout the genome[76]. For genes passing the Bonferroni threshold, we established three groups of control loci that should not be directly affected by *met1*: NA$^{gbM/teM}$ genes (no association between methylation and expression), eQTL$^{gbM/teM}$ genes that lack methylation in the wild type control accession (Col-0), and Col-0 methylated eQTL$^{gbM/teM}$ genes lost due to linked GWA SNP variation. None of the control groups showed significantly altered expression in *met1* plants, nor did any group of -eQTL^{gbM} or +eQTL^{teM} genes (Figures 2D and 2E). Only -eQTL^{teM} and +eQTL^{gbM} genes methylated in Col-0 are significantly affected by *met1* (Figures 2D and 2E).

Col-0 methylated -eQTL^{teM} genes are strongly overexpressed in *met1* (over 6-fold of wild-type; Figure 2D), consistent with the established repressive activity of teM[25,35,36]. In contrast, +eQTL^{gbM} genes are modestly downregulated (expressed at about 80% of wild-type; Figure 2E). Analysis of genes that passed less stringent significance thresholds (0.05 and 0.1 FDR)

280 produced similar results (Figures S10A and S10B). Analysis of additional Col-0 *met1* leaf [49]
281 and inflorescence[77] RNA-seq datasets, and of *met1* microarray data in the WS wild-type
282 background[78], also produced analogous results (Figures S11A-S11F). *MET1* inactivation alters
283 non-CG methylation and histone modifications[79], which could influence gene expression, but
284 are not significantly changed in any relevant gbM gene category (Figures S12A and S12B).
285 The reduced expression of only those +eQTL^{gbM} genes that are methylated in Col-0 or WS and
286 not confounded by genetic variation indicates that gbM promotes gene expression.
287 Furthermore, retained +eQTL^{gbM} genes with higher methylation in parental accessions show
288 stronger downregulation in *met1* (Col-0 and WS; Figure 2F), consistent with gbM
289 quantitatively promoting expression.

290 To further evaluate whether gbM loss alters gene expression, we analyzed a plant that is
291 heterozygous for *met1* (*met1*+/-), has relatively normal TE methylation, and limited gbM loss[49].
292 This plant also contains loss-of-function mutations in two histone H1 genes[49] and therefore
293 expression was analyzed with respect to *h1* mutant controls. Retained +eQTL^{gbM} genes
294 demethylated in this plant have significantly decreased (~35%) expression in comparison to
295 retained +eQTL^{gbM} genes that maintain gbM (Figure 2G), specifically linking gbM loss with
296 reduced expression. Moreover, we isolated six *h1met1*+/+ progeny of *h1met1*+/-. These plants
297 exhibit mosaic demethylation of gbM genes, whereas TE methylation is comparatively normal
298 (Figures S13A and S13B). Retained +eQTL^{gbM} genes demethylated in *h1met1*+/+ plants
299 display significantly reduced (~25%) expression compared to retained +eQTL^{gbM} genes that
300 keep gbM (Figure 2H). Altogether, we find that loss of gbM at retained +eQTL^{gbM} genes
301 consistently produces decreased expression, regardless of the genetic background (Col-0 or
302 WS), tissues (seedlings, leaves, or inflorescence), presence of functional *MET1*, or the extent
303 of global teM or gbM perturbation. Our results therefore constitute strong evidence that gbM
304 directly promotes gene expression. The sensitivity of +eQTL^{gbM} gene expression to gbM also
305 reinforces the conclusion that detection of these eQTLs is not substantially confounded by
306 linked cryptic genetic polymorphism in natural accessions, because the effects of such
307 polymorphism should be independent of DNA methylation.

308 **Intragenic DNA methylation is associated with plant fitness in a hot and dry climate**

309 Modulation of gene expression by natural methylation polymorphism potentiates the
310 possibility that methylation epialleles drive phenotypic variation in natural populations. To
311 uncover how DNA methylation shapes natural phenotypic diversity, we performed epigenome-

wide association (epiGWA) analyses for relative fitness using genic mCG levels in a linear model. We used published relative fitness data of *Arabidopsis* accessions grown in open-sided greenhouses in Madrid and Tübingen[80]. Linear model associations can be confounded by population structure, producing elevated rates of false-positives[81]. Therefore, we used a 'genomic control factor'[82] to correct association statistics (Table S6). We identified two significant (FDR 0.05) epigenetic relative fitness QTLs (rfQTLs) for which mCG is positively correlated with fitness in Madrid (hot climate) under low rainfall and high-density population growth (MLP conditions; Figures 3A and S14). One (rfQTL^{gbM}) corresponds to gbM in *Proline Transporter 1* (*PROT1*; AT2G39890) and the second (rfQTL^{teM}) to teM in an uncharacterized gene encoding a FBD/Leucine Rich Repeat containing protein (AT1G19410). Identification of *PROT1* was particularly intriguing, as regulation of proline accumulation is a hallmark plant response to abiotic stresses including drought[83].

PROT1 is a typical gbM gene methylated in 95.1% of accessions, but *AT1G19410* is unusual, with teM in 73.4% of accessions (Table S7), suggesting that teM in this gene may be functionally important. We identified *PROT1* as a +eQTL^{gbM} at the Bonferroni threshold, and *AT1G19410* as a -eQTL^{teM} at FDR 0.05 (Table S2), consistent with the general effects of gbM (positive) and teM (negative) on gene expression. Tripartite analysis confirmed that *PROT1* and *AT1G19410* mCG levels are significantly associated with expression across accessions (positively for *PROT1* and negatively for *AT1G19410*), and revealed that expression of both genes is significantly associated with relative MLP fitness (again positively for *PROT1* and negatively for *AT1G19410*; Figure 3B). Consistent associations between mCG, fitness, and expression were detected after accounting for SV, in two independent *PROT1* and *AT1G19410* SNP haplogroups, and in haplogroups after accounting for SV (Figure S15A and Table S8). *PROT1* is downregulated by 38% in *met1* (Figure 3C), confirming that gbM promotes *PROT1* expression. *AT1G19410* teM is lost in plants that lack DRM and CMT methyltransferases (Figure S15B), and in such *ddcc* mutants[84] *AT1G19410* expression increases about 7-fold (Figure 3D), confirming the repressive effect of teM.

To understand the relative contribution of epigenetic and genetic variants to trait heritability we performed genetic GWA analyses for relative fitness. We detected two significant genetic QTLs (rfQTLs^{SNP}) only in MLP (Table S9), suggesting that these harsh conditions (heat, drought, crowding) are conducive to identification of fitness QTLs. Notably, rfQTLs$^{gbM/teM}$ do not colocalize with rfQTLs^{SNP} (Figure S16). Heritability estimates indicate that the two

rfQTLs^{SNP} explain 18.9% of the relative fitness variance in MLP environment (Figure 3E). Strikingly, the two epigenetic rfQTLs also account for a substantial proportion of fitness variance (12.1% after accounting for *trans* QTL effects; Figure 3E). These results illustrate that natural epigenetic variation substantially contributes to the heritability of a complex plant trait.

PROT1 and AT1G19410 (*ANAHITA*) regulate plant fitness

The above results make clear predictions about the effects of gene inactivation on fitness in hot and dry conditions: *PROT1* inactivation should reduce fitness, whereas *AT1G19410* inactivation should enhance fitness. Genetic inactivation of *PROT1* indeed caused about 35% fitness reduction under experimentally induced joint heat and drought stress (Figure 3F). Mutant plants produced less biomass and had decreased survival to fruit, but had the same fecundity (seed set) as wild type (Figures 3G and S17A). Consistently, *PROT1* gbM is specifically associated with survival in MLP conditions (Figures S18A-S18C). Thus, the effects of *PROT1* inactivation match those expected from the associations between gbM, gene expression, and relative MLP fitness (Figure 3B).

Inactivation of *AT1G19410* resulted in a slight (~13%) but non-significant increase in relative fitness under heat and drought stress (Figure S17B). However, *AT1G19410* mutants have greatly enhanced (>2-fold) fitness under heat stress alone, with >2-fold increased fecundity and significantly increased fertility (percent of flowers developing siliques; Figures 3H, 3I, and S17B). Therefore, we named *AT1G19410 ANAHITA* (*ANH*) after the ancient Persian goddess of fertility and water. Notably, the association of *ANH* teM is much stronger with fecundity than survival in MLP conditions (Figures S18D-S18F). Thus, although both genes influence relative fitness under experimental conditions, *PROT1* specifically influences survival, whereas *ANH* affects fecundity. These results reveal the power of epiGWA analyses to identify novel regulators of plant fitness.

Flowering trait heritability is comparably influenced by epigenetic and genetic variation

Our results demonstrate that gbM and teM variation influences plant fitness, and thus presumably adaptation to the environment. A trait generally recognized as adaptive in *Arabidopsis* populations is timing of the transition to flowering to synchronize reproduction with the environment[85–87]. Therefore, we performed epiGWA and GWA analyses using published data for nine flowering-related traits: flowering time at 10 °C (FT_10°C) and 16 °C (FT_16°C), number of days for inflorescence stalk to reach 1 cm (DTF2), number of days to

opening of the first flower (DTF3), number of cauline leaves (CLN), number of rosette leaves (RLN), cauline branch number, number of primary inflorescence branches, and length of the primary inflorescence stalk[88]. Due to the structure of flowering phenotypes in *Arabidopsis* populations[89], we could not use a linear model with genomic control to analyze quantitative methylation variation (Figure S19 and Table S10). Instead, we used a mixed linear model (MLM) that effectively corrects population structure[90–93] (Table S11) to analyze associations between phenotypes and binary epiallelic states of genes (UM or gbM; UM or teM).

Six of the traits we examined (FT_10°C, FT_16°C, DTF2, DTF3, CLN, and RLN) are correlated with some shared genetic control[94]. EpiGWA mapping detected eight epigenetic flowering QTLs (fQTLs) at FDR 0.05 (all gbM) for these phenotypes, and no epigenetic QTLs associated with the other flowering traits (Figures 4A and S20). Seven $fQTLs^{gbM}$ contain SVs, but all loci except *AT4G33560* were retained after accounting for SV (Figures S21A-S21G). Two $fQTLs^{gbM}$ localized to the well-known flowering regulators *Flowering locus* C (*FLC*) and *FRIGIDA* (*FRI*)[95–100] (Figure 4A). To examine the roles of genes underlying the remaining six associations (*AT1G26795*, *AT1G51820*, *AT2G18210*, *AT3G09530*, *AT3G43860*, and *AT4G33560*) in transition to flowering, we examined FT_16°C in mutant lines. Mutants in all genes except *AT3G43860* have significantly altered FT_16°C (Figure 4B). Given the extensive analyses of *Arabidopsis* flowering time determinants performed over several decades[101], the discovery of five new flowering time loci underscores the power of epiGWA to identify new functional genes.

GWA analysis using SNPs identified eleven genetic flowering QTLs ($fQTLs^{SNP}$) associated with the six correlated traits, including *FLC* and *FRI* (the only overlaps with $fQTLs^{gbM}$), as well as three other known flowering regulators: *FT*, *SVP*, and *VIN3*[102–104] (Figure S22 and Table S12). Except for FT_10°C, $fQTLs^{SNP}$ explain <10% of trait heritability (Figure 4C). Strikingly, epigenetic variation explains >10% of heritability for three flowering traits before accounting for the effects of *trans* QTLs on gbM variation, and two traits after accounting for *trans* effects (Figure 4C). Accurate estimation of natural variation contribution to trait heritability is challenging[81,105] and we analyzed only the specific genetic and epigenetic QTLs we detected. Hence, the values in Figure 4C should be interpreted with caution. For example, it is very unlikely that genetic variation explains 23% of heritability for FT_16°C but only 3.4% for FT_10°C (Figure 4C). Instead, the key conclusion is that epigenetic gbM variation and genetic variation explain comparable amounts of flowering trait heritability.

FLC gbM explains *FLC* expression and flowering time in *Arabidopsis* populations

We chose two fQTLs^{gbM}, *FLC* and *FRI*, for further analyses. Not only are these extensively studied regulators of flowering time and its natural variation[99,100], but we found them associated with all (*FLC*) or nearly all (*FRI*) of the six correlated flowering phenotypes (Figure 4A). Furthermore, both genes show consistent significant associations between gbM, flowering and gene expression (Figure 4D), and both were detected as +eQTL^{gbM} at the Bonferroni threshold (Table S2). However, the *FRI* +eQTL^{gbM} was lost after accounting for genetic (SNP) variation (Table S3), *FRI* expression is not significantly altered in *met1* plants (Figure S23A), and we did not detect significant associations between *FRI* epiallelic states and FT_16°C or expression within nested genetic populations (Figures S23B and S23C). Furthermore, we found significant linkage disequilibrium[106] (r=0.725, D'=0.824, P<0.0001) between methylation states and *FRI* SNPs, suggesting that *FRI* epigenetic and genetic QTLs are redundant. Therefore, *FRI* was excluded from our estimates of gbM effects on flowering trait heritability (Figure 4C).

In contrast to *FRI*, the *FLC* +eQTL^{gbM} was retained after accounting for genetic variation (Tables S3 and S5; expression in *met1* could not be tested because *FLC* lacks gbM in Col-0, Figure S24A), and we identified significant associations between *FLC* epiallelic states, FT_16°C and expression within a nested population as well as after accounting for SV (Figures 4E, 4F, S21G, and S21H). Because multiple *FLC* alleles affect flowering time or vernalization response[32,96,97], we defined thirteen *FLC* haplotypes (Table S8), all of which contain gbM accessions (Figure 4G), suggesting complex gbM evolution at this locus. GbM accessions display significantly delayed flowering (FT_16°C) within five haplotypes (delay of >18 days in three haplotypes; Figure 4G), and significantly higher *FLC* expression in three of these haplotypes (Figure 4H). These findings demonstrate that gbM variation explains *FLC* expression and flowering time independently of *FLC* SNP and SV variation. Moreover, gbM at *FLC* is not associated with global gbM levels (Figure S24B) and is not influenced by *trans* QTLs in the worldwide population (Figure S24C), indicating that *FLC* gbM variation is epigenetic.

