Gene body methylation mediates epigenetic inheritance of plant traits

Zaigham Shahzad[1], Jonathan D. Moore[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

Cytosine methylation is an epigenetically heritable DNA modification common in plant and animal genes, but the functional and evolutionary significance of gene body methylation (gbM) has remained enigmatic. Here we show that gbM enhances gene expression in *Arabidopsis thaliana*. We also demonstrate that natural gbM variation influences drought and heat tolerance and flowering time by modulating gene expression, including that of *Flowering Locus C* (*FLC*). Notably, epigenetic variation accounts for as much trait heritability in natural populations as DNA sequence polymorphism. Furthermore, we identify gbM variation in numerous genes associated with environmental variables, including a strong association between flowering time, spring atmospheric NO_2 – a by-product of fossil fuel burning – and *FLC* epialleles. Our study demonstrates that gbM is an important modulator of gene expression, and its natural variation fundamentally shapes phenotypic diversity in plant populations. Thus, gbM provides an epigenetic basis for adaptive evolution independent of genetic polymorphism.

DNA methylation can be epigenetically inherited over many generations[1,2] and occurs in transposable elements (TEs) and bodies of transcribed genes[3–6]. Plant TEs are generally methylated in all sequence contexts – CG, CHG, and CHH (H being A, T, or C)[3,4,7,8]. TE methylation induces silencing[7], confers genome stability[6,9], and can influence the expression of neighbouring genes[10–15]. Gene body methylation (gbM) occurs only in the CG context[3,4,16], although genes can also feature TE-like methylation in all contexts (teM)[17,18]. TeM is associated with silencing[5,17,18], but the function of gbM has been extensively debated. GbM is nearly ubiquitous in flowering plants[19,20] and is common in animals[3,4,21]. In both groups, gbM preferentially resides in nucleosome-wrapped DNA within the exons of conserved, constitutively transcribed genes[5,22–25]. Conservation and phenomenological coherence suggest important functions[21]. Indeed, gbM is associated with (small) gene expression differences across plant species[26,27], represses aberrant intragenic transcripts[28], and appears to be under natural selection[26,29,30]. Moreover, loss of methyltransferase function causes developmental abnormalities in honeybees[31], animals in which methylation is principally restricted to gene bodies[32]. However, loss of gbM has not been linked to changes in gene expression in plants or animals[5,33,34], leading to the proposals that gbM is a non-functional and somewhat deleterious by-product of TE methylation (in plants)[5,34,35] or has functions unrelated to gene expression (in animals)[33]. 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 thaliana*[17,18]. Methylation levels of natural accessions are associated with climate[17], suggesting that methylation variation could be of adaptive significance. Furthermore, experimentally induced methylation polymorphism can account for the inheritance of complex *Arabidopsis* traits[36–38], and methylation changes have been linked to adaptation under artificial selection[39]. Due to its stable epigenetic inheritance[40], gbM appears particularly suited to mediate natural phenotypic variation. Nonetheless, whether gbM or any other type of methylation drives phenotypic diversity in natural populations is unknown.

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[17]. 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[21,29,30]. In contrast, the vast majority of genes exhibit teM in ≤10% of accessions (Fig. 1b), suggesting that teM is disfavoured in most genes, likely due to its deleterious effects on expression[17].

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[35] of gbM and teM (Supplementary Fig. 1a). Additionally, intragenic teM prevalence across accessions tends to increase with decreasing gbM conservation under either gbM/teM definition (Fig. 1d and Supplementary Fig. 1b-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[34] or that gbM promotes the transition to teM[35]. 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)[3–5].

