Epigenetic inheritance mediates phenotypic diversity in natural populations

Zaigham Shahzad[1], Jonathan D. Moore[1], Jaemyung Choi[1] and Daniel Zilberman[1,2,*]

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

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

*Correspondence: daniel.zilberman@ist.ac.at

Abstract

Genetic variation is regarded as a prerequisite for evolution. Theoretical models suggest epigenetic information inherited independently of DNA sequence can also enable evolution. However, whether epigenetic inheritance mediates phenotypic evolution in natural populations is unknown. Here we show that natural epigenetic DNA methylation variation in gene bodies regulates genes expression, and thereby influences the natural variation of complex traits in *Arabidopsis thaliana*. Notably, the effects of methylation variation on phenotypic diversity and gene expression variance are comparable with those of DNA sequence polymorphism. We also identify methylation epialleles in numerous genes associated with environmental conditions in native habitats, suggesting that intragenic methylation facilitates adaptation to fluctuating environments. Our results demonstrate that methylation variation fundamentally shapes phenotypic diversity in natural populations and provides an epigenetic basis for adaptive Darwinian evolution independent of genetic polymorphism.

Introduction

The neo-Darwinian or Modern synthesis is at the heart of evolutionary biology (Bowler, 2005). This postulates 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 (Charlesworth et al., 2017; Liu, 2018). Epigenetic information, which can be encoded independently of the DNA sequence, is essential for cell fate determination, development, and environmental responses in eukaryotes (Birnbaum and Roudier, 2017; Elsherbiny and Dobreva, 2021; Greenberg and Bourc'his, 2019; Zhang et al., 2018). In theory, if stably heritable and independent of DNA sequence polymorphism, epigenetic variation could influence inherited traits and contribute to adaptation (Gómez-Schiavon and Buchler, 2019; Kronholm and Collins, 2016; Slatkin, 2009). 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 (Quadrana and Colot, 2016; Shahzad et al., 2020). However, these epialleles are either too unstable for the epigenetic states to be subject to natural selection (Gómez-Schiavon and Buchler, 2019; Kronholm and Collins, 2016; Slatkin, 2009) (such as *Cyc* (Cubas et al., 1999), *D1* (Miura et al., 2009) and *MANTLED* (Ong-Abdullah et al., 2015)), are confounded by *trans* genetic variation (such as *PAI2* (Schmitz et al., 2013) and *IAA7* (Shahzad et al., 2020)), are artificially induced (such as *FWA* (Soppe et al., 2000) and *MANTLED* (Ong-Abdullah et al., 2015)), or evidence is lacking that heritable epiallelic variation at the gene occurs in nature (such as *CNR* (Manning et al., 2006), *VTE3* (Quadrana et al., 2014), *Spm* (Fedoroff et al., 1995) and *D1* (Miura et al., 2009)), 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 in any natural population (Quadrana and Colot, 2016). The addition of epigenetic inheritance to the Modern synthesis would represent a major expansion of evolutionary theory.

DNA methylation can be epigenetically inherited over many generations (Heard and Martienssen, 2014; Quadrana and Colot, 2016) and occurs in transposable elements (TEs) and bodies of transcribed genes (Bewick and Schmitz, 2017; Feng et al., 2010; Jones, 2012; Zemach et al., 2010). Plant TEs are generally methylated in all sequence contexts – CG, CHG, and CHH (H being A, T, or C) (Feng et al., 2010; Law and Jacobsen, 2010; Zemach et al., 2010; Zhang et al., 2018). TE methylation induces silencing (Law and Jacobsen, 2010), confers

genome stability (Du et al., 2015; Jones, 2012), can influence the expression of neighboring genes (Hollister and Gaut, 2009; Hollister et al., 2011; Qiu and Köhler, 2020; Quadrana et al., 2016; Shahzad et al., 2020; Stuart et al., 2016), and its variation has been associated with all known epialleles (Quadrana and Colot, 2016). Gene body methylation (gbM) occurs only in the CG context (Feng et al., 2010; Yan et al., 2015; Zemach et al., 2010), although genes can also feature TE-like methylation in all contexts (teM) (Kawakatsu et al., 2016; Schmitz et al., 2013). TeM is associated with silencing (Bewick and Schmitz, 2017; Kawakatsu et al., 2016; Schmitz et al., 2013), but the function of gbM has been extensively debated. GbM is nearly ubiquitous in flowering plants (Niederhuth et al., 2016; Takuno et al., 2016) and is common in animals (Feng et al., 2010; Zemach et al., 2010; Zilberman, 2017). In both groups, gbM preferentially resides in nucleosome-wrapped DNA within the exons of conserved, constitutively transcribed genes (Bewick and Schmitz, 2017; Bräutigam and Cronk, 2018; Dixon et al., 2016; Lewis et al., 2020; Lyons and Zilberman, 2017). Conservation and phenomenological coherence suggest important functions (Zilberman, 2017). Indeed, gbM is associated with (small) gene expression differences within and between plant species (Dubin et al., 2015; Meng et al., 2016; Muyle et al., 2021; Schmitz et al., 2013; Seymour and Gaut, 2020; Takuno et al., 2017), represses aberrant intragenic transcripts (Choi et al., 2020), and appears to be under natural selection (Muyle et al., 2021; Seymour and Gaut, 2020; Takuno and Gaut, 2013). Moreover, loss of methyltransferase function causes developmental abnormalities in honeybees (Kucharski et al., 2008), animals in which methylation is principally restricted to gene bodies (Harris et al., 2019). However, loss of gbM has not been linked to changes in gene expression in plants or animals (Bewick and Schmitz, 2017; Bewick et al., 2016, 2019), leading to the proposals that gbM is a non-functional and somewhat deleterious by-product of TE methylation (in plants) (Bewick and Schmitz, 2017; Bewick et al., 2016; Zhang et al., 2020) or has functions unrelated to gene expression (in animals) (Bewick et al., 2019). Thus, the functional and evolutionary significance of gbM has been mysterious and controversial.

Recent studies reported extensive natural variation in TE methylation, gbM and teM in *Arabidopsis* (Kawakatsu et al., 2016; Schmitz et al., 2013). Methylation levels of natural accessions are associated with climate (Kawakatsu et al., 2016), suggesting that methylation variation could be of adaptive significance. Furthermore, genetically induced methylation polymorphism can account for the inheritance of complex *Arabidopsis* traits (Cortijo et al., 2014; Furci et al., 2019; Johannes et al., 2009), and methylation changes have been linked to

adaptation under artificial selection (Schmid et al., 2018). Due to its stable epigenetic inheritance (Catoni et al., 2017), 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 (Quadrana and Colot, 2016).

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 (Kawakatsu et al., 2016). GbM conservation is bimodal, with genes most commonly methylated either in over 90% of accessions, or in ≤10% (Fig. 1A). The high frequency of gbM conservation supports the hypothesis that gbM is maintained by natural selection (Muyle et al., 2021; Takuno and Gaut, 2013; Zilberman, 2017). In contrast, the vast majority of genes exhibit teM in ≤10% of accessions (Fig. 1B), suggesting that teM is disfavored in most genes, likely due to its deleterious effects on expression (Kawakatsu et al., 2016). A small fraction (7.3%) of teM genes is methylated in >50% of accessions (Fig. 1B), suggesting that teM effects, potentially including transcriptional silencing (Baduel et al., 2021; Quadrana et al., 2016), might be advantageous in some genes and/or environments.

Numbers of teM and gbM genes are very weakly (negatively) correlated across accessions (Fig. 1C) and are similarly weakly (positively) correlated under more restrictive published definitions (Zhang et al., 2020) of gbM and teM (fig. S1A). Additionally, intragenic teM prevalence across accessions tends to increase with decreasing gbM conservation under either gbM/teM definition (Fig. 1D and fig. S1, B-F), 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 (Bewick et al., 2016) or that gbM promotes the transition to teM (Zhang et al., 2020). Instead, intragenic gbM and teM appear largely independent, which is consistent with many linages having only teM (fungi and some land plants) or only gbM (many invertebrates) (Bewick and Schmitz, 2017; Feng et al., 2010; Zemach et al., 2010).

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 (Kawakatsu et al., 2016) 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 (Fig. 1E, fig. 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* (Fig. S3) (Muyle et al., 2021), 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 (Fig. 1F and fig. S4A). Non-associated gbM genes (NA^{gbM}) and eQTL^{gbM} genes have similar CG dinucleotide composition and length (fig. S5, A and B) – key distinguishing characteristics of gbM genes (Zhang et al., 2020) – indicating that eQTL^{gbM} genes are fairly typical. However, +eQTL^{gbM} genes show lower expression than NA^{gbM} genes (fig. 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 (Dubin et al., 2015; Meng et al., 2016; Schmitz et al., 2013).

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} (table S1). TeM effects on gene expression are stronger than for gbM, as teM explains >20% of expression variance in 46% of eQTL^{teM} Bonferroni genes (Fig. 1F and fig. S4B). The primarily negative associations between teM and expression are consistent

with existing knowledge (Kawakatsu et al., 2016; Schmitz et al., 2013) 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 (fig. 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%) (Fig. 1G and table S3). Similarly, nearly all (98.5%) $-eQTL^{teM}$ were retained (Fig. 1H and table S3). However, 39.7% of $-eQTL^{gbM}$ and 33.3% of $+eQTL^{teM}$ were lost after accounting for SNP variation (Fig. 1, G and H 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* (Kerdaffrec et al., 2016; Lopez-Arboleda et al., 2021; Shahzad et al., 2016; Tang et al., 2018; Zhang et al., 2011). 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 (Fig. 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%; fig. S7, B and C and table S5). Therefore, our results show that intragenic methylation is quantitively 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 SVs (Baduel et al., 2021) within retained epigenetic eQTL genes and 4 kb upstream and downstream. We found that 98.2% of +eQTL^{gbM} are not confounded by SV, either because an eQTL^{SV} is not detected (85.4%), or because a +eQTL^{gbM} was detected in the population without SV (12.8%) (Fig. 1I). Similarly, all -eQTL^{gbM} are retained after accounting for SV (Fig. 1I). These results show that eQTL^{gbM} are very rarely confounded by SV, and are consistent with the findings that gbM variation develops independently of SV in mutation accumulation (MA) lines (Becker et al., 2011; Schmitz et al., 2011), gbM variation in MA lines is highly correlated with that in natural accessions (Vidalis et al., 2016), and gbM is generally neither associated with nor affected by SV in accessions (Stuart et al., 2016).