Arabidopsis flowering time variation is correlated with geographic regions and functionally distinct *FLC* alleles can display restricted geographic distributions[97]. We find that the density of accessions carrying *FLC* UM and gbM epialleles is generally comparable across latitudes, except 48-52° where density of UM accessions is higher, and around 63° (northern Sweden) where *FLC* gbM accessions are more common (Figures S24D-S24F). Taking advantage of the

large number of accessions distributed within 48-52° latitude, we detected a strong association (*P*=7.81E-05) between epiallelic states of *FLC* and flowering time (Figure S24G). The *FLC* fQTL^{gbM} is thus not confounded by links between flowering time and latitude, variation of global gbM (Figure S24B), *FLC* genetic polymorphism (Figures 4E, 4F, 4G, 4H, S21G, and S21H), or *trans* genetic polymorphism (Figure S24C).

GbM of the *LGM* gene explains the transition to flowering in simulated climates

Arabidopsis flowering time variation has been characterized under simulated local climates in Spain and Sweden across two seasonal plantings: spring and summer in 2008 and 2009[86]. EpiGWA identified significant fQTLs^{gbM} (FDR 0.05) associated with flowering time in simulated local climates within two uncharacterized genes, *AT4G15750* and *AT5G49440* (Figures 5A and S25, and Table S13). The *AT4G15750* fQTL^{gbM} shows sensitivity to seasonal planting in Sweden and to climate (Spain vs. Sweden) in summer (Figures S26A-S26D), whereas the *AT5G49440* fQTL^{gbM} is sensitive to seasons in Spain and to climate in spring (Figures 5B, 5C, S26E, S26F). Seasonal and climate sensitivities suggest these fQTLs^{gbM} influence flowering in response to environmental inputs, and hence could underlie local adaptation.

The *AT5G49440* fQTL^{gbM} drew our attention because it is significantly associated with flowering in six simulated climate experiments and with two flowering traits (FT_16°C and CLN) at FDR 0.05 or 0.1 (Table S14). *AT5G49440* has gbM in 6.1% of accessions, teM in 12.9%, UM in 36.4%, and is deleted in the remaining 44.6% (Figure 5D). Flowering (FT_16°C) of gbM accessions is severely delayed (by 29 days on average) compared to teM, UM or deleted accessions (Figure 5D), and hence we named *AT5G49440 Late Flowering due to Gene Body Methylation* (*LGM*). *LGM* is highly expressed in gbM and (more variably) teM accessions, whereas expression is similarly low in UM and deleted accessions (Figure 5E). Col-0 is a teM accession, and *LGM* is strongly downregulated in *met1* (which partially reduces teM; Figures 5F and S27A), indicating that intragenic methylation is required for robust *LGM* expression.

AT5G49430, the gene neighboring *LGM*, is a bromodomain WDR protein, a class that includes known flowering genes[107]. Therefore, the *LGM* association could arise due to effects on *AT5G49430* expression. However, *AT5G49430* is expressed similarly in accessions carrying *LGM* UM, gbM, teM, and deletion (epi)alleles, showing that *AT5G49430* does not underlie the *LGM* association (Figure S27B).

Curiously, although *LGM* is highly expressed in teM accessions (Figure 5E), we did not detect a significant association of teM with flowering time (Figure 5A), and teM, UM and deleted accessions flower similarly early (Figure 5D). Analysis within haplogroups (Table S8) revealed that gbM accessions flower significantly later in the two *LGM* haplotypes in which they are present (Figure 5G). In contrast, teM is not associated with delayed flowering in any haplotype (Figure 5G). Therefore, gbM (but not teM) at *LGM* is associated with delayed flowering independently of genetic variation.

A close inspection revealed that *LGM* is transcribed differently in gbM and teM accessions. The *LGM* locus produces abundant sense and antisense transcripts in Col-0[108]. We found that gbM accessions predominantly produce a longer sense and a shorter antisense transcript, whereas teM accessions produce shorter sense and longer antisense transcripts (Figures 5H and S27C). TeM extends further 5´ than gbM and covers the apparent transcriptional start site (TSS) of the long sense *LGM* mRNA (Figure 5H), which would be expected to silence the TSS[25]. Our results suggest that the position of gbM is crucial for producing a specific set of *LGM* transcripts that greatly delay flowering.

Global gbM levels are similar for *LGM* gbM and non-gbM accessions (Figure 5I), and gbM at *LGM* is not associated with *trans* QTLs (Figure S27D), indicating that gbM at *LGM* is evolving epigenetically. GbM accessions mainly grow in southern Sweden, around 55° latitude, where UM accessions are rare (Figures 5J, S27E, and S27F). EpiGWA analysis using only southern Swedish accessions found a significant association between gbM and flowering time (Figure 5K), showing that links between flowering time and latitude do not confound *LGM* $fQTL^{gbM}$ detection. These results are analogous to the ones we obtained for *FLC* (Figures S24B, S24F, and S24G). Considering the much greater relative flowering delay of gbM accessions grown in simulated Sweden climate compared to Spain climate (Figure 5C), the prevalence of *LGM* gbM in southern Sweden (Figure 5J) suggests that gbM is under selection to delay flowering in this environment.

GbM polymorphism explains natural variation in the accumulation of many minerals

GWA studies have identified numerous genes and genetic variants associated with the diversity of pant physiological traits, such as the accumulation of mineral ions[81,109,110]. Hence, we examined whether intragenic DNA methylation influences the natural variation of minerals in *Arabidopsis* leaves using published data[111]. GWA using levels of 18 minerals identified 31

501 associations, whereas 19 QTLs^{gbM} and 6 QTLs^{teM} were detected (Table S15). Only one of the
502 epigenetic QTLs (HB18, associated with manganese (Mn) accumulation) colocalizes with a
503 genetic QTL (for Rubidium (Rb) accumulation; Figure S28), indicating that – as for MLP
504 fitness and flowering time – the effects of epigenetic and genetic polymorphism on plant traits
505 manifest through distinct gene sets.

506 Next, we estimated the amount of phenotypic variance explained by QTL^{gbM} and QTL^{SNP} (the
507 ambiguous HB18 QTL was excluded from QTL^{gbM} estimates). After accounting for SV, gbM
508 polymorphism explained from 2.6% (arsenic; As) to 15.5% (potassium; K) of mineral
509 accumulation variance (Figure 6A). Consistent with our results for gene expression and other
510 phenotypes, SNP effects are comparable, explaining from 2.7% (K) to 35.3% (molybdenum;
511 Mo) of phenotypic variance (Figure 6A). Whereas several of the genetic QTLs contain genes
512 known to influence mineral accumulation, such as HKT1 for sodium (Na)[110], HMA3 for
513 cadmium (Cd)[109], and MOT1 for Mo[112], epiGWA identified entirely new genes (Figure S28,
514 Table S15). After accounting for SVs, we found that 22 epigenetic QTLs still display strong
515 associations (P<1E-03, MLM) between intragenic DNA methylation and ion variation (Table
516 S15), indicating that these QTLs^{gbM} and QTLs^{teM} are not confounded by SV. Next, we analyzed
517 10 QTL^{gbM} genes associated with the accumulation of 4 of the most easily quantified ions – K,
518 magnesium (Mg), Mn, and zinc (Zn) – using T-DNA insertion mutants. Plants with mutations
519 in 9 of these genes displayed significant changes in the accumulation of relevant minerals
520 ($p \leq 0.07$, 8 genes with $p \leq 0.03$; Figures 6B-6E). Hence, gbM polymorphism substantially
521 explains the natural variation of physiological traits, as well as life history traits such as
522 transition to flowering.

523 **EpiGWA directly identifies causal genes more frequently than GWA**

524 In GWA studies, multiple *cis* SNPs are often associated with phenotypic variation due to LD
525 between these SNPs. The associated genomic regions can be broad with multiple genes
526 underlying QTLs[109,110,113]. SNPs that show the strongest association P value are frequently in
527 non-coding regions or have unclear functional consequences[114]. Therefore, the identification
528 of genes responsible for the relevant phenotypic variation can be challenging, particularly in
529 the absence of candidate genes. To illustrate this point, we investigated published *Arabidopsis*
530 GWA studies for 48 diverse traits that have 57 validated genes (Table S16) underlying SNP
531 QTLs. We define a validated gene as one that causes the expected phenotype when mutated
532 either in the same or other studies. By calculating the number of genes between the SNP with

the lowest *P* value and the validated gene (0 means the SNP is within the validated gene), we found that the top SNP is within the validated gene in only ~54% of cases (Figure 7A). This is likely an overestimate, because cases where researchers failed to validate an association are not always reported.

In contrast to the GWA results, we validated nearly 90% (16/18) of QTL^{gbM} using mutants of the genes that displayed associations between gbM and phenotypes, a difference that is highly statistically significant (Figure 7B). We propose that this high success rate is due to epimutation rates of gbM exceeding genetic mutation rates by about 100,000-fold[61,69,70,115]. Such high turnover should rapidly disrupt linkage between gbM patterns in nearby genes, so that only gbM in the causative gene is associated with phenotypic variance. Considering that the associations obtained with GWA and epiGWA rarely overlap (Figures S16, S22, and S28), gbM-based epiGWA may be a powerful and broadly applicable gene discovery tool, as we illustrate here by identifying 15 new genes affecting six distinct phenotypes (MLP fitness, flowering time, and accumulation of K, Mg, Mn and Zn).

GbM variation may facilitate local adaptation

Arabidopsis grows in a broad range of natural environments and shows extensive local adaptation[116]. As we find that gbM affects diverse plants traits, we tested whether gbM variation may facilitate adaptation to the local environment by performing MLM epiGWA analyses using intragenic epigenetic states (UM and gbM) for 171 environmental variables[117] (latitude, longitude, temperature, precipitation, atmospheric composition, and soil properties; Table S17). We detected 571 significant associations at FDR 0.05 between 232 genes and 115 of these variables (Table S17). SNP-based GWA analyses showed that 76.9% of epigenetic associations do not colocalize with genetic associations (Figure S29A and Table S17). GbM variation in 57 genes is associated with at least three environments (Table S17). EpiGWA *P* values for these genes are strongly correlated for associated environments (Figures 8A and 8B, and Tables S18 and S19), suggesting that multiple correlated environmental conditions impose selection on epiallelic states of individual genes.

Our analysis identified several notable gbM associations with a plausible functional link to environmental adaptation (Figures 8C, 8D, 8E, S29B, S29C, and S29D) that do not overlap genetic associations (Table S17). GbM in *CCS*, which mediates heat stress responses[118], is associated with summer solar insolation, with gbM epialleles much more common in high

insolation environments (Figure 8C). GbM in *CHY1*, which is involved in cold signaling and promotes freezing tolerance[119], is associated with spring minimum temperature, with gbM epialleles rare in environments where temperature drops below -4°C (Figure 8D). GbM in *HUP9*, a regulator of flooding stress response[120], is associated with annual precipitation (Figure S29B). *PYR1* gbM variation is associated with soil excess salts, with high salt soils almost exclusively featuring gbM epialleles (Figure S29C). *PYR1* is an abscisic acid (ABA) receptor[121], and ABA is a central regulator of plant salt stress responses[122]. GbM variation in the calcium sensor *SOS3*[123] associates with soil salinity and sodicity (Figure S29D), which includes measures of calcium carbonate ($CaCO_3$) and gypsum ($CaSO_4 \cdot 2H_2O$). Nearly all accessions from soils with high salinity and sodicity have UM *SOS3* epialleles (Figure S29D). These associations suggest that natural gbM variation facilitates local adaptation in native habitats.

The most striking association we discovered is between *FLC* gbM and springtime concentration of NO_2, an anthropogenic atmospheric pollutant[124], with UM *FLC* alleles prevalent in high NO_2 environments (Figure 8E and Table S20). Because UM *FLC* accessions flower early (Figures 4D-4G), this association predicts that accessions from high NO_2 environments should flower early. Indeed, flowering time (FT_16°C) of laboratory grown *Arabidopsis* accessions is more strongly correlated with atmospheric NO_2 in native environments than with any other environmental variable, including spring frost, spring temperature, precipitation, and latitude (Figure 8F and Table S21). NO_2 concentrations vary regionally, an indicator of air quality over urban and industrial centers[125]. We find that average concentrations of NO_2 across countries show a remarkable linear correlation (R^2=0.67) with flowering time in the laboratory (Figure 8G), suggesting that earlier flowering is advantageous in higher NO_2 environments. Prevalence of the *FLC* UM epiallele in countries is also strongly correlated with NO_2 (R^2=0.68; Figure 8H), and enrichment of the UM allele at 48-52° latitude (Figure S24F) corresponds to Germany, which has especially high NO_2 (Figures 7G and 7H). These results suggest *FLC* gbM variation is selected to adapt flowering time to atmospheric NO_2 (or an unevaluated environmental factor correlated with NO_2).