Intragenic DNA methylation is associated with transcription in natural populations

To investigate the effects of gbM and teM on gene expression, we analysed associations between CG methylation (mCG) and mRNA levels of individual genes 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, Supplementary Fig. 2 and Supplementary Tables 1 and 2) 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 (Supplementary Tables 1 and 2). 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. GbM generally accounts for ≤10% of expression variance (93.7% of eQTL^{gbM} Bonferroni genes), but in exceptional cases gbM explains around 40% of expression variance (Fig. 1f and Supplementary Fig. 3a). Non-associated gbM genes (NA^{gbM}) and eQTL^{gbM} genes have similar CG dinucleotide composition and length (Fig. 1g, h) – key distinguishing characteristics of gbM genes[35] – indicating that eQTL^{gbM} genes are fairly typical. However, +eQTL^{gbM} genes show lower expression than NA^{gbM} genes (Fig. 1i), 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.

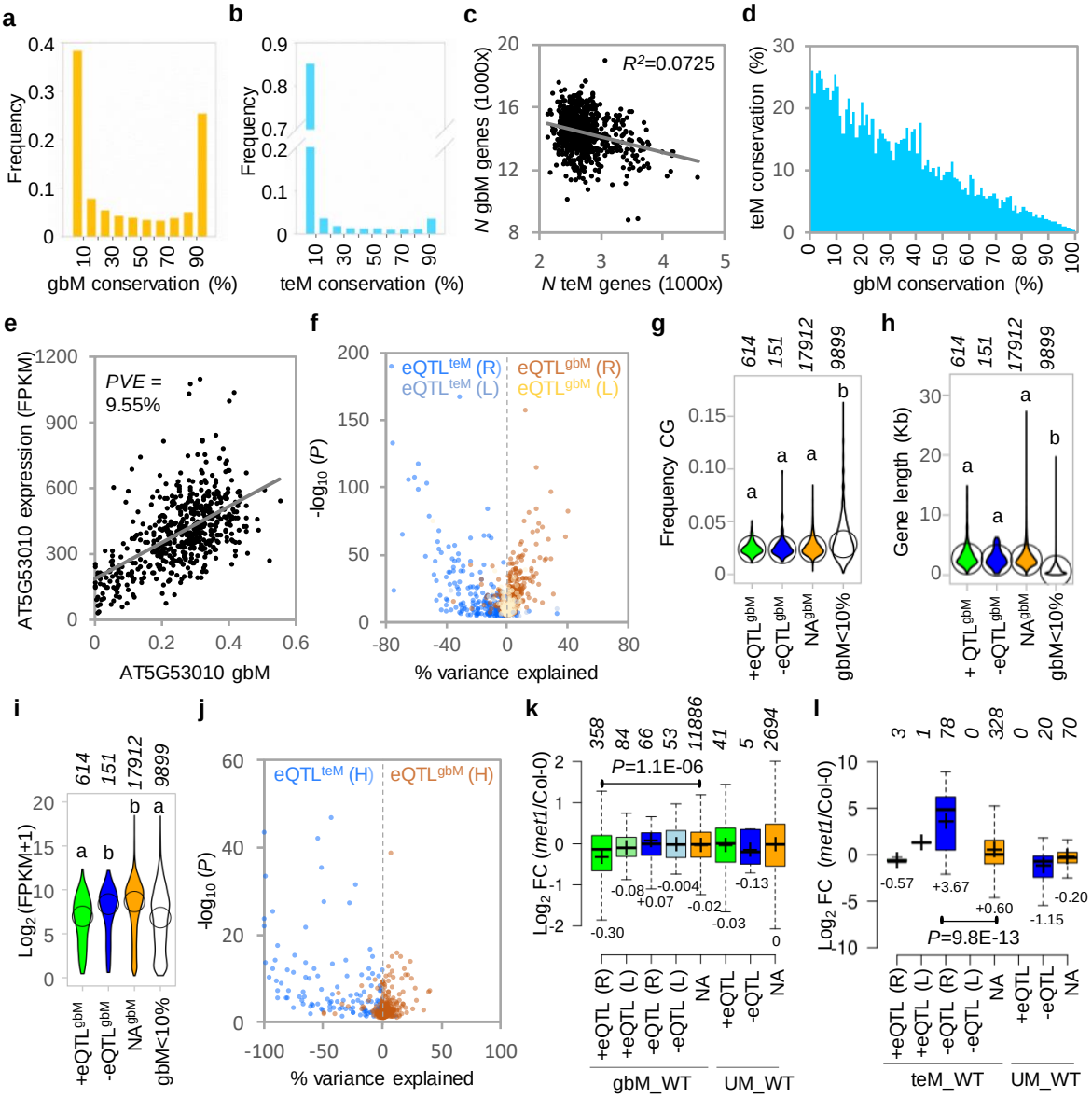

Fig. 1 | GbM and teM variation regulates gene expression. a, b, Frequency distribution of gbM (**a**) and teM (**b**) conservation across 33,056 *Arabidopsis* genes in 948 accessions. **c**, Correlation between the number (*N*) of gbM and teM genes across accessions. **d**, Average teM conservation within 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 retained (R) or lost (L) eQTL$^{gbM/teM}$ (PVE and association *P* values). **g-i**, CG dinucleotide frequency (**g**), gene length (**h**), and gene expression (**i**) of eQTL^{gbM} genes, NA^{gbM} genes, and genes with gbM in <10% of accessions. Different letters signify *P*<0.001, one-way ANOVA, Dunn's test. Numbers of genes within each group are indicated. **j**, GbM and teM effect sizes on eQTL$^{gbM/teM}$ gene expression in the most significant haplogroup (H). **k, l**, Expression in *met1* compared to Col-0 of Bonferroni (α=0.05) eQTL^{gbM} (**k**) and eQTL^{teM} (**l**) genes retained (R) or lost (L) after accounting for genetic variation. Genes methylated (gbM_WT; teM_WT) and unmethylated (UM_WT) in Col-0 were analysed separately. Numbers of genes within each group are indicated. *P*, Wilcoxon rank sum test.