In comparison to eQTL^{gbM}, a higher fraction of eQTL^{teM} (10% +eQTL^{teM} and 4% -eQTL^{teM}) are lost after accounting for SV (Fig. IJ). More frequent confounding of eQTL^{teM} by SV is expected, because teM variation is known to be associated with TE SVs (Quadrana et al., 2016; Schmitz et al., 2013; Stuart et al., 2016), which may promote teM and repress expression (or *vice versa*).

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 (Dubin et al., 2015; Kawakatsu et al., 2016; Sasaki et al., 2019; Schmitz et al., 2013; Shen et al., 2014), and consistently we find 47.4% of retained eQTL^{teM} genes are significantly affected by *trans* QTLs (Fig. 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 (Fig 1L and fig. 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. Analyses of broadly distributed *Arabidopsis* accessions did not identify strong *trans* influences on gbM (Kawakatsu

et al., 2016; Schmitz et al., 2013), whereas a study of Swedish *Arabidopsis* did find strong *trans* effects for around 1,300 gbM genes (Dubin et al., 2015). 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; fig. S8, B-D). 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} (Fig. 1, K and L and fig. S8, A-D). 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 (fig. S8, B-D). This indicates that estimates of *trans* effects on gbM variation are inflated in analyses of small populations, a phenomenon known as the Beavis effect (Beavis, 1998; Xu, 2003). Therefore, *trans* genetic polymorphism explains only a very minor fraction of overall gbM variation. Furthermore, global gbM levels of accessions are weakly associated with gbM levels of individual genes (R^2<0.1 for ~80% genes, linear model; fig. S8E), suggesting that gbM at individual genes tends to evolve independently of *trans* factors that may cause global gbM differences, such as the high overall gbM in *Arabidopsis* accessions from northern Sweden (Kawakatsu et al., 2016). Taken together, our results demonstrate that gbM variation is not strongly affected by *trans* (or *cis*) genetic polymorphism, but is instead primarily epigenetic, as has been previously suggested (Schmitz et al., 2013).

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 (Meng et al., 2016). 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 (Fig. 1, 1-L), 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 (Fig. 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 (Fig. 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 (fig. 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 (Fig. 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 (Fig. 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 (Fig. 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 (Fig. 2C), which is comparable to the effects of gbM on expression variance at retained eQTL^{gbM} (10.8% for +eQTL^{gbM}; 8.7% for -eQTL^{gbM}; Fig. 2A). Analysis of control genes in UM accessions produces a significant increase in expression variance explained by *cis* SNPs (Fig. 2C), which is expected due to the smaller population sizes (Beavis, 1998; Xu, 2003). An analogous (non-significant) increase in expression variance explained by *cis* SNPs is observed in UM populations for retained eQTL^{teM} (fig. 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 (Fig. 2C), indicating that these eQTLs^{SNP} are overwhelmingly detected due to linkage of SNPs with gbM. Importantly, our 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.

MET1 inactivation 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 (Choi et al., 2020). Mutation of the *MET1* methyltransferase causes complete loss of gbM and nearly complete loss of mCG throughout the genome (Catoni et al., 2017). 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 (Fig. 2, D and E). Only -eQTL^{teM} and +eQTL^{gbM} genes methylated in Col-0 are significantly affected by *met1* (Fig. 2, D and E).

Col-0 methylated -eQTL^{teM} genes are strongly overexpressed in *met1* (over 6-fold of wild-type; Fig. 2D), consistent with the established repressive activity of teM (Bewick and Schmitz, 2017; Kawakatsu et al., 2016; Schmitz et al., 2013). In contrast, +eQTL^{gbM} genes are modestly downregulated (expressed at about 80% of wild-type; Fig. 2E). Analysis of genes that passed less stringent significance thresholds (0.05 and 0.1 FDR) produced similar results (fig. S10, A and B). Analysis of additional Col-0 *met1* leaf (Choi et al., 2020) and inflorescence (Oberlin et al., 2017) RNA-seq datasets, and of *met1* microarray data in the WS wild-type background (Li et al., 2011a), also produced analogous results (fig. S11). *MET1* inactivation alters non-CG methylation and histone modifications (Deleris et al., 2012), which could influence gene expression, but are not significantly changed in any relevant gbM gene category (fig. S12). The reduced expression of only those +eQTL^{gbM} genes that are methylated in Col-0 or WS and not confounded by genetic variation indicates that gbM promotes gene expression. Furthermore, retained +eQTL^{gbM} genes with higher methylation in parental accessions show stronger downregulation in *met1* (Col-0 and WS; Fig. 2F), consistent with gbM quantitatively promoting expression.

To further evaluate whether gbM loss alters gene expression, we analyzed a plant that is heterozygous for *met1* (*met1*+/-), has relatively normal TE methylation, and limited gbM loss (Choi et al., 2020). This plant also contains loss-of-function mutations in two histone H1 genes (Choi et al., 2020) and therefore expression was analyzed with respect to *h1* mutant controls.

Retained +eQTL^{gbM} genes demethylated in this plant have significantly decreased (~35%) expression in comparison to retained +eQTL^{gbM} genes that maintain gbM (Fig. 2G), specifically linking gbM loss with reduced expression. Moreover, we isolated six *h1met1*+/+ progeny of *h1met1*+/-. These plants exhibit mosaic demethylation of gbM genes, whereas TE methylation is comparatively normal (fig. S13). Retained +eQTL^{gbM} genes demethylated in *h1met1*+/+ plants display significantly reduced (~25%) expression compared to retained +eQTL^{gbM} genes that keep gbM (Fig. 2H). Altogether, we find that loss of gbM at retained +eQTL^{gbM} genes consistently produces decreased expression, regardless of the genetic background (Col-0 or WS), tissues (seedlings, leaves, or inflorescence), presence of functional *MET1*, or the extent of global teM or gbM perturbation. Our results therefore constitute strong evidence that gbM directly promotes gene expression. The sensitivity of +eQTL^{gbM} gene expression to gbM also reinforces the conclusion that detection of these eQTLs is not substantially confounded by linked cryptic genetic polymorphism in natural accessions, because the effects of such polymorphism should be independent of DNA methylation.

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

Modulation of gene expression by natural methylation polymorphism potentiates the exciting possibility that methylation epialleles drive phenotypic variation in natural populations. To 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 (Exposito-Alonso et al., 2019). Linear model associations can be confounded by population structure, producing elevated rates of false-positives (Atwell et al., 2010). Therefore, we used a 'genomic control factor' (Devlin and Roeder, 1999) 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; Fig. 3A and fig. 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 (Verbruggen and Hermans, 2008).

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*; Fig. 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 (Fig. S15A; table S8). *PROT1* is downregulated by 38% in *met1* (Fig. 3C), confirming that gbM promotes *PROT1* expression. *AT1G19410* teM is lost in plants that lack DRM and CMT methyltransferases (Fig. S15B), and in such *ddcc* mutants (Stroud et al., 2014) *AT1G19410* expression increases about 7-fold (Fig. 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} (Fig. S16). Heritability estimates indicate that the two rfQTLs^{SNP} explain 18.9% of the relative fitness variance in MLP environment (Fig. 3E). Strikingly, the two epigenetic rfQTLs also account for a substantial proportion of fitness variance (12.1% after accounting for *trans* QTL effects; Fig. 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 (Fig. 3F). Mutant plants produced less biomass and had decreased survival to fruit, but had the same fecundity (seed set) as wild type (Fig. 3G and fig. S17A). Consistently, *PROT1* gbM is specifically associated with survival in MLP conditions (Fig. S18, A-C). Thus, the effects of *PROT1*

inactivation match those expected from the associations between gbM, gene expression, and relative MLP fitness (Fig. 3B).

Inactivation of *AT1G19410* resulted in a slight (~13%) but non-significant increase in relative fitness under heat and drought stress (Fig. 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; Fig 3, H and I and fig. 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 (Fig. S18, D-F). 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 key plant traits.

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 (Ågren et al., 2017; Li et al., 2010; Taylor et al., 2019). 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 (The 1001 Genomes Consortium, 2016). Due to the structure of flowering phenotypes in *Arabidopsis* populations (Shindo et al., 2007), we could not use a linear model with genomic control to analyze quantitative methylation variation (fig. S19 and table S10). Instead, we used a mixed linear model (MLM) that effectively corrects population structure (Bradbury et al., 2007; Vilhjálmsson and Nordborg, 2013; Weigel and Nordborg, 2015; Zhang et al., 2010) (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 (Grimm et al., 2017). 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 (Fig. 4A and fig. S20). Seven fQTLs^{gbM} contain SVs, but all loci except *AT4G33560* were retained after accounting for SV (fig. S21, A-G). Two fQTLs^{gbM} localized to the well-known flowering regulators *Flowering locus* C (*FLC*) and *FRIGIDA* (*FRI*) (Adrian et al., 2009; Hepworth et al., 2020; Johanson et al., 2000; Li et al., 2014; Shindo et al., 2005; Whittaker and Dean, 2017) (Fig. 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 (fig. S22). Given the extensive analyses of *Arabidopsis* flowering time determinants performed over several decades (Blümel et al., 2015) , 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* (Corbesier et al., 2007; Gregis et al., 2006; Sung and Amasino, 2004) (fig. S23 and table S12). Except for FT_10°C, fQTLs^{SNP} explain <10% of trait heritability (Fig. 4B). 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 (Fig. 4B). Accurate estimation of natural variation contribution to trait heritability is challenging (Atwell et al., 2010; Gallagher and Chen-Plotkin, 2018) and the values in Fig. 4B 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 (Fig. 4B). Instead, the key conclusion is that epigenetic gbM variation and genetic variation explain comparable amounts of flowering trait heritability.

FLC gbM influences *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 (Adrian et al., 2009; Whittaker and Dean, 2017), but we found them associated with all (*FLC*) or nearly all (*FRI*) of the six correlated flowering phenotypes (Fig. 4A). Furthermore, both genes show consistent significant associations between gbM, flowering and gene expression (Fig. 4C), and both were detected as +eQTL^{gbM} at the Bonferroni threshold (table S2). Importantly, the *FRI* +eQTL^{gbM} was lost after accounting for genetic (SNP) variation (table S3), *FRI* expression is not significantly altered in *met1* plants (fig. S24A), and we did not detect significant associations

between *FRI* epiallelic states and FT_16°C or expression within nested genetic populations (fig. S24, B and C). Furthermore, we found significant linkage disequilibrium (Flint-Garcia et al., 2003) (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 (Fig. 4B).