Discussion

Our results demonstrate that gbM promotes gene expression (Figure 2E-H), and that its natural variation influences at least six important traits: fitness in a hot and dry climate, flowering time, and accumulation of K, Mg, Mn and Zn (Figures 3-6). Notably, epigenetically inherited

596 intragenic methylation accounts for a substantial amount of gene expression and trait variance
597 (Figures 2A, 2C, 3E, 4C, 6A, S9A and S9B). Although estimation of phenotypic variance
598 effects is challenging, our ability to consistently identify and validate associations between
599 methylation and phenotypic variance (Figures 3F-3I, 4B, 6B-6E, and 7B) indicates that the
600 phenotypic effects of intergenic methylation are comparable to those of SNPs.

601 When assessing the effects of epigenetic polymorphism, as we do here, confounding effects of
602 *cis* or *trans* genetic polymorphism must be ruled out. We demonstrate that neither known *cis*
603 SNPs nor SVs can explain our results (Figures 1G-1L, 3E, 4C, and 6A), which leaves
604 undetected cryptic polymorphism. However, to explain our results, *cis* genetic polymorphism
605 must have certain characteristics. First, it must be common in the population, because we
606 discard rare alleles in epiGWA analyses based on epiallelic states. Second, it must be
607 systematically linked with methylation variants but not with known genetic polymorphism
608 (either SNPs or SV), because the vast majority of the epigenetic effects we identify are not
609 confounded by known genetic variation (Figures 1G-1L, 3E, 4C, and 6A), and because
610 epiGWA directly identifies the relevant gene more frequently than GWA (Figure 7B), which is
611 only possible if the causative variant isn't linked with nearby genetic polymorphism. We are
612 not aware of any genetic polymorphism type that fits this description: common polymorphism
613 that is systematically unlinked with nearby SNPs or SVs. Furthermore, lack of linkage with
614 genetic polymorphism, but linkage with gbM polymorphism that evolves ~100,000 times
615 faster[61,69,70,115] isn't theoretically plausible. Thus, an entirely new and highly implausible type
616 of cryptic genetic polymorphism would have to be invoked to explain our results, whereas the
617 rapid turnover of gbM naturally explains the lack of linkage with genetic variation.
618 Furthermore, only a tiny fraction of gbM polymorphism we consider here is explained by *trans*
619 variation (Figures 1L, 3E, 4C, S24C, and S27D), and the reliable verification of phenotypic
620 associations with local gbM polymorphism (Figures 3F-3I, 4B, 6B-6E, S17A, and S17B)
621 indicates that these effects cannot be explained by cryptic *trans* variation that functions
622 independently of gbM. Overall, our results are entirely consistent with published work that
623 shows local gbM variation is primarily epigenetic in *Arabidopsis*[36,61].

624 The epigenetic effects on plant phenotypes we describe here are primarily mediated by gbM
625 (Figures 3A, S20 and S25 and Tables S13-S15). Thus, although studies of epigenetic trait
626 inheritance have focused exclusively on TE methylation or teM[11,22,32,62,126], our results suggest
627 that gbM variation is the epigenetic mechanism primarily responsible for driving phenotypic

variation in natural plant populations. Furthermore, the epigenetic and genetic effects on gene expression and phenotype manifest through distinct sets of genes, or at least their relative strengths at individual loci are sufficiently different for epiGWA to detect 16 new functional genes in this study and new phenotypic associations with the environment (such as flowering time and NO_2). This, and the propensity of epiGWA to directly identify the relevant gene (Figure 7B), make epiGWA a powerful gene discovery tool. We therefore expect epiGWA to reveal new genes that regulate key processes across a wide range of species, including major crops. QTLs controlling less than 10% of phenotypic variance have been successfully exploited to increase rice yield[127–130], suggesting that gbM polymorphism could prove useful for crop improvement.

DNA methylation is mutagenic[131], and its presence in coding sequences likely incurs a fitness cost[39]. The widespread conservation of gbM in plants and animals has therefore presented a mystery. A potential explanation is that gbM allows rapid adaptation to new or changing environments, because gbM epimutation rates are about 10^5 times greater than genetic mutation rates[61,69,70,115]. The direct relationship between gbM and mRNA levels (Figures 1E, 2F, and S2) also indicates that gbM changes should quantitatively alter gene expression, whereas most mutations are expected to be neutral or detrimental to gene function[132]. Thus, gbM variation within a population can rapidly generate a range of gene expression for selection by the environment. This argument implies that the relative gbM contribution to phenotypic variance should decrease with divergence time, as genetic differences accumulate and exert an ever-larger effect. Indeed, mathematical modelling indicates that epigenetic fluctuations that generate local gbM variation operate on thousand-year timescales, whereas overall gbM patterns over longer timescales are genetically determined[61]. Hence, at long timescales, such as those that typically separate species, genetic differences should dominate. However, our results indicate that within the *Arabidopsis* population, epigenetic gbM variation is of comparable importance to genetic variation for driving phenotypic diversity.

The association between atmospheric NO_2, flowering time and *FLC* gbM (Figures 8E-8H) presents a plausible illustration of how gbM variation may facilitate adaptation to a rapidly changing climate. The strong negative association between springtime atmospheric NO_2 in native environments and flowering time under laboratory conditions (Figures 8F and 8G) suggests that accelerated flowering is adaptive under elevated NO_2. Natural genetic variation at *FLC* is a major determinant of flowering time[88,96,97], and is associated with over twenty

environmental variables that are (or may be) related to flowering, including latitude, temperature and precipitation, but not NO_2[117]. The majority of atmospheric NO_2 (>75%) is produced by recent human activities, especially the burning of fossil fuel[124]. Therefore, *Arabidopsis* populations have had to adapt to NO_2 concentrations (or a correlated unexamined environmental variable) changing over a few decades. Genetic adaptation at *FLC* apparently has not yet occurred in response to such rapid environmental alteration, or at least is too weak for detection by GWA analysis. However, epigenetic gbM variation at *FLC* is significantly associated with atmospheric NO_2 (Figures 8E-8H), but not other environmental variables (Table S20), which is consistent with our observation that *FLC* gbM varies independently of DNA sequence (Figure 4G). Therefore, gbM variation at *FLC* has likely facilitated adaptation to anthropogenic NO_2 increases, whereas genetic variation has been involved in adaptation to environmental conditions that vary over longer timescales. This interplay between epigenetic and genetic adaptation is consistent with evolutionary models[8–11] and may be a generally important component of environmental adaptation.

Although this study investigates plant epigenetic inheritance, gbM variation is likely to influence population dynamics in many animal lineages. Direct comparison of plant gbM with vertebrate methylation is challenging because vertebrates have broadly methylated genomes in which methylation is the default state[133]. Furthermore, methylation is extensively reprogrammed during mammalian embryogenesis and germline development[134], which precludes methylation inheritance at most loci[135–137]. However, many invertebrates, from basal cnidarians to chordates, display gbM patterns highly similar to those of plants[23–25,39]. There are also intriguing indications that invertebrate methylation is epigenetically heritable. For example, a cross between two wasp species produces hybrids with chromosomes that retain their ancestral methylation patterns[138], as would be expected if methylation is epigenetically inherited. Moreover, methylation fidelity is high in honeybee sperm compared to somatic tissues [52], and methylation patterns can be stably transmitted from fathers to daughters[139]. These features of invertebrate methylation suggest that epigenetically heritable gbM facilitates environmental adaptation in animals. This may be especially important for species like aphids[140], in which self-compatibility or asexual reproduction limit the generation of genetic diversity through outcrossing. By rapidly creating functional diversity, gbM may allow animal (and plant) evolution to explore a broader range of reproductive strategies.

Acknowledgements: We thank P. Baduel and V. Colot for sharing SV data, A. Muyle for gbM conservation data, and X. Feng, C. Dean, E. Coen and Zilberman lab members for constructive comments on the manuscript. This work was supported by a European Research Council grant (725746) to DZ. This study would not have been possible without *Arabidopsis* 1001 genome, methylome, and transcriptome resources.

Ethics declarations: The authors declare no competing interest.

Supplementary Figures: This file contains material and methods and supplementary figures.

Supplementary Tables: This file contains supplementary tables 1-24.

References

1. Bowler, P.J. (2005). CHAPTER 2 - Variation from Darwin to the Modern Synthesis. In Variation, B. Hallgrímsson and B. K. Hall, eds. (Academic Press), pp. 9–27. 10.1016/B978-012088777-4/50004-1.
2. Charlesworth, D., Barton, N.H., and Charlesworth, B. (2017). The sources of adaptive variation. Proc. R. Soc. B Biol. Sci. *284*, 20162864. 10.1098/rspb.2016.2864.
3. Liu, Y. (2018). Chapter Five - Natural selection and pangenesis: the Darwinian Synthesis of evolution and genetics. In Advances in Genetics Darwin's Pangenesis and Its Rediscovery Part B., D. Kumar, ed. (Academic Press), pp. 121–142. 10.1016/bs.adgen.2018.05.010.
4. Birnbaum, K.D., and Roudier, F. (2017). Epigenetic memory and cell fate reprogramming in plants. Regeneration *4*, 15–20. https://doi.org/10.1002/reg2.73.
5. Elsherbiny, A., and Dobreva, G. (2021). Epigenetic memory of cell fate commitment. Curr. Opin. Cell Biol. *69*, 80–87. 10.1016/j.ceb.2020.12.014.
6. Greenberg, M.V.C., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. Nat. Rev. Mol. Cell Biol. *20*, 590–607. 10.1038/s41580-019-0159-6.
7. Zhang, H., Lang, Z., and Zhu, J.-K. (2018). Dynamics and function of DNA methylation in plants. Nat. Rev. Mol. Cell Biol. *19*, 489–506. 10.1038/s41580-018-0016-z.
8. Gómez-Schiavon, M., and Buchler, N.E. (2019). Epigenetic switching as a strategy for quick adaptation while attenuating biochemical noise. PLOS Comput. Biol. *15*, e1007364. 10.1371/journal.pcbi.1007364.
9. Kronholm, I., and Collins, S. (2016). Epigenetic mutations can both help and hinder adaptive evolution. Mol. Ecol. *25*, 1856–1868. https://doi.org/10.1111/mec.13296.
10. Slatkin, M. (2009). Epigenetic inheritance and the missing heritability problem. Genetics *182*, 845–850. 10.1534/genetics.109.102798.
11. Quadrana, L., and Colot, V. (2016). Plant transgenerational epigenetics. Annu. Rev. Genet. *50*, 467–491. 10.1146/annurev-genet-120215-035254.
12. Shahzad, Z., Eaglesfield, R., Carr, C., and Amtmann, A. (2020). Cryptic variation in RNA-directed DNA-methylation controls lateral root development when auxin signalling is perturbed. Nat. Commun. *11*, 218. 10.1038/s41467-019-13927-3.
13. Cubas, P., Vincent, C., and Coen, E. (1999). An epigenetic mutation responsible for natural variation in floral symmetry. Nature *401*, 157–161. 10.1038/43657.
14. Miura, K., Agetsuma, M., Kitano, H., Yoshimura, A., Matsuoka, M., Jacobsen, S.E., and Ashikari, M. (2009). A metastable DWARF1 epigenetic mutant affecting plant stature in rice. Proc. Natl. Acad. Sci. *106*, 11218–11223. 10.1073/pnas.0901942106.
15. Ong-Abdullah, M., Ordway, J.M., Jiang, N., Ooi, S.-E., Kok, S.-Y., Sarpan, N., Azimi, N., Hashim, A.T., Ishak, Z., Rosli, S.K., et al. (2015). Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. Nature *525*, 533–537. 10.1038/nature15365.
16. Melquist, S., Luff, B., and Bender, J. (1999). Arabidopsis PAI Gene Arrangements, Cytosine Methylation and Expression. Genetics *153*, 401–413. 10.1093/genetics/153.1.401.
17. Soppe, W.J.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J.M. (2000). The late flowering phenotype of FWA mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Mol. Cell *6*, 791–802. 10.1016/S1097-2765(05)00090-0.
18. Manning, K., Tör, M., Poole, M., Hong, Y., Thompson, A.J., King, G.J., Giovannoni, J.J., and Seymour, G.B. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nat. Genet. *38*, 948–952. 10.1038/ng1841.