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}) (Supplementary Table 1). As with gbM, the less stringent FDR thresholds produced more associations, but with a reduced ratio of -eQTL^{teM} to +eQTL^{teM} (Supplementary Table 1). TeM effects on gene expression are stronger than for gbM, as teM explains >20% of expression variance in 37.7% of eQTL^{teM} Bonferroni genes (Fig. 1f and Supplementary Fig. 3b). The primarily negative associations between teM and expression are consistent with existing knowledge[17] and indicate that teM and gbM have fundamentally different effects on gene expression.

GbM is associated with transcription in the absence of local genetic 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 analysed associations between mCG and expression within each population (Supplementary Fig. 4). Using this approach, we found that 82.4% of +eQTL^{gbM} were retained, either because a eQTL^{SNP} is not detected (28.7%), or because a +eQTL^{gbM} was detected in at least one nested population (53.7%) (Fig. 1f, Supplementary Fig. 5a, b and Supplementary Table 3). Similarly, nearly all (98.5%) -eQTL^{teM} were retained (Fig. 1f, Supplementary Fig. 5a, c and Supplementary Table 3). However, 39.7% of -eQTL^{gbM} and 33.3% of +eQTL^{teM} were lost after accounting for genetic variation (Fig. 1f, Supplementary Fig. 5a-c and Supplementary Table 3). 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 genetic variation.

To further disentangle the effects of epigenetic variation and *cis* genetic polymorphism, we defined haplogroups that share SNPs within retained epigenetic eQTL genes and 4 kb upstream and downstream. This analysis created numerous small groups (Supplementary Table 4), and consequently many associations are expected to be lost due to lack of statistical power (Supplementary Fig. 5d). However, 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. 1j, Supplementary Fig 5e, f and Supplementary Table 5). Moreover, the direction and magnitude of gbM and teM expression effects in haplogroups are comparable to those for the entire population (Fig. 1f, j). Hence, intragenic methylation is quantitively associated with gene expression independently of local genetic polymorphism.