In contrast to *FRI*, the *FLC* +eQTL^{gbM} was retained after accounting for genetic variation (table S3 and S5; expression in *met1* could not be tested because *FLC* lacks gbM in Col-0, fig. S25A), 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 (Fig. 4, D and E and fig. S21, G and H). Because multiple *FLC* alleles affect flowering time or vernalization response (Hepworth et al., 2020; Li et al., 2014; Quadrana et al., 2016), we defined thirteen *FLC* haplotypes (table S8), all of which contain gbM accessions (Fig. 4F), 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; Fig. 4F), and significantly higher *FLC* expression in three of these haplotypes (Fig. 4G). These findings demonstrate that gbM variation influences *FLC* expression and flowering time independently of *FLC* SNP and SV variation. Moreover, gbM at *FLC* is not associated with global gbM levels (Fig. 4H) and is not influenced by *trans* QTLs in the worldwide population (fig. S25B), 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 (Li et al., 2014). 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 (Fig. 4I and fig. S25, C and D). 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 (Fig. 4J). The *FLC* fQTL^{gbM} is thus not confounded by links between flowering time and latitude, variation of global gbM (Fig. 4H), *FLC* genetic polymorphism (Fig. 4, D-G and fig. S21, G and H), or *trans* genetic polymorphism (fig. S25B).

GbM of a novel locus (*LGM*) affects the transition to flowering

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 (Li et al., 2010). EpiGWA identified significant fQTLs^{gbM} (FDR 0.05) associated with flowering time in simulated local climates within two uncharacterized genes, *AT4G15750* and *AT5G49440* (Fig. 5A, fig. S26, and table S13). The *AT4G15750* fQTL^{gbM} shows sensitivity to seasonal planting in Sweden and to climate (Spain vs. Sweden) in summer (fig. S27, A-D), whereas the *AT5G49440* fQTL^{gbM} is sensitive to seasons in Spain and to climate in spring (Fig. 5, B and C and fig. S27, E and F). 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% (Fig. 5D). Flowering (FT_16°C) of gbM accessions is severely delayed (by 29 days on average) compared to teM, UM or deleted accessions (Fig. 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 (Fig. 5E). Col-0 is a teM accession, and *LGM* is strongly downregulated in *met1* (which partially reduces teM; Fig. 5F and fig. S28A), 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 (van Nocker and Ludwig, 2003). 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 (fig. S28B).

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

(Fig. 5G). Therefore, gbM (but not teM) at *LGM* delays 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 (Stolc et al., 2005). 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 (Fig. 5H and fig. S28C). TeM extends further 5´ than gbM and covers the apparent transcriptional start site (TSS) of the long sense *LGM* mRNA (Fig. 5H), which would be expected to silence the TSS (Bewick and Schmitz, 2017). Our results indicate 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 (Fig. 5I), and gbM at *LGM* is not associated with *trans* QTLs (fig. S28D), indicating that gbM at *LGM* is evolving epigenetically. GbM accessions mainly grow in southern Sweden, around 55° latitude, where UM accessions are rare (Fig. 5J and fig. S28, E and F). EpiGWA analysis using only southern Swedish accessions found a significant association between gbM and flowering time (Fig. 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* (Fig. 4, H-J). Considering the much greater relative flowering delay of gbM accessions grown in Sweden compared to Spain (Fig. 5C), the prevalence of *LGM* gbM in southern Sweden (Fig. 5J) suggests that gbM is under selection to delay flowering in this environment.

GbM variation may facilitate local adaptation

Arabidopsis grows in a broad range of natural environments and shows extensive local adaptation (Ågren and Schemske, 2012). Sensitivities of the *AT4G15750* and *LGM* fQTLs^{gbM} to climate (Fig. 5, B and C and fig. S27) suggest that gbM may facilitate adaptation to the local environment. To explore this possibility, we performed MLM epiGWA analyses using intragenic epigenetic states (UM and gbM) for 171 environmental variables (Ferrero-Serrano and Assmann, 2019) (latitude, longitude, temperature, precipitation, atmospheric composition, and soil properties; table S15). We detected 571 significant associations at FDR 0.05 between 232 genes and 115 of these variables (table S15). SNP-based GWA analyses showed that 76.9% of epigenetic associations do not colocalize with genetic associations (Fig. S29A; table S15). GbM variation in 57 genes is associated with at least three environments (table S15).

EpiGWA *P* values for these genes are strongly correlated for associated environments (Fig. 6, A and B and tables S16 and S17), 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 (Fig. 6, C-E and fig. S29, B-D) that do not overlap genetic associations (table S15). GbM in *CCS*, which mediates heat stress responses (Guan et al., 2013), is associated with summer solar insolation, with gbM epialleles much more common in high insolation environments (Fig. 6C). GbM in *CHY1*, which is involved in cold signaling and promotes freezing tolerance (Dong et al., 2009), is associated with spring minimum temperature, with gbM epialleles rare in environments where temperature drops below -4°C (Fig. 6D). GbM in *HUP9*, a regulator of flooding stress response (Lee et al., 2011), is associated with annual precipitation (Fig. S29B). *PYR1* gbM variation is associated with soil excess salts, with high salt soils almost exclusively featuring gbM epialleles (Fig. S29C). *PYR1* is an abscisic acid (ABA) receptor (Cutler et al., 2010), and ABA is a central regulator of plant salt stress responses (Yu et al., 2020). GbM variation in the calcium sensor *SOS3* (Sánchez-Barrena et al., 2005) associates with soil salinity and sodicity (Fig. 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 (Fig. 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 (Lelieveld et al., 2015), with UM *FLC* alleles prevalent in high NO_2 environments (Fig. 6E and table S18). Because UM *FLC* accessions flower early (Fig. 4, D-G), 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 (Fig. 6F and table S19). NO_2 concentrations vary regionally, an indicator of air quality over urban and industrial centers (Larkin et al., 2017). 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 (Fig. 6G), indicating 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; Fig. 6H), and enrichment of the UM

allele at 48-52° latitude (Fig. 4I) corresponds to Germany, which has especially high NO_2 (Fig. 6, G and H). 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, and that its natural variation influences two important traits: fitness in a hot and dry climate and flowering time. Strikingly, epigenetically inherited intragenic methylation accounts for about as much gene expression and trait variance as genetic polymorphism (Fig. 2, A and C, 3E and 4B). Except at *ANH*, effects on tested traits are mediated entirely by gbM. Thus, although studies of epigenetic trait inheritance have focused exclusively on TE methylation or teM (Baduel et al., 2021; Heard and Martienssen, 2014; Quadrana and Colot, 2016; Quadrana et al., 2016; Slotkin and Martienssen, 2007), gbM variation appears 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 new functional genes (*PROT1*, *ANH*, *LGM, AT1G26795, AT1G51820, AT2G18210, AT3G09530*) and new phenotypic associations with the environment (flowering time and NO_2). We therefore expect epiGWA analyses 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 (Fujita et al., 2013; Li et al., 2011b; McCouch et al., 2016; Yano et al., 2019), suggesting that gbM polymorphism could prove useful for crop improvement.

DNA methylation is mutagenic (Robertson and A.Jones, 2000), and its presence in coding sequences likely incurs a fitness cost (Zilberman, 2017). 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. GbM epimutation rates are about 10^5 times greater than genetic mutation rates (Becker et al., 2011; Schmitz et al., 2011; Van der Graaf et al., 2015). The direct relationship between gbM and mRNA levels (Fig. 1E, 2F and fig. S2) also indicates that gbM changes should quantitatively alter gene expression, whereas most mutations are expected to be neutral or detrimental to gene function (Eyre-Walker and Keightley, 2007). 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. At long timescales, such as those that typically separate species, genetic differences should dominate. However, our results indicate that within plant species, 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 (Fig. 6, E-H) 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 (Fig. 6, F and G) suggests that accelerated flowering is adaptive under elevated NO_2. Natural genetic variation at *FLC* is a major determinant of flowering time (Hepworth et al., 2020; Li et al., 2014; The 1001 Genomes Consortium, 2016), 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 (Ferrero-Serrano and Assmann, 2019). The majority of atmospheric NO_2 (>75%) is produced by recent human activities, especially the burning of fossil fuel (Lelieveld et al., 2015). 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 (Fig. 6, E-H), but not other environmental variables (table S18), which is consistent with our observation that *FLC* gbM varies independently of DNA sequence (Fig. 4F). 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 published evolutionary models (Gómez-Schiavon and Buchler, 2019; Kronholm and Collins, 2016; Quadrana and Colot, 2016; Slatkin, 2009) and is likely 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 (Luo et al., 2018). Furthermore, methylation is extensively reprogrammed during mammalian embryogenesis and germline development

(Zeng and Chen, 2019), which precludes methylation inheritance at most loci (Horsthemke, 2018; Kazachenka et al., 2018; Nadeau, 2015). However, many invertebrates, from basal cnidarians to chordates, display gbM patterns highly similar to those of plants (Bewick and Schmitz, 2017; Feng et al., 2010; Zemach et al., 2010; Zilberman, 2017). 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 (Wang et al., 2016), as would be expected if methylation is epigenetically inherited. Moreover, methylation fidelity is high in honeybee sperm compared to somatic tissues (Harris et al., 2019), and methylation patterns can be stably transmitted from fathers to daughters (Yagound et al., 2020). These features of invertebrate methylation suggest that epigenetically heritable gbM facilitates environmental adaptation in animals. This may be especially important for species like aphids (Mathers et al., 2019), 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 information: This file contains Materials and Methods and fig. S1-S29.

Supplementary Tables: This file contains tables S1-S22.

References

Adrian, J., Torti, S., and Turck, F. (2009). From decision to commitment: the molecular memory of flowering. Mol. Plant *2*, 628–642.

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.

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.

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.

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.

Beavis, W.D. (1998). QTL Analyses: Power, Precision, and Accuracy. In Molecular Dissection of Complex Traits, (CRC Press), pp. 145–162.

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.

Bewick, A.J., and Schmitz, R.J. (2017). Gene body DNA methylation in plants. Curr. Opin. Plant Biol. *36*, 103–110.

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.

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.

Birnbaum, K.D., and Roudier, F. (2017). Epigenetic memory and cell fate reprogramming in plants. Regeneration *4*, 15–20.