19. Quadrana, L., Almeida, J., Asís, R., Duffy, T., Dominguez, P.G., Bermúdez, L., Conti, G., Corrêa da Silva, J.V., Peralta, I.E., Colot, V., et al. (2014). Natural occurring epialleles determine vitamin E accumulation in tomato fruits. Nat. Commun. *5*, 4027. 10.1038/ncomms5027.
20. Fedoroff, N., Schläppi, M., and Raina, R. (1995). Epigenetic regulation of the maize Spm transposon. BioEssays *17*, 291–297. 10.1002/bies.950170405.
21. Baduel, P., and Colot, V. (2021). The epiallelic potential of transposable elements and its evolutionary significance in plants. Philos. Trans. R. Soc. B Biol. Sci. *376*, 20200123. 10.1098/rstb.2020.0123.
22. Heard, E., and Martienssen, R.A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. Cell *157*, 95–109. 10.1016/j.cell.2014.02.045.
23. Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916–919. 10.1126/science.1186366.
24. Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., et al. (2010). Conservation and divergence of methylation patterning in plants and animals. Proc. Natl. Acad. Sci. *107*, 8689–8694. 10.1073/pnas.1002720107.
25. Bewick, A.J., and Schmitz, R.J. (2017). Gene body DNA methylation in plants. Curr. Opin. Plant Biol. *36*, 103–110. 10.1016/j.pbi.2016.12.007.
26. Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492. 10.1038/nrg3230.
27. Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. *11*, 204–220. 10.1038/nrg2719.
28. Du, J., Johnson, L.M., Jacobsen, S.E., and Patel, D.J. (2015). DNA methylation pathways and their crosstalk with histone methylation. Nat. Rev. Mol. Cell Biol. *16*, 519–532. 10.1038/nrm4043.
29. Qiu, Y., and Köhler, C. (2020). Mobility connects: transposable elements wire new transcriptional networks by transferring transcription factor binding motifs. Biochem. Soc. Trans. *48*, 1005–1017. 10.1042/BST20190937.
30. Hollister, J.D., and Gaut, B.S. (2009). Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. *19*, 1419–1428. 10.1101/gr.091678.109.
31. Hollister, J.D., Smith, L.M., Guo, Y.-L., Ott, F., Weigel, D., and Gaut, B.S. (2011). Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata. Proc. Natl. Acad. Sci. *108*, 2322–2327. 10.1073/pnas.1018222108.
32. Quadrana, L., Bortolini Silveira, A., Mayhew, G.F., LeBlanc, C., Martienssen, R.A., Jeddeloh, J.A., and Colot, V. (2016). The Arabidopsis thaliana mobilome and its impact at the species level. eLife *5*, e15716. 10.7554/eLife.15716.
33. Stuart, T., Eichten, S.R., Cahn, J., Karpievitch, Y.V., Borevitz, J.O., and Lister, R. (2016). Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. eLife *5*, e20777. 10.7554/eLife.20777.
34. Yan, H., Bonasio, R., Simola, D.F., Liebig, J., Berger, S.L., and Reinberg, D. (2015). DNA methylation in social insects: how epigenetics can control behavior and longevity. Annu. Rev. Entomol. *60*, 435–452. 10.1146/annurev-ento-010814-020803.
35. Kawakatsu, T., Huang, S.C., Jupe, F., Sasaki, E., Schmitz, R.J., Urich, M.A., Castanon, R., Nery, J.R., Barragan, C., He, Y., et al. (2016). Epigenomic diversity in a global collection of Arabidopsis thaliana accessions. Cell *166*, 492–505. 10.1016/j.cell.2016.06.044.
36. Schmitz, R.J., Schultz, M.D., Urich, M.A., Nery, J.R., Pelizzola, M., Libiger, O., Alix, A., McCosh, R.B., Chen, H., Schork, N.J., et al. (2013). Patterns of population epigenomic diversity. Nature *495*, 193–198. 10.1038/nature11968.

37. Niederhuth, C.E., Bewick, A.J., Ji, L., Alabady, M.S., Kim, K.D., Li, Q., Rohr, N.A., Rambani, A., Burke, J.M., Udall, J.A., et al. (2016). Widespread natural variation of DNA methylation within angiosperms. Genome Biol. *17*, 194. 10.1186/s13059-016-1059-0.
38. Takuno, S., Ran, J.-H., and Gaut, B.S. (2016). Evolutionary patterns of genic DNA methylation vary across land plants. Nat. Plants *2*, 1–7. 10.1038/nplants.2015.222.
39. Zilberman, D. (2017). An evolutionary case for functional gene body methylation in plants and animals. Genome Biol. *18*, 87. 10.1186/s13059-017-1230-2.
40. Lyons, D.B., and Zilberman, D. (2017). DDM1 and Lsh remodelers allow methylation of DNA wrapped in nucleosomes. eLife *6*, e30674. 10.7554/eLife.30674.
41. Lewis, S.H., Ross, L., Bain, S.A., Pahita, E., Smith, S.A., Cordaux, R., Miska, E.A., Lenhard, B., Jiggins, F.M., and Sarkies, P. (2020). ------Widespread conservation and lineage-specific diversification of genome-wide DNA methylation patterns across arthropods. PLOS Genet. *16*, e1008864. 10.1371/journal.pgen.1008864.
42. Dixon, G.B., Bay, L.K., and Matz, M.V. (2016). Evolutionary consequences of DNA methylation in a basal metazoan. Mol. Biol. Evol. *33*, 2285–2293. 10.1093/molbev/msw100.
43. Bräutigam, K., and Cronk, Q. (2018). DNA methylation and the evolution of developmental complexity in plants. Front. Plant Sci. *9*. 10.3389/fpls.2018.01447.
44. Dubin, M.J., Zhang, P., Meng, D., Remigereau, M.-S., Osborne, E.J., Paolo Casale, F., Drewe, P., Kahles, A., Jean, G., Vilhjálmsson, B., et al. (2015). DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. eLife *4*, e05255. 10.7554/eLife.05255.
45. Meng, D., Dubin, M., Zhang, P., Osborne, E.J., Stegle, O., Clark, R.M., and Nordborg, M. (2016). Limited contribution of DNA methylation variation to expression regulation in Arabidopsis thaliana. PLOS Genet. *12*, e1006141. 10.1371/journal.pgen.1006141.
46. Muyle, A., Ross-Ibarra, J., Seymour, D.K., and Gaut, B.S. (2021). Gene body methylation is under selection in Arabidopsis thaliana. Genetics *https://doi.org/10.1093/genetics/iyab061*. 10.1093/genetics/iyab061.
47. Seymour, D.K., and Gaut, B.S. (2020). Phylogenetic shifts in gene body methylation correlate with gene expression and reflect trait conservation. Mol. Biol. Evol. *37*, 31–43. 10.1093/molbev/msz195.
48. Takuno, S., Seymour, D.K., and Gaut, B.S. (2017). The evolutionary dynamics of orthologs that shift in gene body methylation between Arabidopsis species. Mol. Biol. Evol. *34*, 1479–1491. 10.1093/molbev/msx099.
49. Choi, J., Lyons, D.B., Kim, M.Y., Moore, J.D., and Zilberman, D. (2020). DNA methylation and Histone H1 jointly repress transposable elements and aberrant intragenic transcripts. Mol. Cell *77*, 310-323.e7. 10.1016/j.molcel.2019.10.011.
50. Takuno, S., and Gaut, B.S. (2013). Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. Proc. Natl. Acad. Sci. *110*, 1797–1802. 10.1073/pnas.1215380110.
51. Kucharski, R., Maleszka, J., Foret, S., and Maleszka, R. (2008). Nutritional control of reproductive status in honeybees via DNA methylation. Science *319*, 1827–1830. 10.1126/science.1153069.
52. Harris, K.D., Lloyd, J.P.B., Domb, K., Zilberman, D., and Zemach, A. (2019). DNA methylation is maintained with high fidelity in the honey bee germline and exhibits global non-functional fluctuations during somatic development. Epigenetics Chromatin *12*, 62. 10.1186/s13072-019-0307-4.
53. Bewick, A.J., Sanchez, Z., Mckinney, E.C., Moore, A.J., Moore, P.J., and Schmitz, R.J. (2019). Dnmt1 is essential for egg production and embryo viability in the large milkweed bug, Oncopeltus fasciatus. Epigenetics Chromatin *12*, 6. 10.1186/s13072-018-0246-5.
54. Bewick, A.J., Ji, L., Niederhuth, C.E., Willing, E.-M., Hofmeister, B.T., Shi, X., Wang, L., Lu, Z., Rohr, N.A., Hartwig, B., et al. (2016). On the origin and evolutionary consequences of gene body DNA methylation. Proc. Natl. Acad. Sci. *113*, 9111–9116. 10.1073/pnas.1604666113.

55. Zhang, Y., Wendte, J.M., Ji, L., and Schmitz, R.J. (2020). Natural variation in DNA methylation homeostasis and the emergence of epialleles. Proc. Natl. Acad. Sci. *117*, 4874–4884. 10.1073/pnas.1918172117.
56. Cortijo, S., Wardenaar, R., Colomé-Tatché, M., Gilly, A., Etcheverry, M., Labadie, K., Caillieux, E., Hospital, F., Aury, J.-M., Wincker, P., et al. (2014). Mapping the epigenetic basis of complex traits. Science *343*, 1145–1148. 10.1126/science.1248127.
57. Johannes, F., Porcher, E., Teixeira, F.K., Saliba-Colombani, V., Simon, M., Agier, N., Bulski, A., Albuisson, J., Heredia, F., Audigier, P., et al. (2009). Assessing the impact of transgenerational epigenetic variation on complex traits. PLOS Genet. *5*, e1000530. 10.1371/journal.pgen.1000530.
58. Furci, L., Jain, R., Stassen, J., Berkowitz, O., Whelan, J., Roquis, D., Baillet, V., Colot, V., Johannes, F., and Ton, J. (2019). Identification and characterisation of hypomethylated DNA loci controlling quantitative resistance in Arabidopsis. eLife *8*, e40655. 10.7554/eLife.40655.
59. Schmid, M.W., Heichinger, C., Coman Schmid, D., Guthörl, D., Gagliardini, V., Bruggmann, R., Aluri, S., Aquino, C., Schmid, B., Turnbull, L.A., et al. (2018). Contribution of epigenetic variation to adaptation in Arabidopsis. Nat. Commun. *9*, 4446. 10.1038/s41467-018-06932-5.
60. Kronholm, I., Bassett, A., Baulcombe, D., and Collins, S. (2017). Epigenetic and Genetic Contributions to Adaptation in Chlamydomonas. Mol. Biol. Evol. *34*, 2285–2306. 10.1093/molbev/msx166.
61. Briffa, A., Hollwey, E., Shahzad, Z., Moore, J.D., Lyons, D.B., Howard, M., and Zilberman, D. (2023). Millennia-long epigenetic fluctuations generate intragenic DNA methylation variance in Arabidopsis populations. Cell Syst. 10.1016/j.cels.2023.10.007.
62. Baduel, P., Leduque, B., Ignace, A., Gy, I., Gil, J., Loudet, O., Colot, V., and Quadrana, L. (2021). Genetic and environmental modulation of transposition shapes the evolutionary potential of Arabidopsis thaliana. Genome Biol. *22*, 138. 10.1186/s13059-021-02348-5.
63. Kerdaffrec, E., Filiault, D.L., Korte, A., Sasaki, E., Nizhynska, V., Seren, Ü., and Nordborg, M. (2016). Multiple alleles at a single locus control seed dormancy in Swedish Arabidopsis. eLife *5*, e22502. 10.7554/eLife.22502.
64. Lopez-Arboleda, W.A., Reinert, S., Nordborg, M., and Korte, A. (2021). Global genetic heterogeneity in adaptive traits. Mol. Biol. Evol. 10.1093/molbev/msab208.
65. Shahzad, Z., Canut, M., Tournaire-Roux, C., Martinière, A., Boursiac, Y., Loudet, O., and Maurel, C. (2016). A potassium-dependent oxygen sensing pathway regulates plant root hydraulics. Cell *167*, 87-98.e14. 10.1016/j.cell.2016.08.068.
66. Tang, N., Shahzad, Z., Lonjon, F., Loudet, O., Vailleau, F., and Maurel, C. (2018). Natural variation at XND1 impacts root hydraulics and trade-off for stress responses in Arabidopsis. Nat. Commun. *9*, 3884. 10.1038/s41467-018-06430-8.
67. Zhang, X., Cal, A.J., and Borevitz, J.O. (2011). Genetic architecture of regulatory variation in Arabidopsis thaliana. Genome Res. *21*, 725–733. 10.1101/gr.115337.110.
68. Hufford, M.B., Seetharam, A.S., Woodhouse, M.R., Chougule, K.M., Ou, S., Liu, J., Ricci, W.A., Guo, T., Olson, A., Qiu, Y., et al. (2021). De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. Science *373*, 655–662.
69. Becker, C., Hagmann, J., Müller, J., Koenig, D., Stegle, O., Borgwardt, K., and Weigel, D. (2011). Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature *480*, 245–249. 10.1038/nature10555.
70. Schmitz, R.J., Schultz, M.D., Lewsey, M.G., O'Malley, R.C., Urich, M.A., Libiger, O., Schork, N.J., and Ecker, J.R. (2011). Transgenerational epigenetic instability is a source of novel methylation variants. Science *334*, 369–373. 10.1126/science.1212959.
71. Vidalis, A., Živković, D., Wardenaar, R., Roquis, D., Tellier, A., and Johannes, F. (2016). Methylome evolution in plants. Genome Biol. *17*, 264. 10.1186/s13059-016-1127-5.