MET1 inactivation specifically downregulates +eQTL^{gbM} genes

To determine whether intragenic DNA methylation directly affects gene expression, we analysed published RNA sequencing data from *met1* mutant plants and wild type controls[28]. Mutation of the *MET1* methyltransferase causes complete loss of gbM and nearly complete loss of mCG throughout the genome[40]. 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. 1k, l). Only +eQTL^{gbM} and -eQTL^{teM} genes methylated in Col-0 were significantly affected by *met1* (Fig. 1k, l). Col-0 methylated +eQTL^{gbM} genes were modestly downregulated (expressed at about 80% of wild-type) and -eQTL^{teM} genes were strongly overexpressed (over 12-fold of wild-type). Analysis of genes that passed less stringent significance thresholds (0.05 and 0.1 FDR) produced similar results (Supplementary Fig. 6). The strong upregulation of -eQTL^{teM} genes in *met1* is unsurprising and consistent with the established repressive activity of teM[5,17,18]. Crucially, the reduced expression of only those +eQTL^{gbM} genes that are methylated in Col-0 and not confounded by genetic variation demonstrates that gbM directly promotes gene expression.

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

Modulation of gene expression by natural gbM and teM polymorphism potentiates the exciting possibility that methylation epialleles can drive phenotypic variation in natural populations. To uncover how DNA methylation shapes 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[41]. Linear model associations can be confounded by population structure, producing elevated rates of false-positives[42]. Therefore, we used a 'genomic control factor'[43] to correct association statistics (Supplementary Table 6). 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. 2a and Supplementary Fig. 7). 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[44].