Blümel, M., Dally, N., and Jung, C. (2015). Flowering time regulation in crops—what did we learn from Arabidopsis? Curr. Opin. Biotechnol. *32*, 121–129.

Bowler, P.J. (2005). CHAPTER 2 - Variation from Darwin to the Modern Synthesis. In Variation, B. Hallgrímsson, and B.K. Hall, eds. (Burlington: Academic Press), pp. 9–27.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics *23*, 2633–2635.

Bräutigam, K., and Cronk, Q. (2018). DNA methylation and the evolution of developmental complexity in plants. Front. Plant Sci. *9*.

684 Catoni, M., Griffiths, J., Becker, C., Zabet, N.R., Bayon, C., Dapp, M., Lieberman-Lazarovich, M.,
685 Weigel, D., and Paszkowski, J. (2017). DNA sequence properties that predict susceptibility to
686 epiallelic switching. EMBO J. *36*, 617–628.

687 Charlesworth, D., Barton, N.H., and Charlesworth, B. (2017). The sources of adaptive variation. Proc.
688 R. Soc. B Biol. Sci. *284*, 20162864.

689 Choi, J., Lyons, D.B., Kim, M.Y., Moore, J.D., and Zilberman, D. (2020). DNA methylation and Histone
690 H1 jointly repress transposable elements and aberrant intragenic transcripts. Mol. Cell *77*, 310-
691 323.e7.

692 Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L.,
693 Turnbull, C., et al. (2007). FT protein movement contributes to long-distance signaling in floral
694 induction of Arabidopsis. Science *316*, 1030–1033.

695 Cortijo, S., Wardenaar, R., Colomé-Tatché, M., Gilly, A., Etcheverry, M., Labadie, K., Caillieux, E.,
696 Hospital, F., Aury, J.-M., Wincker, P., et al. (2014). Mapping the epigenetic basis of complex traits.
697 Science *343*, 1145–1148.

698 Cubas, P., Vincent, C., and Coen, E. (1999). An epigenetic mutation responsible for natural variation
699 in floral symmetry. Nature *401*, 157–161.

700 Cutler, S.R., Rodriguez, P.L., Finkelstein, R.R., and Abrams, S.R. (2010). Abscisic acid: emergence of a
701 core signaling network. Annu. Rev. Plant Biol. *61*, 651–679.

702 Deleris, A., Stroud, H., Bernatavichute, Y., Johnson, E., Klein, G., Schubert, D., and Jacobsen, S.E.
703 (2012). Loss of the DNA Methyltransferase MET1 Induces H3K9 Hypermethylation at PcG Target
704 Genes and Redistribution of H3K27 Trimethylation to Transposons in Arabidopsis thaliana. PLOS
705 Genet. *8*, e1003062.

706 Devlin, B., and Roeder, K. (1999). Genomic control for association studies. Biometrics *55*, 997–1004.

707 Dixon, G.B., Bay, L.K., and Matz, M.V. (2016). Evolutionary consequences of DNA methylation in a
708 basal metazoan. Mol. Biol. Evol. *33*, 2285–2293.

709 Dong, C.-H., Zolman, B.K., Bartel, B., Lee, B., Stevenson, B., Agarwal, M., and Zhu, J.-K. (2009).
710 Disruption of Arabidopsis CHY1 reveals an important role of metabolic status in plant cold stress
711 signaling. Mol. Plant *2*, 59–72.

712 Du, J., Johnson, L.M., Jacobsen, S.E., and Patel, D.J. (2015). DNA methylation pathways and their
713 crosstalk with histone methylation. Nat. Rev. Mol. Cell Biol. *16*, 519–532.

714 Dubin, M.J., Zhang, P., Meng, D., Remigereau, M.-S., Osborne, E.J., Paolo Casale, F., Drewe, P.,
715 Kahles, A., Jean, G., Vilhjálmsson, B., et al. (2015). DNA methylation in Arabidopsis has a genetic
716 basis and shows evidence of local adaptation. ELife *4*, e05255.

717 Elsherbiny, A., and Dobreva, G. (2021). Epigenetic memory of cell fate commitment. Curr. Opin. Cell
718 Biol. *69*, 80–87.

719 Exposito-Alonso, M., Burbano, H.A., Bossdorf, O., Nielsen, R., and Weigel, D. (2019). Natural
720 selection on the Arabidopsis thaliana genome in present and future climates. Nature *573*, 126–129.

721 Eyre-Walker, A., and Keightley, P.D. (2007). The distribution of fitness effects of new mutations. Nat.
722 Rev. Genet. *8*, 610–618.

723 Fedoroff, N., Schläppi, M., and Raina, R. (1995). Epigenetic regulation of the maize Spm transposon.
724 BioEssays *17*, 291–297.

725 Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H.,
726 Halpern, M.E., et al. (2010). Conservation and divergence of methylation patterning in plants and
727 animals. Proc. Natl. Acad. Sci. *107*, 8689–8694.

728 Ferrero-Serrano, Á., and Assmann, S.M. (2019). Phenotypic and genome-wide association with the
729 local environment of Arabidopsis. Nat. Ecol. Evol. *3*, 274–285.

730 Flint-Garcia, S.A., Thornsberry, J.M., and Buckler, E.S. (2003). Structure of linkage disequilibrium in
731 plants. Annu. Rev. Plant Biol. *54*, 357–374.

732 Fujita, D., Trijatmiko, K.R., Tagle, A.G., Sapasap, M.V., Koide, Y., Sasaki, K., Tsakirpaloglou, N.,
733 Gannaban, R.B., Nishimura, T., Yanagihara, S., et al. (2013). NAL1 allele from a rice landrace greatly
734 increases yield in modern indica cultivars. Proc. Natl. Acad. Sci. *110*, 20431–20436.

735 Furci, L., Jain, R., Stassen, J., Berkowitz, O., Whelan, J., Roquis, D., Baillet, V., Colot, V., Johannes, F.,
736 and Ton, J. (2019). Identification and characterisation of hypomethylated DNA loci controlling
737 quantitative resistance in Arabidopsis. ELife *8*, e40655.

738 Gallagher, M.D., and Chen-Plotkin, A.S. (2018). The post-GWAS era: from association to function.
739 Am. J. Hum. Genet. *102*, 717–730.

740 Gómez-Schiavon, M., and Buchler, N.E. (2019). Epigenetic switching as a strategy for quick
741 adaptation while attenuating biochemical noise. PLOS Comput. Biol. *15*, e1007364.

742 Greenberg, M.V.C., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian
743 development and disease. Nat. Rev. Mol. Cell Biol. *20*, 590–607.

744 Gregis, V., Sessa, A., Colombo, L., and Kater, M.M. (2006). AGL24, SHORT VEGETATIVE PHASE, and
745 APETALA1 redundantly control AGAMOUS during early stages of flower development in Arabidopsis.
746 Plant Cell *18*, 1373–1382.

747 Grimm, D.G., Roqueiro, D., Salomé, P.A., Kleeberger, S., Greshake, B., Zhu, W., Liu, C., Lippert, C.,
748 Stegle, O., Schölkopf, B., et al. (2017). EasyGWAS: a cloud-based platform for comparing the results
749 of genome-wide association studies. Plant Cell *29*, 5–19.

750 Guan, Q., Lu, X., Zeng, H., Zhang, Y., and Zhu, J. (2013). Heat stress induction of miR398 triggers a
751 regulatory loop that is critical for thermotolerance in Arabidopsis. Plant J. *74*, 840–851.

752 Harris, K.D., Lloyd, J.P.B., Domb, K., Zilberman, D., and Zemach, A. (2019). DNA methylation is
753 maintained with high fidelity in the honey bee germline and exhibits global non-functional
754 fluctuations during somatic development. Epigenetics Chromatin *12*, 62.

755 Heard, E., and Martienssen, R.A. (2014). Transgenerational epigenetic inheritance: myths and
756 mechanisms. Cell *157*, 95–109.

757 Hepworth, J., Antoniou-Kourounioti, R.L., Berggren, K., Selga, C., Tudor, E.H., Yates, B., Cox, D.,
758 Collier Harris, B.R., Irwin, J.A., Howard, M., et al. (2020). Natural variation in autumn expression is
759 the major adaptive determinant distinguishing Arabidopsis FLC haplotypes. ELife *9*, e57671.

760 Hollister, J.D., and Gaut, B.S. (2009). Epigenetic silencing of transposable elements: A trade-off
761 between reduced transposition and deleterious effects on neighboring gene expression. Genome
762 Res. *19*, 1419–1428.

763 Hollister, J.D., Smith, L.M., Guo, Y.-L., Ott, F., Weigel, D., and Gaut, B.S. (2011). Transposable
764 elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana
765 and Arabidopsis lyrata. Proc. Natl. Acad. Sci. *108*, 2322–2327.

766 Horsthemke, B. (2018). A critical view on transgenerational epigenetic inheritance in humans. Nat.
767 Commun. *9*, 2973.

768 Johannes, F., Porcher, E., Teixeira, F.K., Saliba-Colombani, V., Simon, M., Agier, N., Bulski, A.,
769 Albuisson, J., Heredia, F., Audigier, P., et al. (2009). Assessing the impact of transgenerational
770 epigenetic variation on complex traits. PLOS Genet. *5*, e1000530.

771 Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., and Dean, C. (2000). Molecular analysis of
772 FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. Science *290*, 344–
773 347.

774 Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat.
775 Rev. Genet. *13*, 484–492.

776 Kawakatsu, T., Huang, S.C., Jupe, F., Sasaki, E., Schmitz, R.J., Urich, M.A., Castanon, R., Nery, J.R.,
777 Barragan, C., He, Y., et al. (2016). Epigenomic diversity in a global collection of Arabidopsis thaliana
778 accessions. Cell *166*, 492–505.

779 Kazachenka, A., Bertozzi, T.M., Sjoberg-Herrera, M.K., Walker, N., Gardner, J., Gunning, R., Pahita, E.,
780 Adams, S., Adams, D., and Ferguson-Smith, A.C. (2018). Identification, characterization, and
781 heritability of murine metastable epialleles: implications for non-genetic inheritance. Cell *175*, 1259-
782 1271.e13.

783 Kerdaffrec, E., Filiault, D.L., Korte, A., Sasaki, E., Nizhynska, V., Seren, Ü., and Nordborg, M. (2016).
784 Multiple alleles at a single locus control seed dormancy in Swedish Arabidopsis. ELife *5*, e22502.