72. Shen, X., Jonge, J.D., Forsberg, S.K.G., Pettersson, M.E., Sheng, Z., Hennig, L., and Carlborg, Ö. (2014). Natural CMT2 variation is associated with genome-wide methylation changes and temperature seasonality. PLOS Genet. *10*, e1004842. 10.1371/journal.pgen.1004842.
73. Sasaki, E., Kawakatsu, T., Ecker, J.R., and Nordborg, M. (2019). Common alleles of CMT2 and NRPE1 are major determinants of CHH methylation variation in Arabidopsis thaliana. PLOS Genet. *15*, e1008492. 10.1371/journal.pgen.1008492.
74. Beavis, W.D. (1998). QTL Analyses: Power, Precision, and Accuracy. In Molecular Dissection of Complex Traits (CRC Press), pp. 145–162.
75. Xu, S. (2003). Theoretical basis of the Beavis effect. Genetics *165*, 2259–2268.
76. Catoni, M., Griffiths, J., Becker, C., Zabet, N.R., Bayon, C., Dapp, M., Lieberman-Lazarovich, M., Weigel, D., and Paszkowski, J. (2017). DNA sequence properties that predict susceptibility to epiallelic switching. EMBO J. *36*, 617–628. 10.15252/embj.201695602.
77. Oberlin, S., Sarazin, A., Chevalier, C., Voinnet, O., and Marí-Ordóñez, A. (2017). A genome-wide transcriptome and translatome analysis of Arabidopsis transposons identifies a unique and conserved genome expression strategy for Ty1/Copia retroelements. Genome Res. *27*, 1549–1562. 10.1101/gr.220723.117.
78. Li, W., Liu, H., Cheng, Z.J., Su, Y.H., Han, H.N., Zhang, Y., and Zhang, X.S. (2011). DNA Methylation and Histone Modifications Regulate De Novo Shoot Regeneration in Arabidopsis by Modulating WUSCHEL Expression and Auxin Signaling. PLOS Genet. *7*, e1002243. 10.1371/journal.pgen.1002243.
79. Deleris, A., Stroud, H., Bernatavichute, Y., Johnson, E., Klein, G., Schubert, D., and Jacobsen, S.E. (2012). Loss of the DNA Methyltransferase MET1 Induces H3K9 Hypermethylation at PcG Target Genes and Redistribution of H3K27 Trimethylation to Transposons in Arabidopsis thaliana. PLOS Genet. *8*, e1003062. 10.1371/journal.pgen.1003062.
80. Exposito-Alonso, M., Burbano, H.A., Bossdorf, O., Nielsen, R., and Weigel, D. (2019). Natural selection on the Arabidopsis thaliana genome in present and future climates. Nature *573*, 126–129. 10.1038/s41586-019-1520-9.
81. Atwell, S., Huang, Y.S., Vilhjálmsson, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M., Hu, T.T., et al. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature *465*, 627–631. 10.1038/nature08800.
82. Devlin, B., and Roeder, K. (1999). Genomic control for association studies. Biometrics *55*, 997–1004. https://doi.org/10.1111/j.0006-341X.1999.00997.x.
83. Verbruggen, N., and Hermans, C. (2008). Proline accumulation in plants: a review. Amino Acids *35*, 753–759. 10.1007/s00726-008-0061-6.
84. Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat. Struct. Mol. Biol. *21*, 64–72. 10.1038/nsmb.2735.
85. Ågren, J., Oakley, C.G., Lundemo, S., and Schemske, D.W. (2017). Adaptive divergence in flowering time among natural populations of Arabidopsis thaliana: Estimates of selection and QTL mapping. Evolution *71*, 550–564. https://doi.org/10.1111/evo.13126.
86. Li, Y., Huang, Y., Bergelson, J., Nordborg, M., and Borevitz, J.O. (2010). Association mapping of local climate-sensitive quantitative trait loci in Arabidopsis thaliana. Proc. Natl. Acad. Sci. *107*, 21199–21204. 10.1073/pnas.1007431107.
87. Taylor, M.A., Wilczek, A.M., Roe, J.L., Welch, S.M., Runcie, D.E., Cooper, M.D., and Schmitt, J. (2019). Large-effect flowering time mutations reveal conditionally adaptive paths through fitness landscapes in Arabidopsis thaliana. Proc. Natl. Acad. Sci. *116*, 17890–17899. 10.1073/pnas.1902731116.
88. The 1001 Genomes Consortium (2016). 1,135 genomes reveal the global pattern of polymorphism in Arabidopsis thaliana. Cell *166*, 481–491. 10.1016/j.cell.2016.05.063.

946 89. Shindo, C., Bernasconi, G., and Hardtke, C.S. (2007). Natural genetic variation in Arabidopsis:
947 tools, traits and prospects for evolutionary ecology. Ann. Bot. *99*, 1043–1054.
948 10.1093/aob/mcl281.
949 90. Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. (2007).
950 TASSEL: software for association mapping of complex traits in diverse samples.
951 Bioinformatics *23*, 2633–2635. 10.1093/bioinformatics/btm308.
952 91. Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J.,
953 Arnett, D.K., Ordovas, J.M., et al. (2010). Mixed linear model approach adapted for genome-
954 wide association studies. Nat. Genet. *42*, 355–360. 10.1038/ng.546.
955 92. Vilhjálmsson, B.J., and Nordborg, M. (2013). The nature of confounding in genome-wide
956 association studies. Nat. Rev. Genet. *14*, 1–2. 10.1038/nrg3382.
957 93. Weigel, D., and Nordborg, M. (2015). Population genomics for understanding adaptation in
958 wild plant species. Annu. Rev. Genet. *49,* 315–338. 10.1146/annurev-genet-120213-092110.
959 94. Grimm, D.G., Roqueiro, D., Salomé, P.A., Kleeberger, S., Greshake, B., Zhu, W., Liu, C.,
960 Lippert, C., Stegle, O., Schölkopf, B., et al. (2017). EasyGWAS: a cloud-based platform for
961 comparing the results of genome-wide association studies. Plant Cell *29*, 5–19.
962 10.1105/tpc.16.00551.
963 95. Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., and Dean, C. (2000). Molecular
964 analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time.
965 Science *290*, 344–347. 10.1126/science.290.5490.344.
966 96. Hepworth, J., Antoniou-Kourounioti, R.L., Berggren, K., Selga, C., Tudor, E.H., Yates, B., Cox,
967 D., Collier Harris, B.R., Irwin, J.A., Howard, M., et al. (2020). Natural variation in autumn
968 expression is the major adaptive determinant distinguishing Arabidopsis FLC haplotypes.
969 eLife *9*, e57671. 10.7554/eLife.57671.
970 97. Li, P., Filiault, D., Box, M.S., Kerdaffrec, E., Oosterhout, C. van, Wilczek, A.M., Schmitt, J.,
971 McMullan, M., Bergelson, J., Nordborg, M., et al. (2014). Multiple FLC haplotypes defined by
972 independent cis-regulatory variation underpin life history diversity in Arabidopsis thaliana.
973 Genes Dev. *28*, 1635–1640. 10.1101/gad.245993.114.
974 98. Shindo, C., Aranzana, M.J., Lister, C., Baxter, C., Nicholls, C., Nordborg, M., and Dean, C.
975 (2005). Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time
976 of Arabidopsis. Plant Physiol. *138*, 1163–1173. 10.1104/pp.105.061309.
977 99. Adrian, J., Torti, S., and Turck, F. (2009). From decision to commitment: the molecular
978 memory of flowering. Mol. Plant *2*, 628–642. 10.1093/mp/ssp031.
979 100. Whittaker, C., and Dean, C. (2017). The FLC locus: a platform for discoveries in epigenetics
980 and adaptation. Annu. Rev. Cell Dev. Biol. *33*, 555–575. 10.1146/annurev-cellbio-100616-
981 060546.
982 101. Blümel, M., Dally, N., and Jung, C. (2015). Flowering time regulation in crops—what did we
983 learn from Arabidopsis? Curr. Opin. Biotechnol. *32*, 121–129. 10.1016/j.copbio.2014.11.023.
984 102. Sung, S., and Amasino, R.M. (2004). Vernalization in Arabidopsis thaliana is mediated by the
985 PHD finger protein VIN3. Nature *427*, 159–164. 10.1038/nature02195.
986 103. Gregis, V., Sessa, A., Colombo, L., and Kater, M.M. (2006). AGL24, SHORT VEGETATIVE
987 PHASE, and APETALA1 redundantly control AGAMOUS during early stages of flower
988 development in Arabidopsis. Plant Cell *18*, 1373–1382. 10.1105/tpc.106.041798.
989 104. Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S.,
990 Gissot, L., Turnbull, C., et al. (2007). FT protein movement contributes to long-distance
991 signaling in floral induction of Arabidopsis. Science *316*, 1030–1033.
992 10.1126/science.1141752.
993 105. Gallagher, M.D., and Chen-Plotkin, A.S. (2018). The post-GWAS era: from association to
994 function. Am. J. Hum. Genet. *102*, 717–730. 10.1016/j.ajhg.2018.04.002.

106. Flint-Garcia, S.A., Thornsberry, J.M., and Buckler, E.S. (2003). Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. *54*, 357–374. 10.1146/annurev.arplant.54.031902.134907.
107. van Nocker, S., and Ludwig, P. (2003). The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function. BMC Genomics *4*, 50. 10.1186/1471-2164-4-50.
108. Stolc, V., Samanta, M.P., Tongprasit, W., Sethi, H., Liang, S., Nelson, D.C., Hegeman, A., Nelson, C., Rancour, D., Bednarek, S., et al. (2005). Identification of transcribed sequences in Arabidopsis thaliana by using high-resolution genome tiling arrays. Proc. Natl. Acad. Sci. *102*, 4453–4458. 10.1073/pnas.0408203102.
109. Chao, D.-Y., Silva, A., Baxter, I., Huang, Y.S., Nordborg, M., Danku, J., Lahner, B., Yakubova, E., and Salt, D.E. (2012). Genome-Wide Association Studies Identify Heavy Metal ATPase3 as the Primary Determinant of Natural Variation in Leaf Cadmium in Arabidopsis thaliana. PLOS Genet. *8*, e1002923. 10.1371/journal.pgen.1002923.
110. Baxter, I., Brazelton, J.N., Yu, D., Huang, Y.S., Lahner, B., Yakubova, E., Li, Y., Bergelson, J., Borevitz, J.O., Nordborg, M., et al. (2010). A Coastal Cline in Sodium Accumulation in Arabidopsis thaliana Is Driven by Natural Variation of the Sodium Transporter AtHKT1;1. PLOS Genet. *6*, e1001193. 10.1371/journal.pgen.1001193.
111. Campos, A.C.A.L., van Dijk, W.F.A., Ramakrishna, P., Giles, T., Korte, P., Douglas, A., Smith, P., and Salt, D.E. (2021). 1,135 ionomes reveal the global pattern of leaf and seed mineral nutrient and trace element diversity in Arabidopsis thaliana. Plant J. *106*, 536–554. 10.1111/tpj.15177.
112. Tomatsu, H., Takano, J., Takahashi, H., Watanabe-Takahashi, A., Shibagaki, N., and Fujiwara, T. (2007). An Arabidopsis thaliana high-affinity molybdate transporter required for efficient uptake of molybdate from soil. Proc. Natl. Acad. Sci. *104*, 18807–18812. 10.1073/pnas.0706373104.
113. Sasaki, E., Köcher, T., Filiault, D.L., and Nordborg, M. (2021). Revisiting a GWAS peak in Arabidopsis thaliana reveals possible confounding by genetic heterogeneity. Heredity *127*, 245–252. 10.1038/s41437-021-00456-3.
114. Lee, P.H., Lee, C., Li, X., Wee, B., Dwivedi, T., and Daly, M. (2018). Principles and methods of in-silico prioritization of non-coding regulatory variants. Hum. Genet. *137*, 15–30. 10.1007/s00439-017-1861-0.
115. Van der Graaf, A., Wardenaar, R., Neumann, D.A., Taudt, A., Shaw, R.G., Jansen, R.C., Schmitz, R.J., Colomé-Tatché, M., and Johannes, F. (2015). Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. Proc. Natl. Acad. Sci. *112*, 6676–6681. 10.1073/pnas.1424254112.
116. Ågren, J., and Schemske, D.W. (2012). Reciprocal transplants demonstrate strong adaptive differentiation of the model organism Arabidopsis thaliana in its native range. New Phytol. *194*, 1112–1122. https://doi.org/10.1111/j.1469-8137.2012.04112.x.
117. Ferrero-Serrano, Á., and Assmann, S.M. (2019). Phenotypic and genome-wide association with the local environment of Arabidopsis. Nat. Ecol. Evol. *3*, 274–285. 10.1038/s41559-018-0754-5.
118. Guan, Q., Lu, X., Zeng, H., Zhang, Y., and Zhu, J. (2013). Heat stress induction of miR398 triggers a regulatory loop that is critical for thermotolerance in Arabidopsis. Plant J. *74*, 840–851. https://doi.org/10.1111/tpj.12169.
119. Dong, C.-H., Zolman, B.K., Bartel, B., Lee, B., Stevenson, B., Agarwal, M., and Zhu, J.-K. (2009). Disruption of Arabidopsis CHY1 reveals an important role of metabolic status in plant cold stress signaling. Mol. Plant *2*, 59–72. 10.1093/mp/ssn063.
120. Lee, S.C., Mustroph, A., Sasidharan, R., Vashisht, D., Pedersen, O., Oosumi, T., Voesenek, L.A.C.J., and Bailey-Serres, J. (2011). Molecular characterization of the submergence