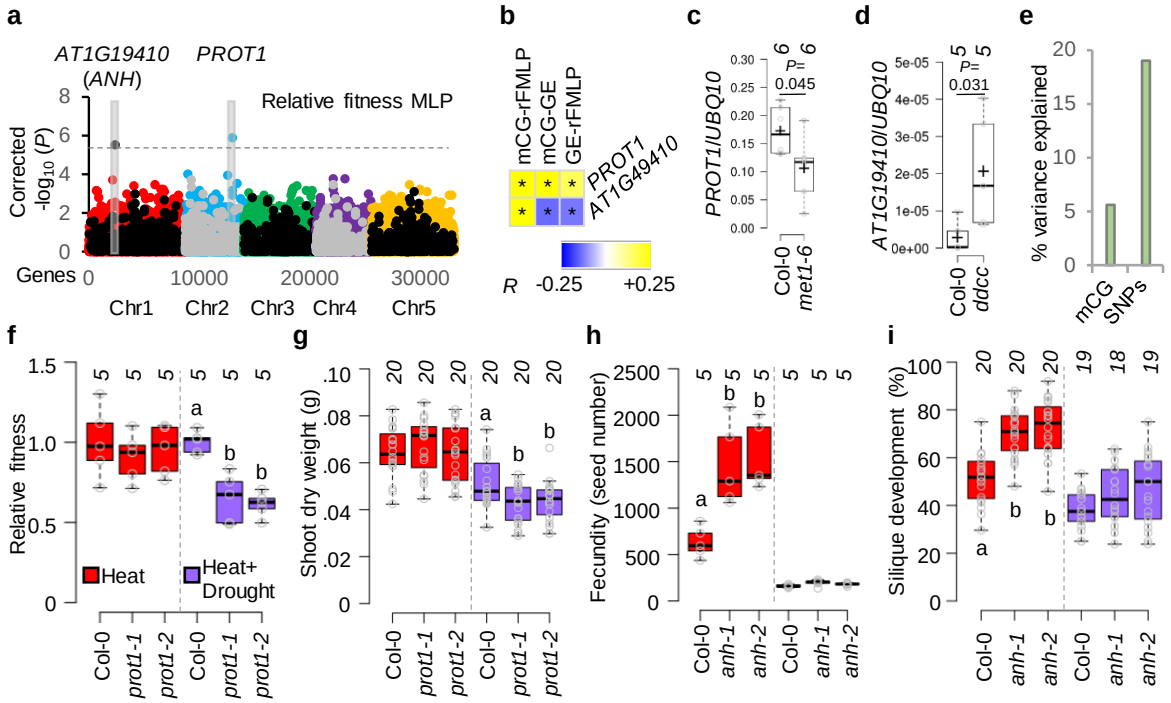

Fig. 2 | Intragenic DNA methylation affects relative fitness. a, Manhattan plot of epiGWA mapping for relative fitness in MLP conditions. GbM markers are depicted with coloured dots and teM markers are shown using grey and black dots. Horizontal dashed line shows 0.05 FDR. **b**, Tripartite associations between intragenic DNA methylation (mCG), gene expression (GE), and relative fitness in MLP (rFMLP). *P<0.05, Pearson's correlation test. **c**, **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. **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.

PROT1 is a typical gbM gene methylated in 95.1% of accessions, but *AT1G19410* is unusual, with teM in 73.4% of accessions, suggesting that teM in this gene may be functionally important (Supplementary Table 7). We identified *PROT1* as a +eQTL^{gbM} at the Bonferroni threshold, and *AT1G19410* as a -eQTL^{teM} at FDR 0.05 (Supplementary Table 2), 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. 2b). Consistent associations between mCG, fitness, and expression were detected in two independent haplogroups of *PROT1* and *AT1G19410* (Supplementary Fig. 8a and Supplementary Table 8). *PROT1* is downregulated by 38% in *met1* (Fig. 2c), confirming that gbM promotes *PROT1* expression. *AT1G19410* teM is lost in plants that lack DRM and CMT methyltransferases (Supplementary Fig. 8b), and in such *ddcc* mutants[45] *AT1G19410* expression increases about 7-fold (Fig. 2d), 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 (Supplementary Table 9), 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} (Supplementary Fig. 9). Heritability estimates indicate that the two rfQTLs^{SNP} explain 18.9% of the relative fitness variance in MLP environment (Fig. 2e). Strikingly, the two epigenetic rfQTLs also account for a substantial proportion (5.6%) of fitness variance (Fig. 2e). 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. 2f). Mutant plants produced less biomass and had decreased survival to fruit, but had the same fecundity (seed set) as wild type (Fig. 2g and Supplementary Fig. 10a). Consistently, *PROT1* gbM is specifically associated with survival in MLP conditions (Supplementary Fig. 11a-c). Thus, the

effects of *PROT1* inactivation match those expected from the associations between gbM, gene expression, and relative MLP fitness (Fig. 2b).

Inactivation of *AT1G19410* resulted in a slight (~13%) but non-significant increase in relative fitness under heat and drought stress (Supplementary Fig. 10b). 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. 2h, i, and Supplementary Fig. 10b). 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 (Supplementary Fig. 11d-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[46–48]. 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[49]. Due to the structure of flowering phenotypes in *Arabidopsis* populations[50], we could not use a linear model with genomic control to analyse quantitative methylation variation (Supplementary Fig. 12 and Supplementary Table 10). Instead, we used a mixed linear model (MLM) that effectively corrects population structure[51–54] (Supplementary Table 11) to analyse 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[55]. 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. 3a, Supplementary Fig. 13). Two fQTLs^{gbM} localized to the well-known flowering regulators *Flowering locus* C (*FLC*) and *FRIGIDA*

(*FRI*)[56–61] (Fig. 3a). GWA analysis 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*[62–64] (Supplementary Fig. 14 and Supplementary Table 12). Except for FT_10°C, fQTLs^{SNP} explain <10% of trait heritability (Fig. 3b). Strikingly, epigenetic variation explains >10% of heritability for four flowering traits (Fig. 3b). Accurate estimation of natural variation contribution to trait heritability is challenging[42,65] and the values in Fig. 3b 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. 3b). Instead, the key conclusion is that epigenetic gbM variation explains roughly the same fraction of flowering trait heritability as genetic variation.