785 Kronholm, I., and Collins, S. (2016). Epigenetic mutations can both help and hinder adaptive
786 evolution. Mol. Ecol. *25*, 1856–1868.

787 Kucharski, R., Maleszka, J., Foret, S., and Maleszka, R. (2008). Nutritional control of reproductive
788 status in honeybees via DNA methylation. Science *319*, 1827–1830.

789 Larkin, A., Geddes, J.A., Martin, R.V., Xiao, Q., Liu, Y., Marshall, J.D., Brauer, M., and Hystad, P.
790 (2017). Global Land Use Regression Model for Nitrogen Dioxide Air Pollution. Environ. Sci. Technol.
791 *51*, 6957–6964.

792 Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation
793 patterns in plants and animals. Nat. Rev. Genet. *11*, 204–220.

794 Lee, S.C., Mustroph, A., Sasidharan, R., Vashisht, D., Pedersen, O., Oosumi, T., Voesenek, L.A.C.J., and
795 Bailey-Serres, J. (2011). Molecular characterization of the submergence response of the Arabidopsis
796 thaliana ecotype Columbia. New Phytol. *190*, 457–471.

797 Lelieveld, J., Beirle, S., Hörmann, C., Stenchikov, G., and Wagner, T. (2015). Abrupt recent trend
798 changes in atmospheric nitrogen dioxide over the Middle East. Sci. Adv. *1*, e1500498.

799 Lewis, S.H., Ross, L., Bain, S.A., Pahita, E., Smith, S.A., Cordaux, R., Miska, E.A., Lenhard, B., Jiggins,
800 F.M., and Sarkies, P. (2020). Widespread conservation and lineage-specific diversification of genome-
801 wide DNA methylation patterns across arthropods. PLOS Genet. *16*, e1008864.

802 Li, P., Filiault, D., Box, M.S., Kerdaffrec, E., Oosterhout, C. van, Wilczek, A.M., Schmitt, J., McMullan,
803 M., Bergelson, J., Nordborg, M., et al. (2014). Multiple FLC haplotypes defined by independent cis-
804 regulatory variation underpin life history diversity in Arabidopsis thaliana. Genes Dev. *28*, 1635–
805 1640.

806 Li, W., Liu, H., Cheng, Z.J., Su, Y.H., Han, H.N., Zhang, Y., and Zhang, X.S. (2011a). DNA Methylation
807 and Histone Modifications Regulate De Novo Shoot Regeneration in Arabidopsis by Modulating
808 WUSCHEL Expression and Auxin Signaling. PLOS Genet. *7*, e1002243.

809 Li, Y., Huang, Y., Bergelson, J., Nordborg, M., and Borevitz, J.O. (2010). Association mapping of local
810 climate-sensitive quantitative trait loci in Arabidopsis thaliana. Proc. Natl. Acad. Sci. *107*, 21199–
811 21204.

812 Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., Shao, D., Xu, C., Li, X., Xiao, J., et al. (2011b). Natural
813 variation in GS5 plays an important role in regulating grain size and yield in rice. Nat. Genet. *43*,
814 1266–1269.

815 Liu, Y. (2018). Chapter Five - Natural selection and pangenesis: the Darwinian Synthesis of evolution
816 and genetics. In Advances in Genetics, D. Kumar, ed. (Academic Press), pp. 121–142.

817 Lopez-Arboleda, W.A., Reinert, S., Nordborg, M., and Korte, A. (2021). Global genetic heterogeneity
818 in adaptive traits. Mol. Biol. Evol.

819 Luo, C., Hajkova, P., and Ecker, J.R. (2018). Dynamic DNA methylation: In the right place at the right
820 time. Science *361*, 1336–1340.

821 Lyons, D.B., and Zilberman, D. (2017). DDM1 and Lsh remodelers allow methylation of DNA wrapped
822 in nucleosomes. ELife *6*, e30674.

823 Manning, K., Tör, M., Poole, M., Hong, Y., Thompson, A.J., King, G.J., Giovannoni, J.J., and Seymour,
824 G.B. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription
825 factor inhibits tomato fruit ripening. Nat. Genet. *38*, 948–952.

826 Mathers, T.C., Mugford, S.T., Percival-Alwyn, L., Chen, Y., Kaithakottil, G., Swarbreck, D., Hogenhout,
827 S.A., and Oosterhout, C. van (2019). Sex-specific changes in the aphid DNA methylation landscape.
828 Mol. Ecol. *28*, 4228–4241.

829 McCouch, S.R., Wright, M.H., Tung, C.-W., Maron, L.G., McNally, K.L., Fitzgerald, M., Singh, N.,
830 DeClerck, G., Agosto-Perez, F., Korniliev, P., et al. (2016). Open access resources for genome-wide
831 association mapping in rice. Nat. Commun. *7*, 10532.

832 Meng, D., Dubin, M., Zhang, P., Osborne, E.J., Stegle, O., Clark, R.M., and Nordborg, M. (2016).
833 Limited contribution of DNA methylation variation to expression regulation in Arabidopsis thaliana.
834 PLOS Genet. *12*, e1006141.

835 Miura, K., Agetsuma, M., Kitano, H., Yoshimura, A., Matsuoka, M., Jacobsen, S.E., and Ashikari, M.
836 (2009). A metastable DWARF1 epigenetic mutant affecting plant stature in rice. Proc. Natl. Acad. Sci.
837 *106*, 11218–11223.

838 Muyle, A., Ross-Ibarra, J., Seymour, D.K., and Gaut, B.S. (2021). Gene body methylation is under
839 selection in Arabidopsis thaliana. Genetics *https://doi.org/10.1093/genetics/iyab061*.

840 Nadeau, J.H. (2015). The nature of evidence for and against epigenetic inheritance. Genome Biol. *16*,
841 137.

842 Niederhuth, C.E., Bewick, A.J., Ji, L., Alabady, M.S., Kim, K.D., Li, Q., Rohr, N.A., Rambani, A., Burke,
843 J.M., Udall, J.A., et al. (2016). Widespread natural variation of DNA methylation within angiosperms.
844 Genome Biol. *17*, 194.

845 van Nocker, S., and Ludwig, P. (2003). The WD-repeat protein superfamily in Arabidopsis:
846 conservation and divergence in structure and function. BMC Genomics *4*, 50.

847 Oberlin, S., Sarazin, A., Chevalier, C., Voinnet, O., and Marí-Ordóñez, A. (2017). A genome-wide
848 transcriptome and translatome analysis of Arabidopsis transposons identifies a unique and
849 conserved genome expression strategy for Ty1/Copia retroelements. Genome Res. *27*, 1549–1562.

850 Ong-Abdullah, M., Ordway, J.M., Jiang, N., Ooi, S.-E., Kok, S.-Y., Sarpan, N., Azimi, N., Hashim, A.T.,
851 Ishak, Z., Rosli, S.K., et al. (2015). Loss of Karma transposon methylation underlies the mantled
852 somaclonal variant of oil palm. Nature *525*, 533–537.

853 Qiu, Y., and Köhler, C. (2020). Mobility connects: transposable elements wire new transcriptional
854 networks by transferring transcription factor binding motifs. Biochem. Soc. Trans. *48*, 1005–1017.

855 Quadrana, L., and Colot, V. (2016). Plant transgenerational epigenetics. Annu. Rev. Genet. *50*, 467–
856 491.

857 Quadrana, L., Almeida, J., Asís, R., Duffy, T., Dominguez, P.G., Bermúdez, L., Conti, G., Corrêa da Silva,
858 J.V., Peralta, I.E., Colot, V., et al. (2014). Natural occurring epialleles determine vitamin E
859 accumulation in tomato fruits. Nat. Commun. *5*, 4027.

860 Quadrana, L., Bortolini Silveira, A., Mayhew, G.F., LeBlanc, C., Martienssen, R.A., Jeddeloh, J.A., and
861 Colot, V. (2016). The Arabidopsis thaliana mobilome and its impact at the species level. ELife *5*,
862 e15716.

863 Robertson, K.D., and A.Jones, P. (2000). DNA methylation: past, present and future directions.
864 Carcinogenesis *21*, 461–467.

865 Sánchez-Barrena, M.J., Martínez-Ripoll, M., Zhu, J.-K., and Albert, A. (2005). The structure of the
866 Arabidopsis thaliana SOS3: molecular mechanism of sensing calcium for salt stress response. J. Mol.
867 Biol. *345*, 1253–1264.

868 Sasaki, E., Kawakatsu, T., Ecker, J.R., and Nordborg, M. (2019). Common alleles of CMT2 and NRPE1
869 are major determinants of CHH methylation variation in Arabidopsis thaliana. PLOS Genet. *15*,
870 e1008492.

871 Schmid, M.W., Heichinger, C., Coman Schmid, D., Guthörl, D., Gagliardini, V., Bruggmann, R., Aluri, S.,
872 Aquino, C., Schmid, B., Turnbull, L.A., et al. (2018). Contribution of epigenetic variation to adaptation
873 in Arabidopsis. Nat. Commun. *9*, 4446.

874 Schmitz, R.J., Schultz, M.D., Lewsey, M.G., O'Malley, R.C., Urich, M.A., Libiger, O., Schork, N.J., and
875 Ecker, J.R. (2011). Transgenerational epigenetic instability is a source of novel methylation variants.
876 Science *334*, 369–373.

877 Schmitz, R.J., Schultz, M.D., Urich, M.A., Nery, J.R., Pelizzola, M., Libiger, O., Alix, A., McCosh, R.B.,
878 Chen, H., Schork, N.J., et al. (2013). Patterns of population epigenomic diversity. Nature *495*, 193–
879 198.

880 Seymour, D.K., and Gaut, B.S. (2020). Phylogenetic shifts in gene body methylation correlate with
881 gene expression and reflect trait conservation. Mol. Biol. Evol. *37*, 31–43.

882 Shahzad, Z., Canut, M., Tournaire-Roux, C., Martinière, A., Boursiac, Y., Loudet, O., and Maurel, C.
883 (2016). A potassium-dependent oxygen sensing pathway regulates plant root hydraulics. Cell *167*,
884 87-98.e14.

885 Shahzad, Z., Eaglesfield, R., Carr, C., and Amtmann, A. (2020). Cryptic variation in RNA-directed DNA-
886 methylation controls lateral root development when auxin signalling is perturbed. Nat. Commun. *11*,
887 218.