1045 response of the Arabidopsis thaliana ecotype Columbia. New Phytol. *190*, 457–471.
1046 https://doi.org/10.1111/j.1469-8137.2010.03590.x.
1047 121. Cutler, S.R., Rodriguez, P.L., Finkelstein, R.R., and Abrams, S.R. (2010). Abscisic acid:
1048 emergence of a core signaling network. Annu. Rev. Plant Biol. *61*, 651–679.
1049 10.1146/annurev-arplant-042809-112122.
1050 122. Yu, Z., Duan, X., Luo, L., Dai, S., Ding, Z., and Xia, G. (2020). How plant hormones mediate salt
1051 stress responses. Trends Plant Sci. *25*, 1117–1130. 10.1016/j.tplants.2020.06.008.
1052 123. Sánchez-Barrena, M.J., Martínez-Ripoll, M., Zhu, J.-K., and Albert, A. (2005). The structure of
1053 the Arabidopsis thaliana SOS3: molecular mechanism of sensing calcium for salt stress
1054 response. J. Mol. Biol. *345*, 1253–1264. 10.1016/j.jmb.2004.11.025.
1055 124. Lelieveld, J., Beirle, S., Hörmann, C., Stenchikov, G., and Wagner, T. (2015). Abrupt recent
1056 trend changes in atmospheric nitrogen dioxide over the Middle East. Sci. Adv. *1*, e1500498.
1057 10.1126/sciadv.1500498.
1058 125. Larkin, A., Geddes, J.A., Martin, R.V., Xiao, Q., Liu, Y., Marshall, J.D., Brauer, M., and Hystad,
1059 P. (2017). Global Land Use Regression Model for Nitrogen Dioxide Air Pollution. Environ. Sci.
1060 Technol. *51*, 6957–6964. 10.1021/acs.est.7b01148.
1061 126. Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic
1062 regulation of the genome. Nat. Rev. Genet. *8*, 272–285. 10.1038/nrg2072.
1063 127. McCouch, S.R., Wright, M.H., Tung, C.-W., Maron, L.G., McNally, K.L., Fitzgerald, M., Singh,
1064 N., DeClerck, G., Agosto-Perez, F., Korniliev, P., et al. (2016). Open access resources for
1065 genome-wide association mapping in rice. Nat. Commun. *7*, 10532. 10.1038/ncomms10532.
1066 128. Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., Shao, D., Xu, C., Li, X., Xiao, J., et al. (2011).
1067 Natural variation in GS5 plays an important role in regulating grain size and yield in rice. Nat.
1068 Genet. *43*, 1266–1269. 10.1038/ng.977.
1069 129. Yano, K., Morinaka, Y., Wang, F., Huang, P., Takehara, S., Hirai, T., Ito, A., Koketsu, E.,
1070 Kawamura, M., Kotake, K., et al. (2019). GWAS with principal component analysis identifies a
1071 gene comprehensively controlling rice architecture. Proc. Natl. Acad. Sci. *116*, 21262–21267.
1072 10.1073/pnas.1904964116.
1073 130. Fujita, D., Trijatmiko, K.R., Tagle, A.G., Sapasap, M.V., Koide, Y., Sasaki, K., Tsakirpaloglou, N.,
1074 Gannaban, R.B., Nishimura, T., Yanagihara, S., et al. (2013). NAL1 allele from a rice landrace
1075 greatly increases yield in modern indica cultivars. Proc. Natl. Acad. Sci. *110*, 20431–20436.
1076 10.1073/pnas.1310790110.
1077 131. Robertson, K.D., and A.Jones, P. (2000). DNA methylation: past, present and future
1078 directions. Carcinogenesis *21*, 461–467. 10.1093/carcin/21.3.461.
1079 132. Eyre-Walker, A., and Keightley, P.D. (2007). The distribution of fitness effects of new
1080 mutations. Nat. Rev. Genet. *8*, 610–618. 10.1038/nrg2146.
1081 133. Luo, C., Hajkova, P., and Ecker, J.R. (2018). Dynamic DNA methylation: In the right place at
1082 the right time. Science *361*, 1336–1340. 10.1126/science.aat6806.
1083 134. Zeng, Y., and Chen, T. (2019). DNA methylation reprogramming during mammalian
1084 development. Genes *10*. 10.3390/genes10040257.
1085 135. Nadeau, J.H. (2015). The nature of evidence for and against epigenetic inheritance. Genome
1086 Biol. *16*, 137. 10.1186/s13059-015-0709-y.
1087 136. Horsthemke, B. (2018). A critical view on transgenerational epigenetic inheritance in
1088 humans. Nat. Commun. *9*, 2973. 10.1038/s41467-018-05445-5.
1089 137. Kazachenka, A., Bertozzi, T.M., Sjoberg-Herrera, M.K., Walker, N., Gardner, J., Gunning, R.,
1090 Pahita, E., Adams, S., Adams, D., and Ferguson-Smith, A.C. (2018). Identification,
1091 characterization, and heritability of murine metastable epialleles: implications for non-
1092 genetic inheritance. Cell *175*, 1259-1271.e13. 10.1016/j.cell.2018.09.043.
1093 138. Wang, X., Werren, J.H., and Clark, A.G. (2016). Allele-specific transcriptome and methylome
1094 analysis reveals stable iInheritance and cis-regulation of DNA methylation in Nasonia. PLOS
1095 Biol. *14*, e1002500. 10.1371/journal.pbio.1002500.

1096 139. Yagound, B., Remnant, E.J., Buchmann, G., and Oldroyd, B.P. (2020). Intergenerational
1097 transfer of DNA methylation marks in the honey bee. Proc. Natl. Acad. Sci. *117*, 32519–
1098 32527. 10.1073/pnas.2017094117.
1099 140. Mathers, T.C., Mugford, S.T., Percival-Alwyn, L., Chen, Y., Kaithakottil, G., Swarbreck, D.,
1100 Hogenhout, S.A., and Oosterhout, C. van (2019). Sex-specific changes in the aphid DNA
1101 methylation landscape. Mol. Ecol. *28*, 4228–4241. https://doi.org/10.1111/mec.15216.
1102

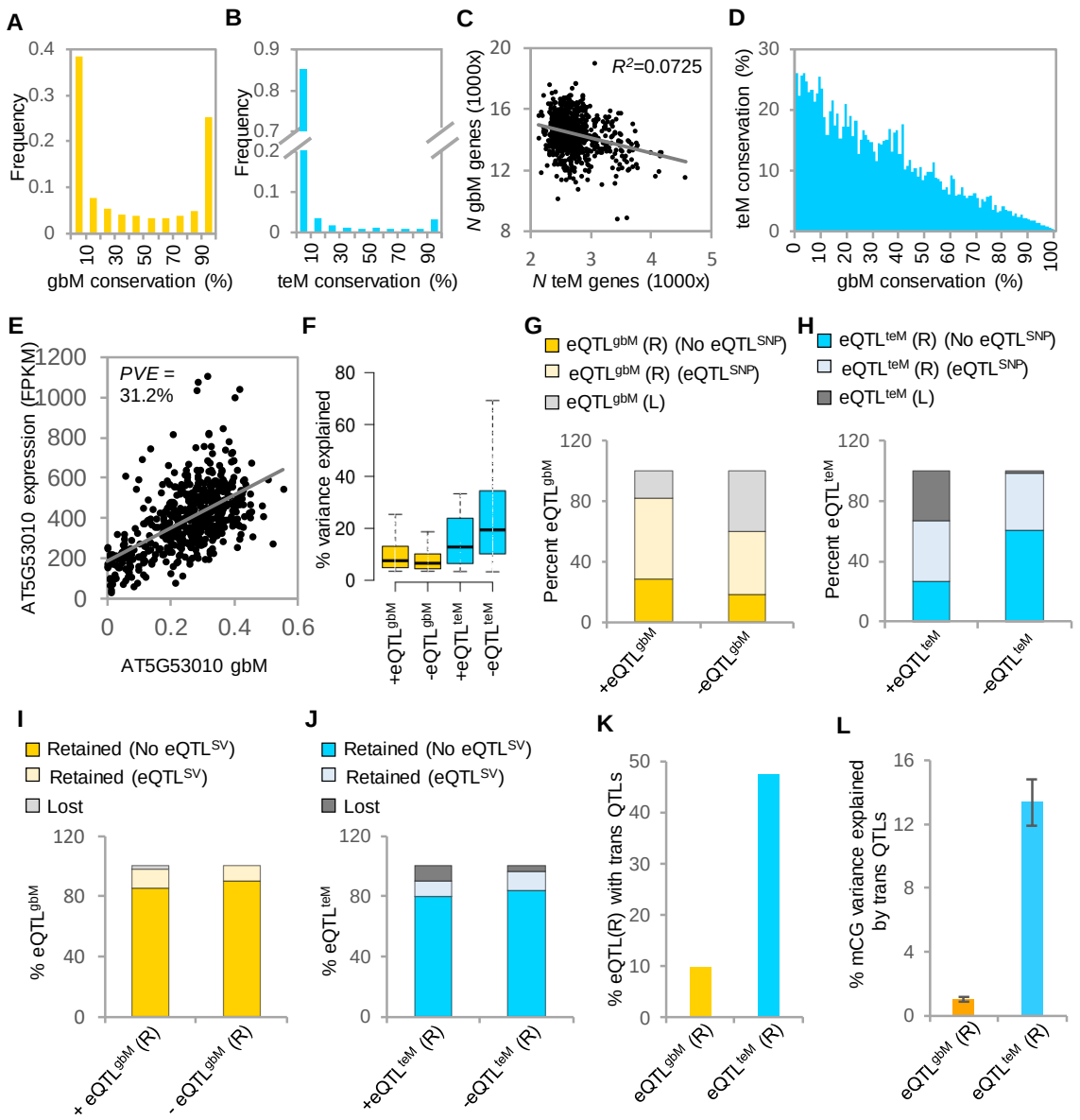

Figure 1. GbM is associated with gene expression independently of genetic variation in natural populations. (**A** and **B**) Frequency distribution of gbM (**A**) and teM (**B**) conservation across 33,056 *Arabidopsis* genes in 948 accessions. (**C**) Correlation between the number (*N*) of gbM and teM genes across accessions. (**D**) Average teM conservation within gbM conservation percentiles. (**E**) GbM and expression of AT5G53010 across accessions. Percent expression variance explained (*PVE*) by gbM is indicated. (**F**) Effect sizes of gbM and teM on expression of positive and negative eQTL$^{gbM/teM}$ genes. (**G** and **H**) Frequency of retained or lost eQTL^{gbM} (**G**) and eQTL^{teM} (**H**) genes after accounting for expression GWA SNPs. (**I** and **J**) Frequency of retained or lost eQTL^{gbM} (**I**) and eQTL^{teM} (**J**) genes after accounting for structural variation (SV) effects on expression. (**K**) Percentage of eQTL$^{gbM/teM}$ retained (R) genes that have *trans* genetic QTLs associated with the variation of gbM or teM. (**L**) Average effects (± standard error) of *trans* polymorphism on mCG variation in all retained eQTL$^{gbM/teM}$ genes.

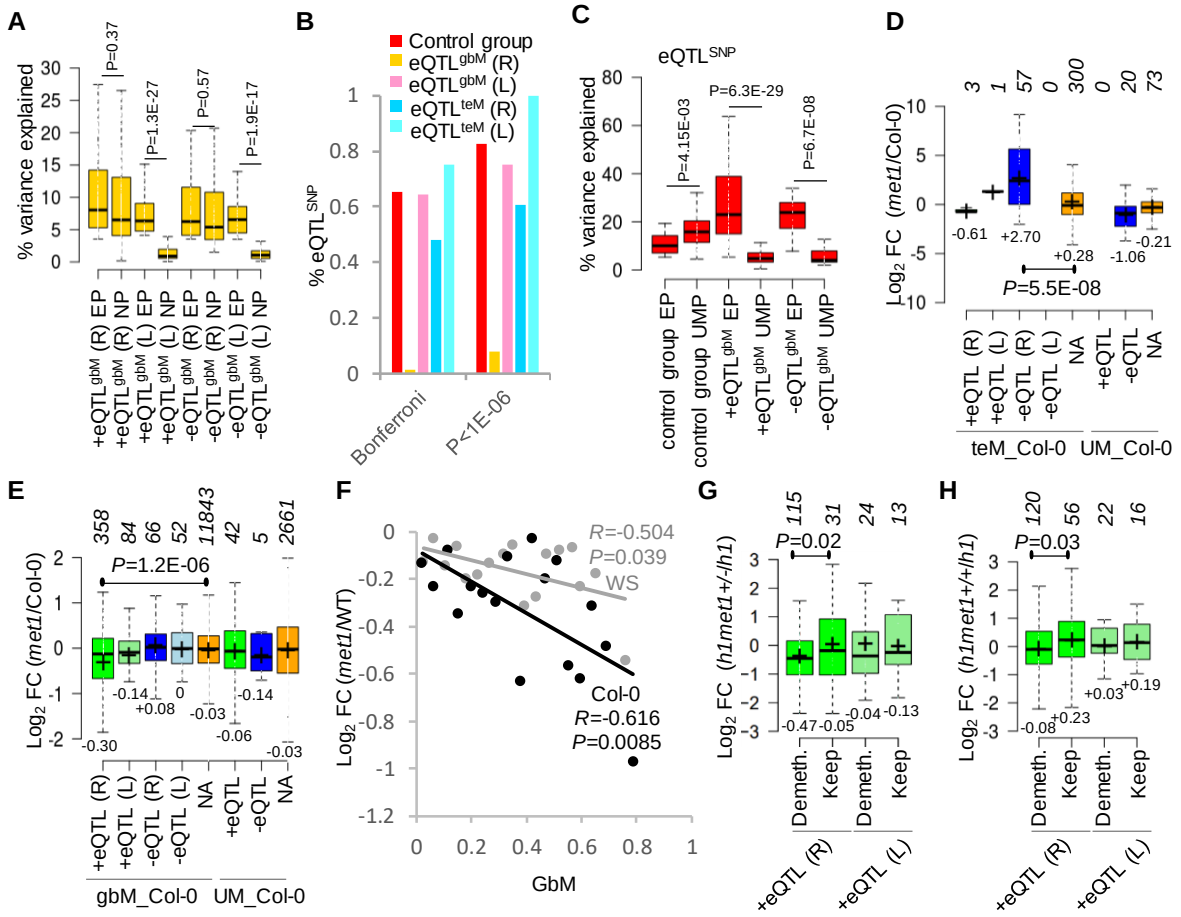

Figure 2. GbM promotes gene expression. (**A**) Effect sizes of gbM on expression of positive and negative retained (R) or lost (L) eQTL^{gbM} genes in the entire population (EP) and after accounting for expression GWA SNPs in nested populations (NP). (**B**) Percent of eQTL^{SNP} detected as significant at the Bonferroni threshold and an arbitrary relaxed threshold (*P*<1E-06) after accounting for methylation variation. (**C**) Effects of SNPs on expression variation of a subset of control group (NA^{gbm}) genes and retained eQTL^{gbM} genes in the entire (EP) and unmethylated populations (UMP). (**D** and **E**) Expression in *met1* seedlings compared to Col-0 of Bonferroni (α=0.05) eQTL^{teM} (**D**) and eQTL^{gbM} (**E**) genes retained (R) or lost (L) after accounting for genetic variation. Genes methylated (gbM_Col-0; teM_Col-0) and unmethylated (UM_Col-0) in Col-0 were analyzed separately. Numbers of genes within each group are indicated. *P*, Wilcoxon rank sum test. (**F**) Relationship between gbM levels of retained +eQTL^{gbM} genes in Col-0 and WS wild type (WT) plants and $\log_2$ fold change in expression in *met1* compared to wild type. Genes were grouped by gbM levels. *R* and *P* values correspond to Pearson's correlation. (**G** and **H**) Expression in *h1met1+/-* (**G**) and *h1met1+/+* (**H**) compared to *h1* of Bonferroni (α=0.05) retained (R) or lost (L) eQTL^{gbM} genes that are either demethylated or keep methylation. *P*, Student's t-test.