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[60,61], but we found them associated with all (*FLC*) or nearly all (*FRI*) of the six correlated flowering phenotypes (Fig. 3a). Furthermore, both genes show consistent significant associations between gbM, flowering and gene expression (Fig. 3c), and both were detected as +eQTL^{gbM} at the Bonferroni threshold (Supplementary Table 2). Importantly, the *FRI* +eQTL^{gbM} was lost after accounting for genetic variation (Supplementary Table 3), *FRI* expression is not significantly altered in *met1* plants (Supplementary Fig. 15a), and we did not detect significant associations between *FRI* epiallelic states and FT_16°C or expression within nested genetic populations (Supplementary Fig. 15b, c). Furthermore, we found significant linkage disequilibrium[66] (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. 3b).

In contrast to *FRI*, the *FLC* +eQTL^{gbM} was retained after accounting for genetic variation (Supplementary Table 3 and 5; expression in *met1* could not be tested because *FLC* lacks gbM in Col-0, Supplementary Fig. 16a), and we identified significant associations between *FLC* epiallelic states, FT_16°C and expression within a nested population (Fig. 3d, e). Because multiple *FLC* alleles affect flowering time or vernalisation response[14,57,58], we defined thirteen *FLC* haplotypes (Supplementary Table 8), all of which contain gbM accessions (Fig. 3f), suggesting complex gbM evolution at this locus. GbM accessions display significantly delayed