888 Shen, X., Jonge, J.D., Forsberg, S.K.G., Pettersson, M.E., Sheng, Z., Hennig, L., and Carlborg, Ö. (2014).
889 Natural CMT2 variation is associated with genome-wide methylation changes and temperature
890 seasonality. PLOS Genet. *10*, e1004842.

891 Shindo, C., Aranzana, M.J., Lister, C., Baxter, C., Nicholls, C., Nordborg, M., and Dean, C. (2005). Role
892 of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. Plant
893 Physiol. *138*, 1163–1173.

894 Shindo, C., Bernasconi, G., and Hardtke, C.S. (2007). Natural genetic variation in Arabidopsis: tools,
895 traits and prospects for evolutionary ecology. Ann. Bot. *99*, 1043–1054.

896 Slatkin, M. (2009). Epigenetic inheritance and the missing heritability problem. Genetics *182*, 845–
897 850.

898 Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the
899 genome. Nat. Rev. Genet. *8*, 272–285.

900 Soppe, W.J.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and
901 Peeters, A.J.M. (2000). The late flowering phenotype of FWA mutants is caused by gain-of-function
902 epigenetic alleles of a homeodomain gene. Mol. Cell *6*, 791–802.

903 Stolc, V., Samanta, M.P., Tongprasit, W., Sethi, H., Liang, S., Nelson, D.C., Hegeman, A., Nelson, C.,
904 Rancour, D., Bednarek, S., et al. (2005). Identification of transcribed sequences in Arabidopsis
905 thaliana by using high-resolution genome tiling arrays. Proc. Natl. Acad. Sci. *102*, 4453–4458.

906 Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014). Non-
907 CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat. Struct. Mol. Biol. *21*,
908 64–72.

909 Stuart, T., Eichten, S.R., Cahn, J., Karpievitch, Y.V., Borevitz, J.O., and Lister, R. (2016). Population
910 scale mapping of transposable element diversity reveals links to gene regulation and epigenomic
911 variation. ELife *5*, e20777.

912 Sung, S., and Amasino, R.M. (2004). Vernalization in Arabidopsis thaliana is mediated by the PHD
913 finger protein VIN3. Nature *427*, 159–164.

914 Takuno, S., and Gaut, B.S. (2013). Gene body methylation is conserved between plant orthologs and
915 is of evolutionary consequence. Proc. Natl. Acad. Sci. *110*, 1797–1802.

916 Takuno, S., Ran, J.-H., and Gaut, B.S. (2016). Evolutionary patterns of genic DNA methylation vary
917 across land plants. Nat. Plants *2*, 1–7.

918 Takuno, S., Seymour, D.K., and Gaut, B.S. (2017). The evolutionary dynamics of orthologs that shift in
919 gene body methylation between Arabidopsis species. Mol. Biol. Evol. *34*, 1479–1491.

920 Tang, N., Shahzad, Z., Lonjon, F., Loudet, O., Vailleau, F., and Maurel, C. (2018). Natural variation at
921 XND1 impacts root hydraulics and trade-off for stress responses in Arabidopsis. Nat. Commun. *9*,
922 3884.

923 Taylor, M.A., Wilczek, A.M., Roe, J.L., Welch, S.M., Runcie, D.E., Cooper, M.D., and Schmitt, J. (2019).
924 Large-effect flowering time mutations reveal conditionally adaptive paths through fitness landscapes
925 in Arabidopsis thaliana. Proc. Natl. Acad. Sci. *116*, 17890–17899.

926 The 1001 Genomes Consortium (2016). 1,135 genomes reveal the global pattern of polymorphism in
927 Arabidopsis thaliana. Cell *166*, 481–491.

928 Van der Graaf, A., Wardenaar, R., Neumann, D.A., Taudt, A., Shaw, R.G., Jansen, R.C., Schmitz, R.J.,
929 Colomé-Tatché, M., and Johannes, F. (2015). Rate, spectrum, and evolutionary dynamics of
930 spontaneous epimutations. Proc. Natl. Acad. Sci. *112*, 6676–6681.

931 Verbruggen, N., and Hermans, C. (2008). Proline accumulation in plants: a review. Amino Acids *35*,
932 753–759.

933 Vidalis, A., Živković, D., Wardenaar, R., Roquis, D., Tellier, A., and Johannes, F. (2016). Methylome
934 evolution in plants. Genome Biol. *17*, 264.

935 Vilhjálmsson, B.J., and Nordborg, M. (2013). The nature of confounding in genome-wide association
936 studies. Nat. Rev. Genet. *14*, 1–2.

937 Wang, X., Werren, J.H., and Clark, A.G. (2016). Allele-specific transcriptome and methylome analysis
938 reveals stable iInheritance and cis-regulation of DNA methylation in Nasonia. PLOS Biol. *14*,
939 e1002500.

940 Weigel, D., and Nordborg, M. (2015). Population genomics for understanding adaptation in wild
941 plant species. Annu. Rev. Genet. *49*, 315–338.

942 Whittaker, C., and Dean, C. (2017). The FLC locus: a platform for discoveries in epigenetics and
943 adaptation. Annu. Rev. Cell Dev. Biol. *33*, 555–575.

944 Xu, S. (2003). Theoretical basis of the Beavis effect. Genetics *165*, 2259–2268.

945 Yagound, B., Remnant, E.J., Buchmann, G., and Oldroyd, B.P. (2020). Intergenerational transfer of
946 DNA methylation marks in the honey bee. Proc. Natl. Acad. Sci. *117*, 32519–32527.

947 Yan, H., Bonasio, R., Simola, D.F., Liebig, J., Berger, S.L., and Reinberg, D. (2015). DNA methylation in
948 social insects: how epigenetics can control behavior and longevity. Annu. Rev. Entomol. *60*, 435–452.

949 Yano, K., Morinaka, Y., Wang, F., Huang, P., Takehara, S., Hirai, T., Ito, A., Koketsu, E., Kawamura, M.,
950 Kotake, K., et al. (2019). GWAS with principal component analysis identifies a gene comprehensively
951 controlling rice architecture. Proc. Natl. Acad. Sci. *116*, 21262–21267.

952 Yu, Z., Duan, X., Luo, L., Dai, S., Ding, Z., and Xia, G. (2020). How plant hormones mediate salt stress
953 responses. Trends Plant Sci. *25*, 1117–1130.

954 Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of
955 eukaryotic DNA methylation. Science *328*, 916–919.

956 Zeng, Y., and Chen, T. (2019). DNA methylation reprogramming during mammalian development.
957 Genes *10*.

958 Zhang, H., Lang, Z., and Zhu, J.-K. (2018). Dynamics and function of DNA methylation in plants. Nat.
959 Rev. Mol. Cell Biol. *19*, 489–506.

960 Zhang, X., Cal, A.J., and Borevitz, J.O. (2011). Genetic architecture of regulatory variation in
961 Arabidopsis thaliana. Genome Res. *21*, 725–733.

962 Zhang, Y., Wendte, J.M., Ji, L., and Schmitz, R.J. (2020). Natural variation in DNA methylation
963 homeostasis and the emergence of epialleles. Proc. Natl. Acad. Sci. *117*, 4874–4884.

964 Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett,
965 D.K., Ordovas, J.M., et al. (2010). Mixed linear model approach adapted for genome-wide association
966 studies. Nat. Genet. *42*, 355–360.

967 Zilberman, D. (2017). An evolutionary case for functional gene body methylation in plants and
968 animals. Genome Biol. *18*, 87.

969

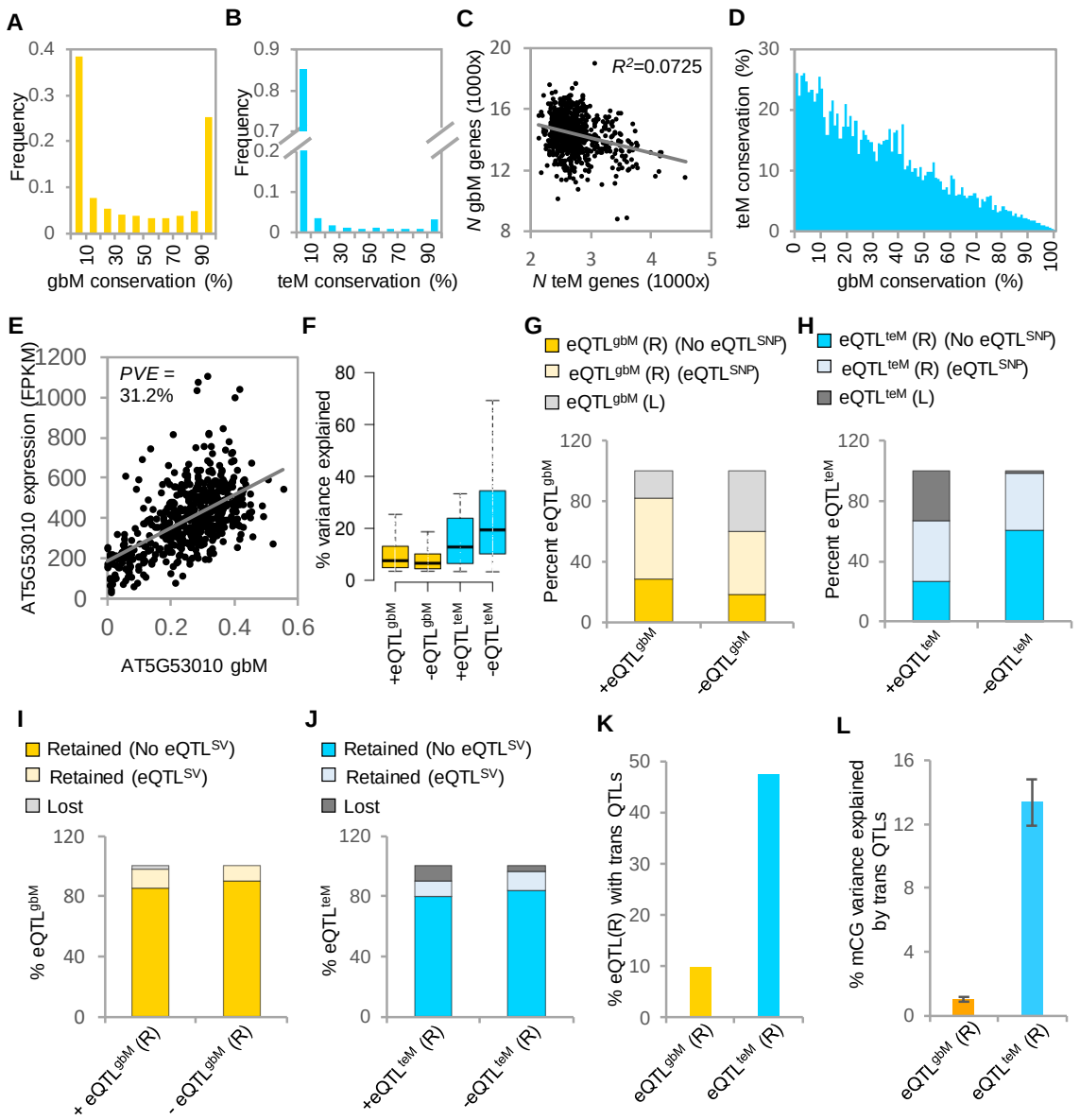

Fig. 1. GbM is associated with gene expression independently of genetic variation in natural populations. (**A** and **B**) Frequency distribution of gbM (**A**) and teM (**B**) conservation across 33,056 *Arabidopsis* genes in 948 accessions. (**C**) Correlation between the number (*N*) of gbM and teM 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.