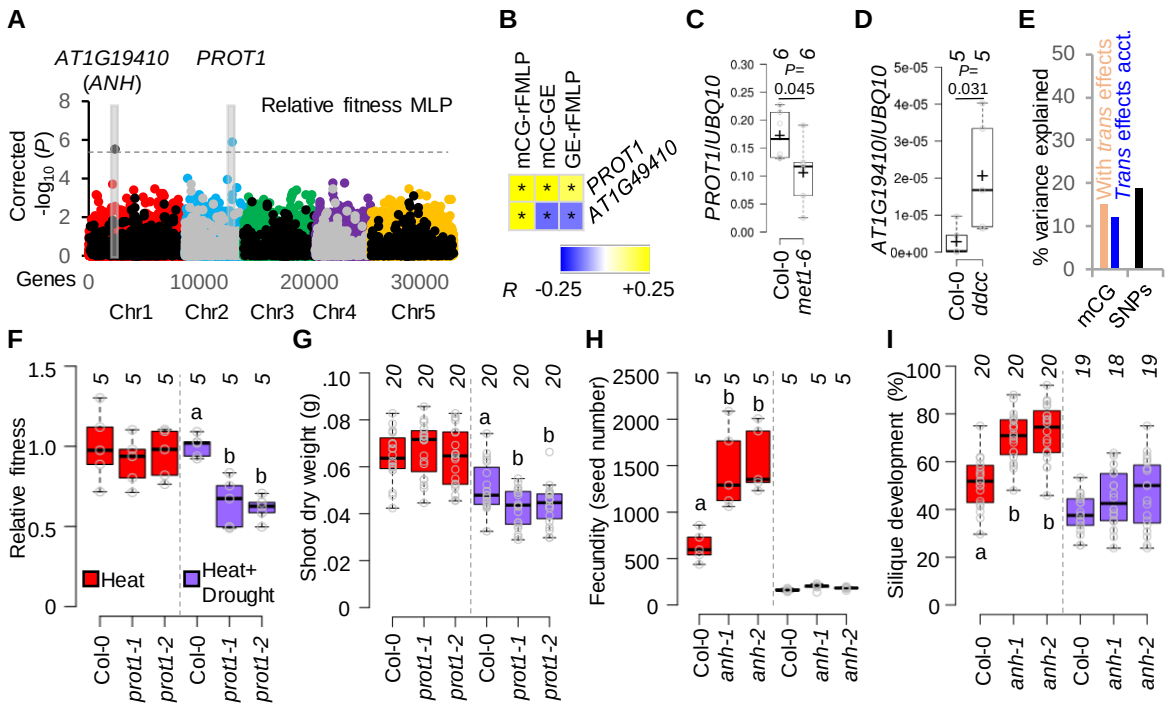

Figure 3. Intragenic DNA methylation affects relative fitness. (**A**) Manhattan plot of epiGWA mapping for relative fitness in MLP conditions. GbM markers are depicted with colored dots and teM markers are shown using grey and black dots. Horizontal dashed line shows 0.05 FDR. (**B**) Tripartite associations between intragenic DNA methylation (mCG), gene expression (GE), and relative fitness in MLP (rFMLP). *P<0.05, Pearson's correlation test. (**C** and **D**) Transcript levels of *PROT1* (**C**) or *AT1G19410* (*ANH*; **D**) relative to *UBQ10* in Col-0 and indicated mutants, assessed by qRT-PCR in six (**C**) or five (**D**) biological replicates. *P*, two-tailed Student's t test. (**E**) Percent variation in MLP relative fitness explained by epigenetic (mCG) and genetic (SNPs) QTLs. Effects of *trans* genetic QTLs on mCG variation were accounted (acct.). (**F**-**I**) Fitness (**F**) or shoot dry weight (**G**) of *prot1* mutants (two independent alleles) and fecundity (**H**) or fertility (**I**) of *anh* mutants (two independent alleles) relative to Col-0 under heat or joint heat and drought stress. Numbers of independent experiments are indicated for fitness (**F**) and fecundity (**H**), and plant numbers are indicated for shoot weight (**G**) and fertility (**I**). Different letters signify *P*<0.05, one-way ANOVA, Tukey's test.

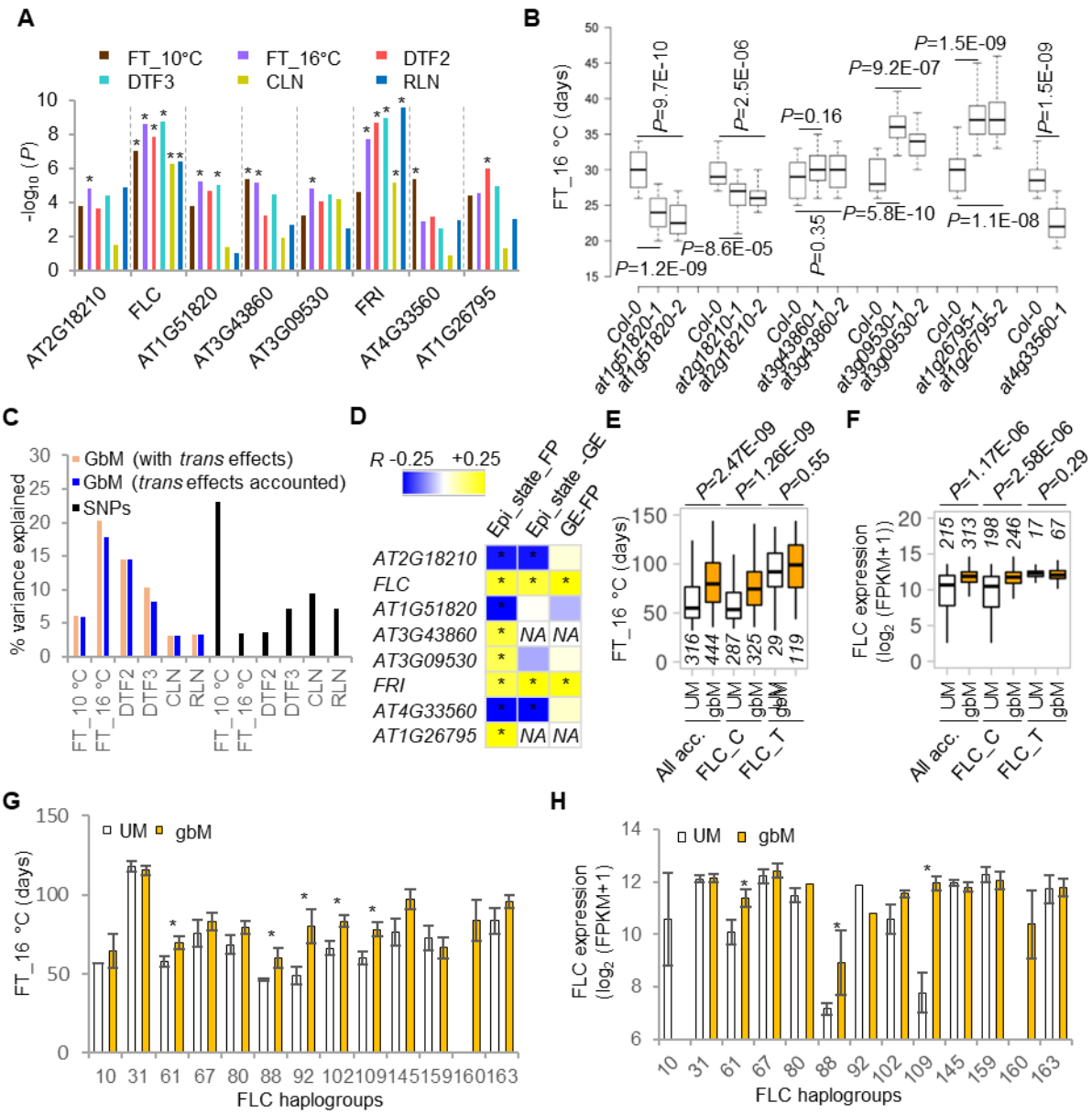

Figure 4. **Flowering time is comparably influenced by natural genetic and epigenetic variation**. (**A**) *P* values (-$\log_{10}$) of epigenetic flowering QTLs. *Significant at 0.05 FDR. (**B**) Genetic inactivation of fQTL^{gbM} genes causes altered flowering in Col-0. Flowering time of Col-0 and knock-out mutants of *AT1G51820*, *AT2G18210*, *AT3G43860*, *AT3G09530*, *AT1G26795*, and *AT4G33560*. Plants were grown at 16 °C. *P*, two-tailed t test. (**C**) Percent variation in flowering phenotypes explained by epigenetic (gbM) and genetic (SNPs) QTLs. Effects of *trans* genetic QTLs on gbM variation are accounted. Effect of *cis* GWA SNP and SV on fQTL^{gbM} are also accounted. (**D**) Tripartite associations between epiallelic states (Epi_state), gene expression (GE), and flowering phenotypes (FP). FT_10°C was used for AT4G33560, DTF2 for AT1G26795, and FT_16°C for the remaining fQTLs^{gbM}. *P<0.05, MLM epiGWA for Epi_state-FP and Epi_state-GE and Pearson's correlation test for GE-FP. (**E** and **F**) Association of FLC epiallelic states with FT_16°C (**E**) and FLC expression (**F**) in all accessions and nested populations (FLC_C and FLC_T) fixed for FLC SNP associated with flowering time. Numbers of accessions for each group are indicated. *P* values correspond to MLM epiGWA. (**G** and **H**) Association of FLC epiallelic states with FT_16°C (**G**) and FLC expression (**H**) in thirteen FLC haplogroups. Only accessions without TE polymorphism around FLC are considered to account for SV. *P<0.05, two-tailed Student's t-test.

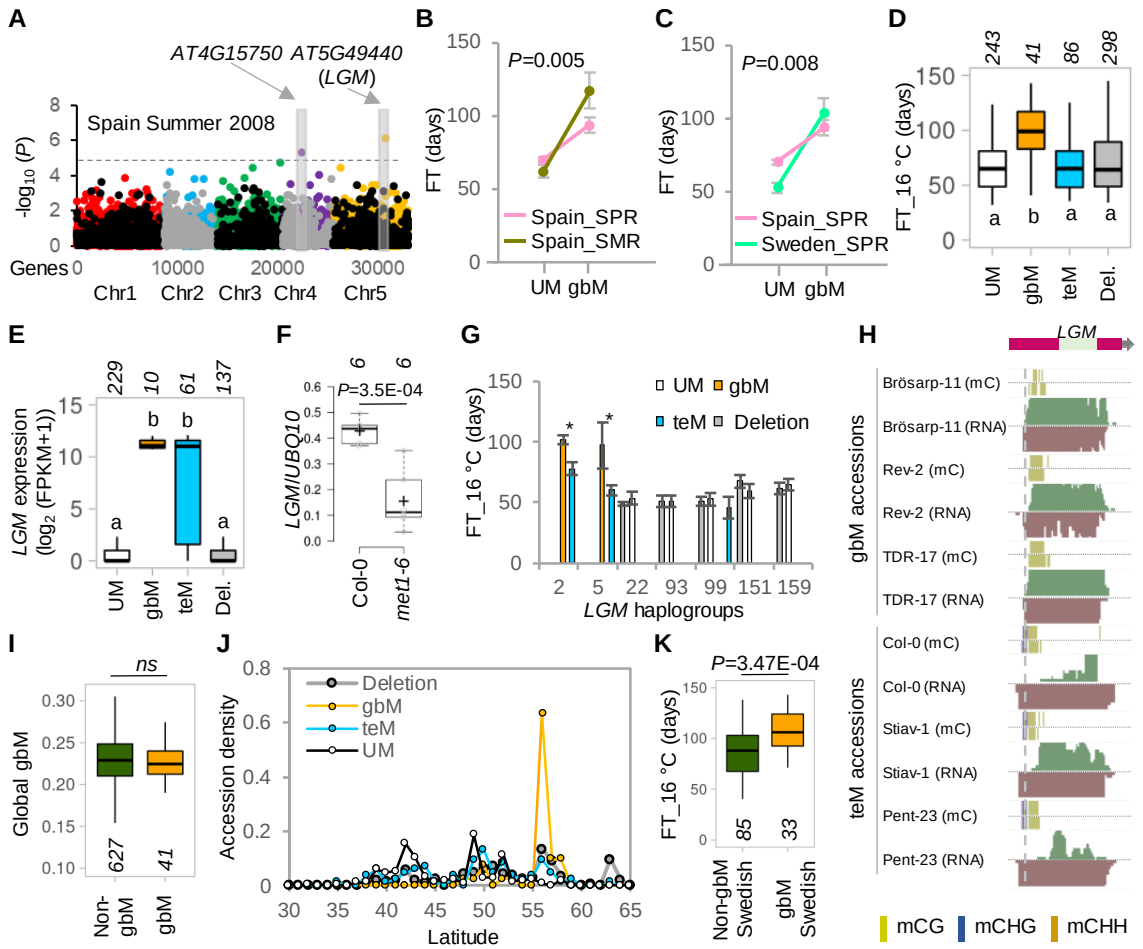

Figure 5. GbM in *LGM* delays flowering. (**A**) Manhattan plot of MLM epiGWA mapping for flowering time in the 2008 Spain summer planting. GbM markers are depicted with colored dots and teM markers are shown using grey and black dots. Horizontal dashed line shows 0.05 FDR. (**B** and **C**) Sensitivity of AT5G49440 (*LGM*) flowering time (FT) QTL^{gbM} to seasons (spring (SPR), summer (SMR)) in Spain (**B**) and spring climates (Spain, Sweden; **C**). *P*, two-way ANOVA. (**D** and **E**) Association of flowering (FT_16°C; **D**) and *LGM* expression (**E**) with *LGM* UM, gbM, teM, and deletion (epi)alleles in the indicated numbers of accessions. Different letters denote significant accession group differences evaluated by pairwise MLM epiGWA. (**F**) Transcript levels of *LGM* relative to *UBQ10* in Col-0 and *met1*, assessed by qRT-PCR in six biological replicates. *P*, two-tailed Student's t test. (**G**) Effects of *LGM* gbM on FT_16° in *LGM* haplogroups. *P<0.05, two-tailed Student's t-test. (**H**) AnnoJ browser snapshots of methylation along with sense (green) and antisense (brown) transcripts for representative *LGM* gbM and teM accessions. Vertical dashed line shows the long sense transcript start site. The light green box corresponds to *LGM* coding sequence. (**I**) Genome-wide average gbM levels of accessions with gbM or non-gbM *LGM*. Accession numbers are indicated. *ns* = non-significant, two-tailed Student's t test. (**J**) Density of *LGM* UM, gbM, teM, and deletion accessions across latitude. (**K**) Association of *LGM* gbM and non-gbM (epi)alleles with FT_16°C in southern Swedish accessions. Accession numbers are indicated. *P* value corresponds to MLM epiGWA.

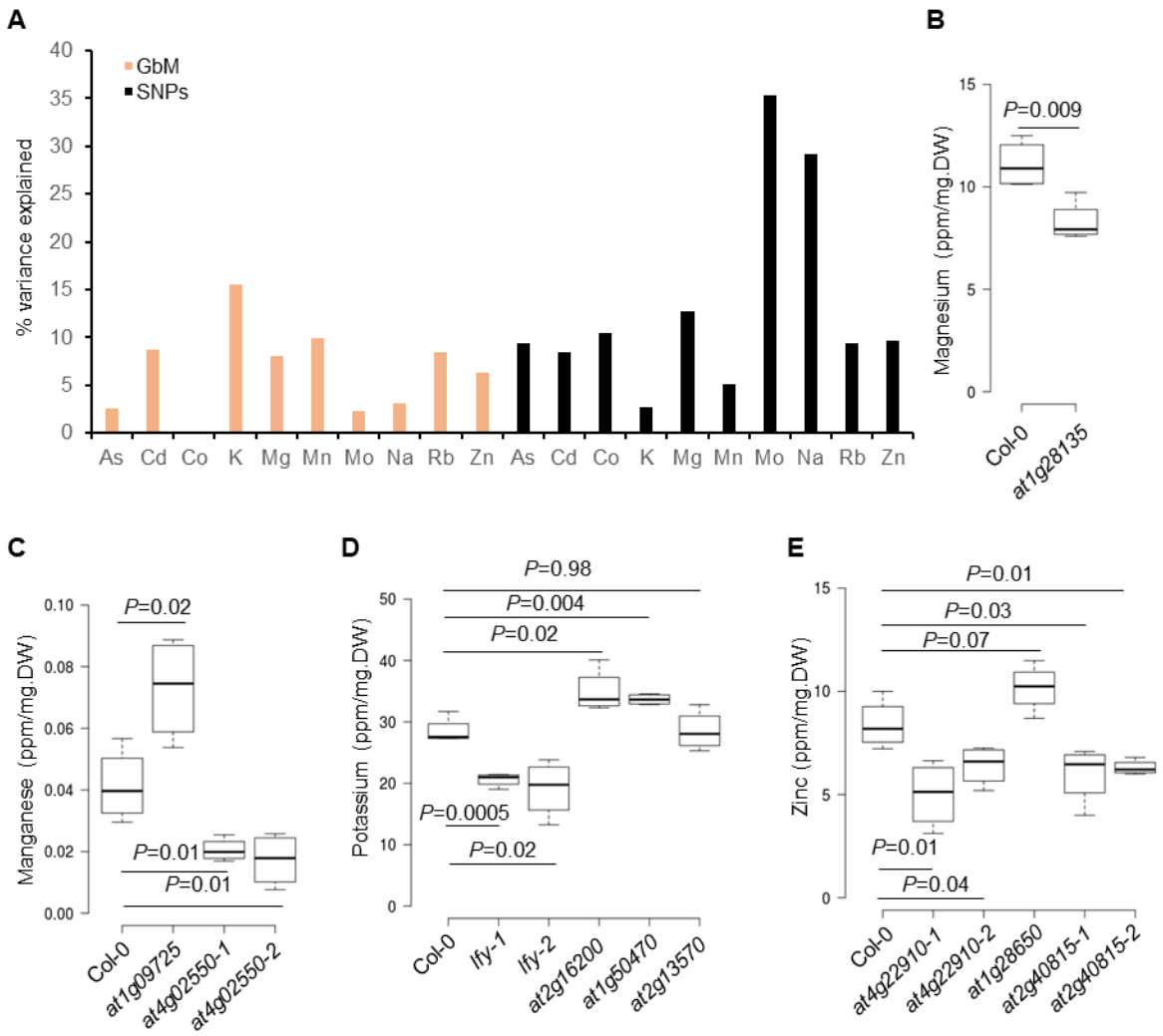

Figure 6. GbM affects mineral accumulation in natural *Arabidopsis* populations. (A) Percent variation in mineral accumulation phenotypes explained by epigenetic (gbM) and genetic (SNPs) QTLs. **(B-E)** Magnesium (**B**), manganese (**C**), potassium (**D**), and zinc (**E**) levels of Col-0 and mutants of indicated genes. Number of biological replicates is 4. *P*, two-tailed t test.

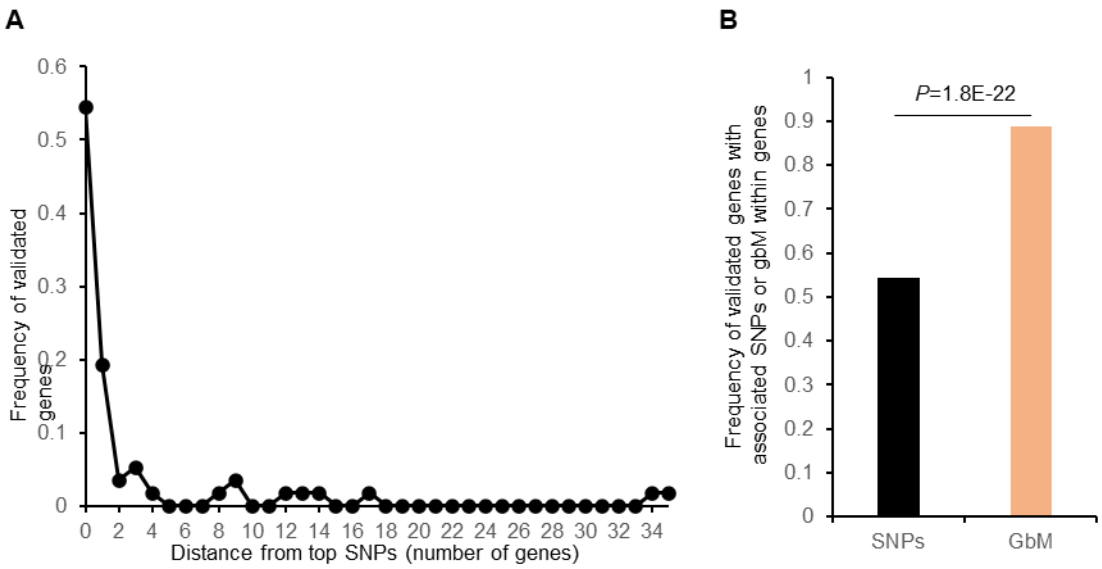

Figure 7. EpiGWA directly identifies causal genes more frequently than GWA. (**A**) Distribution of validated gene distances (in number of genes) from the SNP with the lowest association *P* value in GWA analysis for flowering time, ionome, root growth and development, plant regeneration, stress tolerance, seed longevity, metabolic, and DNA methylation related traits. (**B**) Frequency of SNPs or gbM variants within validated genes. For SNP-based GWA, this encompasses the intergenic region from the end of the upstream gene to the beginning of the downstream gene.

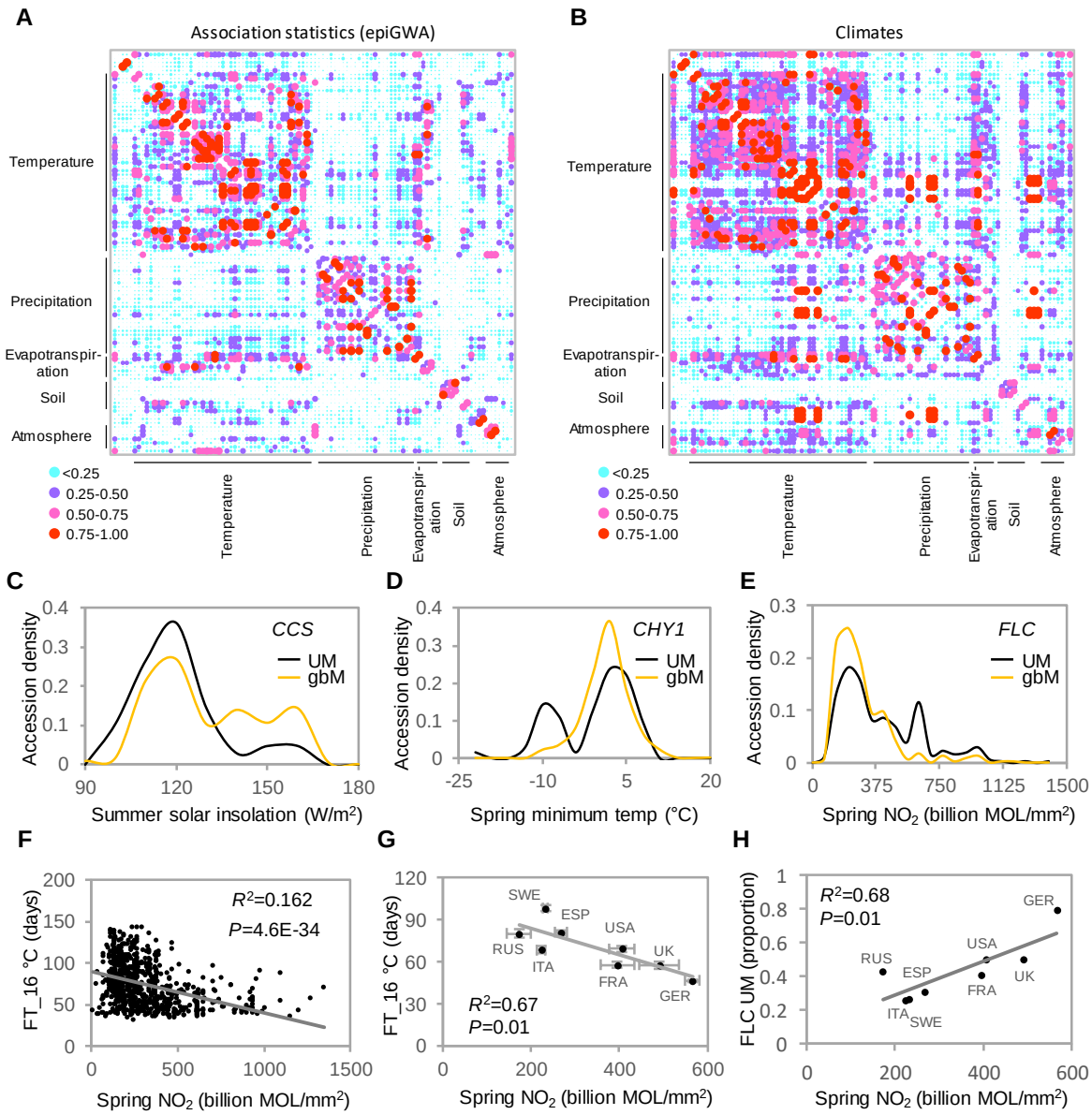

Figure 8. GbM variation is associated with geoclimatic variables. (**A** and **B**) Correlation (R^2) matrices of epiGWA *P* values (**A**) and environmental variables (**B**) for 57 genes identified in at least three epiGWA analyses. See Tables S17 and S18 for individual environment labels in order. (**C**-**E**) Associations between gbM and environmental data for *CCS* (**C**), *CHY1* (**D**), and *FLC* (**E**). (**F**) Correlation between springtime atmospheric NO_2 and flowering time (FT_ 16°C) of individual accessions. (**G**) Average (± standard error) FT_ 16°C and NO_2 concentrations in Sweden (SWE), Russia (RUS), Italy (ITA), Spain (ESP), United States of America (USA), France (FRA), United Kingdom (UK), and Germany (GER). (**H**) Prevalence of *FLC* UM epiallele as a function of NO_2.