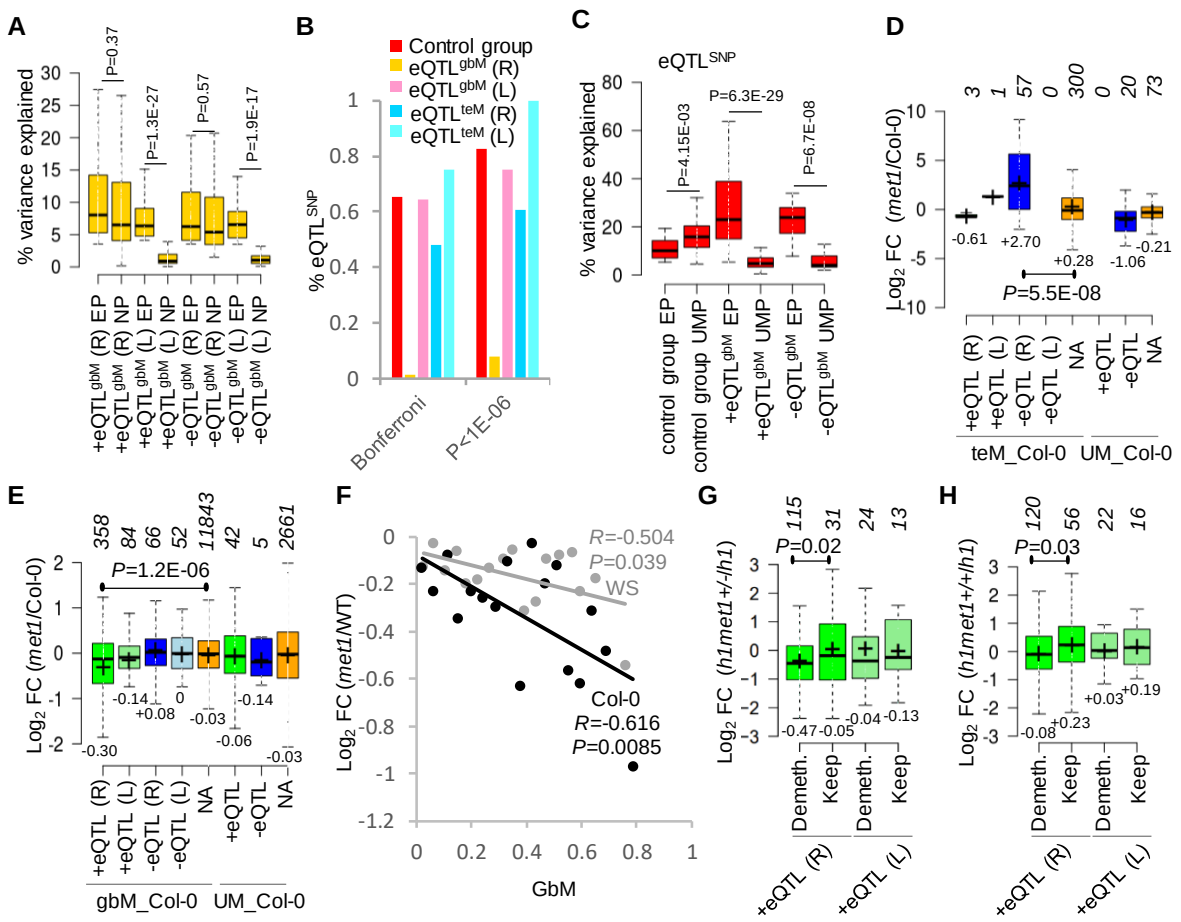

Fig. 2. GbM promotes gene expression. (**A**) Effect sizes of gbM on expression of positive and negative retained (R) or lost (L) eQTL^{gbM} genes in the entire population (EP) and after accounting for expression GWA SNPs in nested populations (NP). (**B**) Percent of eQTL^{SNP} detected as significant 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.

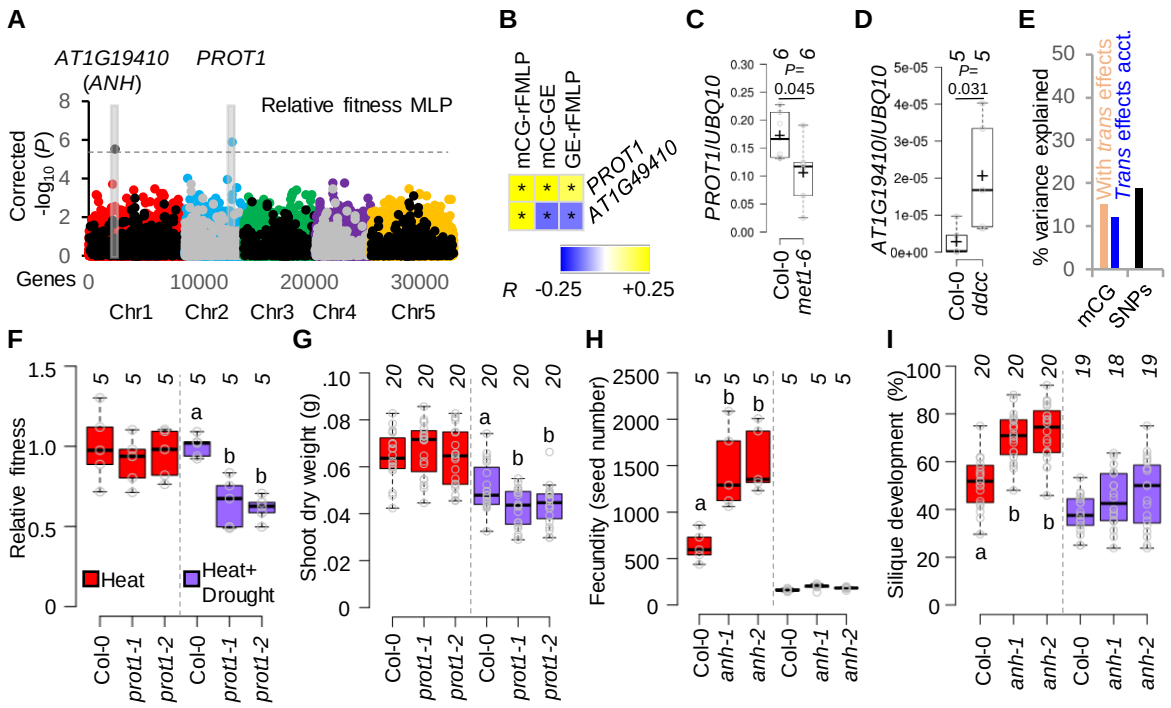

Fig. 3. Intragenic DNA methylation affects relative fitness. (**A**) Manhattan plot of epiGWA mapping for relative fitness in MLP conditions. GbM markers are depicted with colored dots and teM markers are shown using grey and black dots. Horizontal dashed line shows 0.05 FDR. (**B**) Tripartite associations 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.

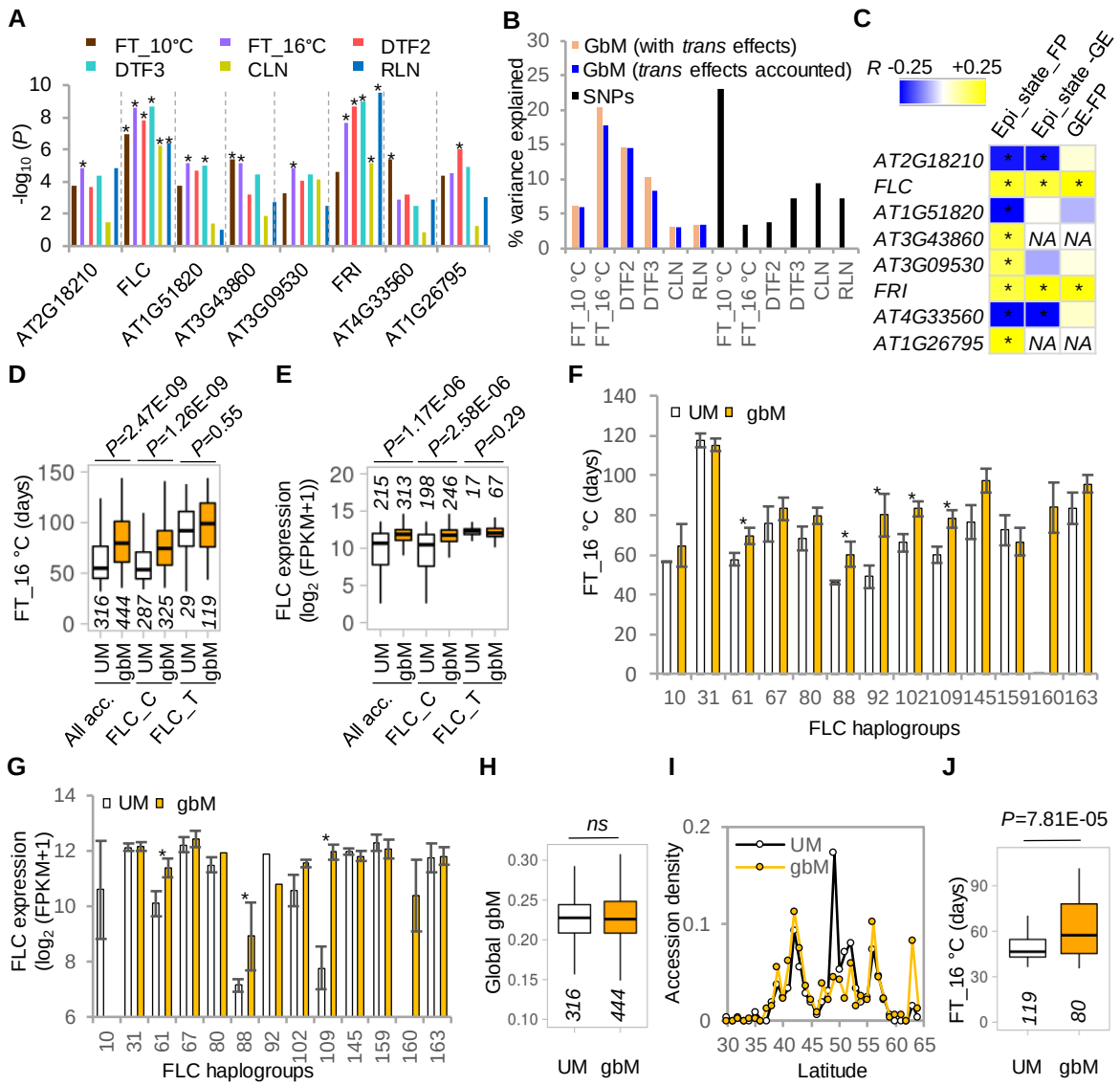

Fig. 4. Flowering time is similarly influenced by natural genetic and epigenetic variation. (**A**) *P* values (-log$_{10}$) of epigenetic flowering QTLs. *Significant at 0.05 FDR. (**B**) Percent variation in flowering phenotypes explained by epigenetic (gbM) and genetic (SNPs) QTLs. Effects of *trans* genetic QTLs on gbM variation are accounted. Effect of *cis* GWA SNP and SV on fQTL^{gbM} are also accounted. (**C**) 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. (**D** and **E**) Association of FLC epiallelic states with FT_16°C (**D**) and FLC expression (**E**) 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. (**F** and **G**) Association of FLC epiallelic states with FT_16°C (**F**) and FLC expression (**G**) 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. (**H**) Genome-wide average gbM levels of accessions with UM or gbM *FLC*. Accession numbers are indicated. *ns* = non-significant, two-tailed Student's t test. (**I**) Density of FLC UM and gbM accessions across latitude. (**J**) Association of FLC epiallelic states with FT_16°C within 48-52° latitude. Accession numbers are indicated. *P* value corresponds to MLM epiGWA.

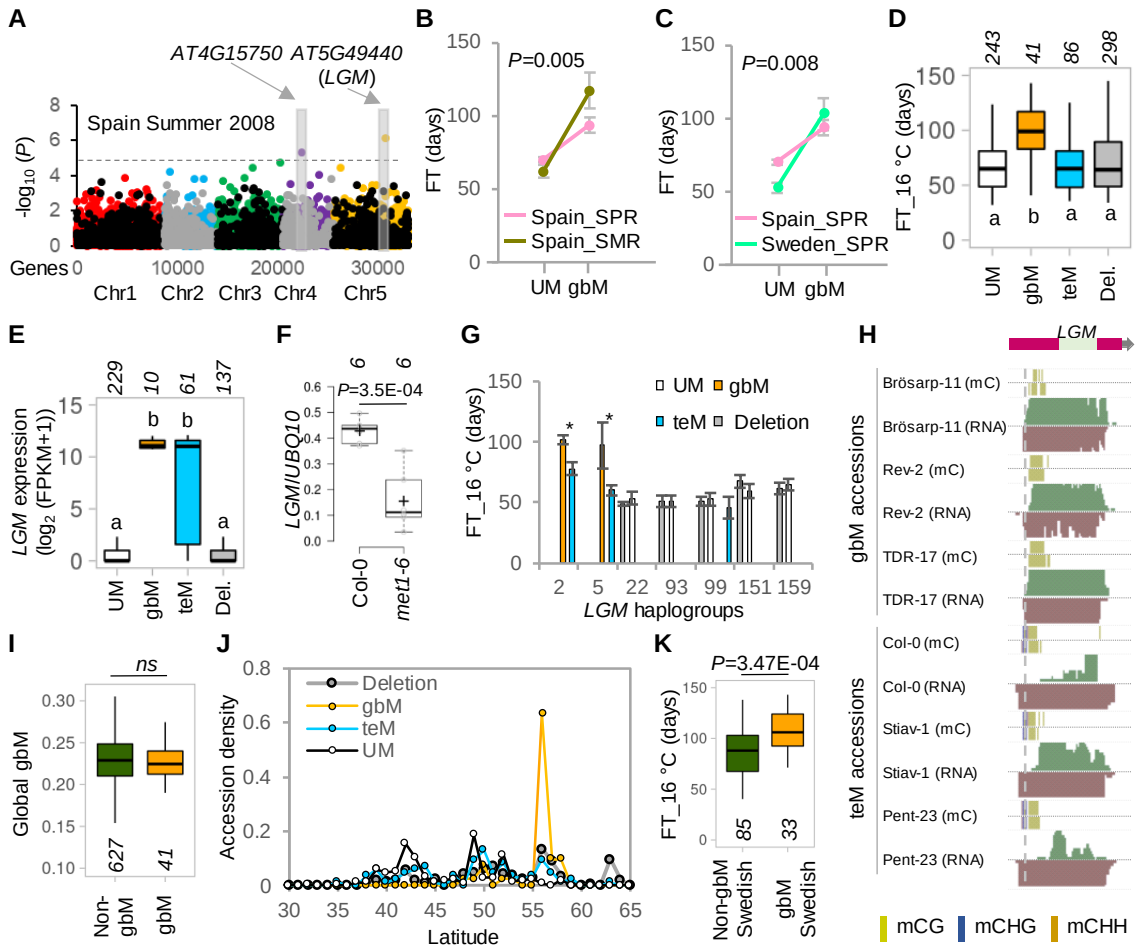

Fig. 5. GbM in *LGM* delays flowering. (**A**) Manhattan plot of MLM epiGWA mapping for flowering time in the 2008 Spain summer planting. GbM markers are depicted with colored dots and teM markers are shown using grey and black dots. Horizontal dashed line shows 0.05 FDR. (**B** and **C**) Sensitivity 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. (**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.

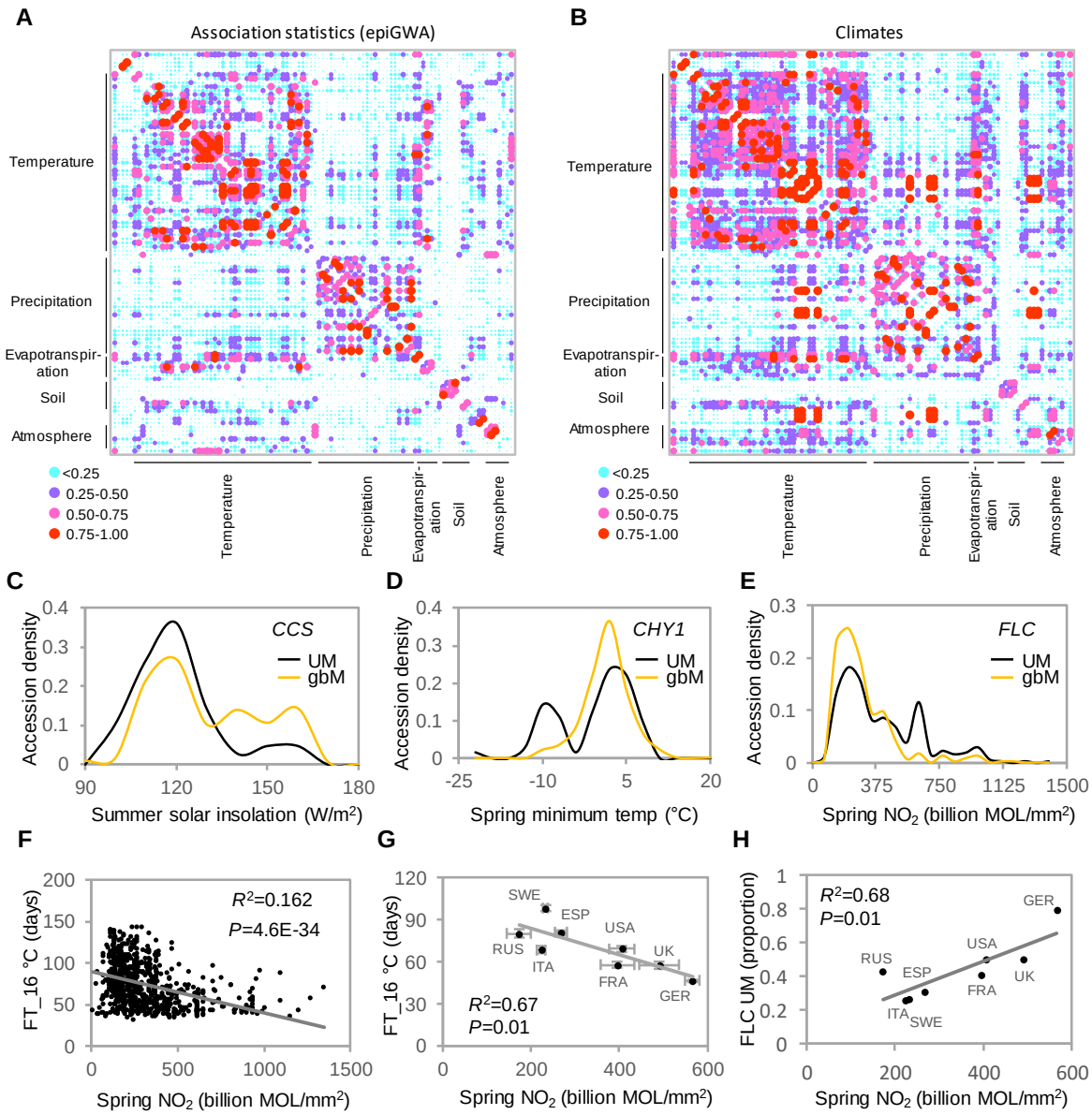

Fig. 6. GbM variation is associated with geoclimatic variables. (**A** and **B**) Correlation (R^2) matrices of epiGWA *P* values (**A**) and environmental variables (**B**) for 57 genes identified in at least three epiGWA analyses. See Supplementary Tables 16 and 17 for individual environment labels 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.