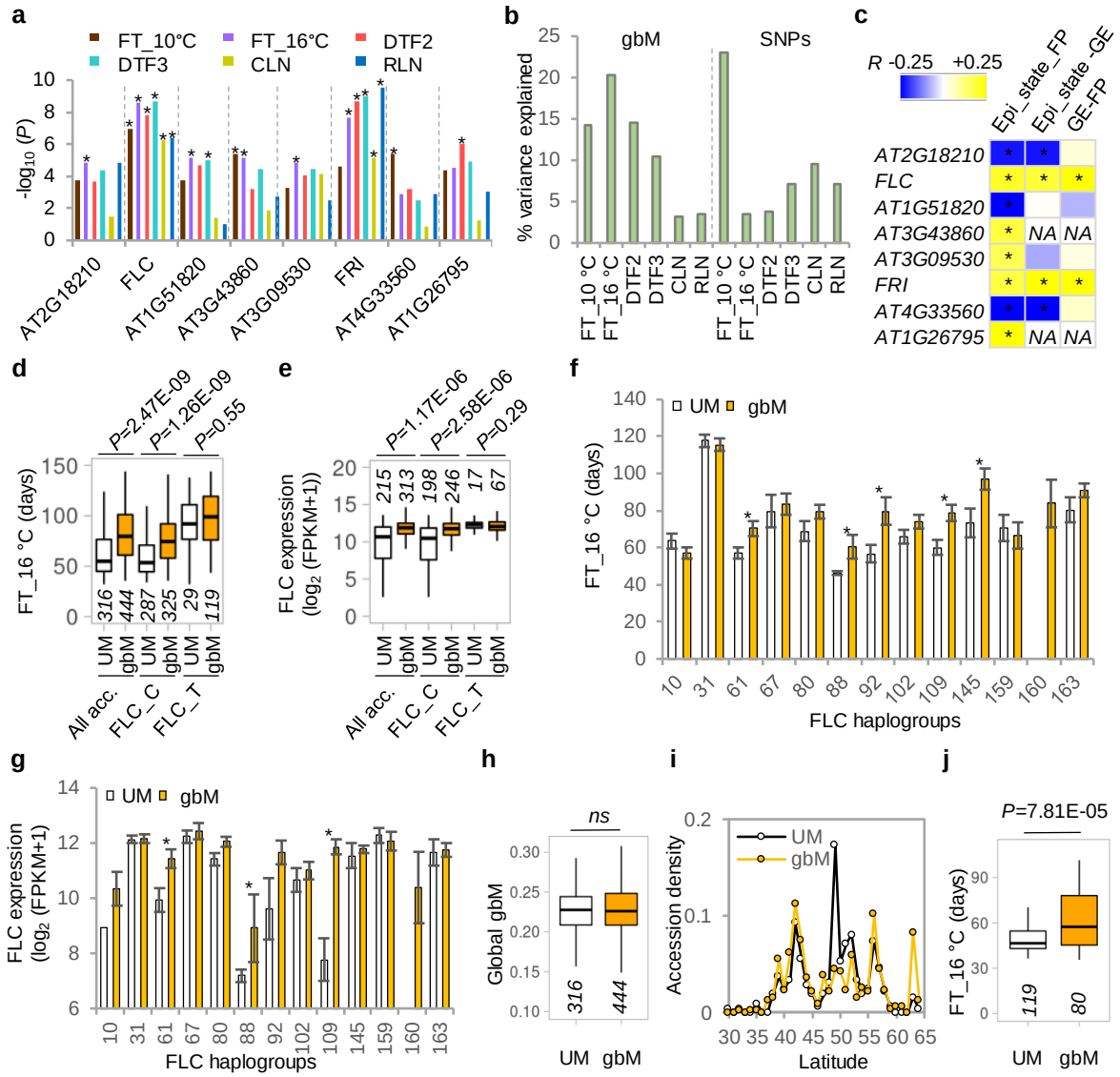

Fig. 3 | Flowering time is similarly influenced by natural genetic and epigenetic variation. a, *P* values ($-\log_{10}$) of epigenetic flowering QTLs. *Significant at 0.05 FDR. **b**, Percent variation in flowering phenotypes explained by epigenetic (gbM) and genetic (SNPs) QTLs. **c**, Tripartite associations 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**, **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**, **g**, Association of FLC epiallelic states with FT_16°C (**f**) and FLC expression (**g**) in thirteen FLC haplogroups. *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.

flowering (FT_16°C) within five haplotypes (delay of >18 days in three haplotypes; Fig. 3f), and significantly higher *FLC* expression in three of these haplotypes (Fig. 3g). These findings demonstrate that gbM variation influences *FLC* expression and flowering time independently of *FLC* genetic variation. GbM at *FLC* is also not associated with global gbM levels (Fig. 3h), suggesting that *FLC* methylation evolves independently of the genetic factors that control overall gbM.

Arabidopsis flowering time variation is correlated with geographic regions and functionally distinct *FLC* alleles can display restricted geographic distributions[58]. 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. 3i and Supplementary Fig. 16b, c). 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. 3j). The *FLC* fQTL^{gbM} is thus not confounded by links between flowering time and latitude, variation of global gbM (Fig. 3h), or *FLC* genetic polymorphism (Fig. 3d-g).

GbM of a novel locus (*LGM*) regulates flowering time

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[47]. EpiGWA identified significant fQTLs^{gbM} (FDR 0.05) associated with flowering time in simulated local climates within two uncharacterized genes, *AT4G15750* and *AT5G49440* (Fig. 4a, Supplementary Fig. 17 and Supplementary Table 13). The *AT4G15750* fQTL^{gbM} shows sensitivity to seasonal planting in Sweden and to climate (Spain vs. Sweden) in summer (Supplementary Fig. 18a-d), whereas the *AT5G49440* fQTL^{gbM} is sensitive to seasons in Spain and to climate in spring (Fig. 4b, c and Supplementary Fig. 18e, 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 (Supplementary Table 14). *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. 4d). Flowering (FT_16°C) of gbM accessions is severely delayed (by 29 days on average) compared

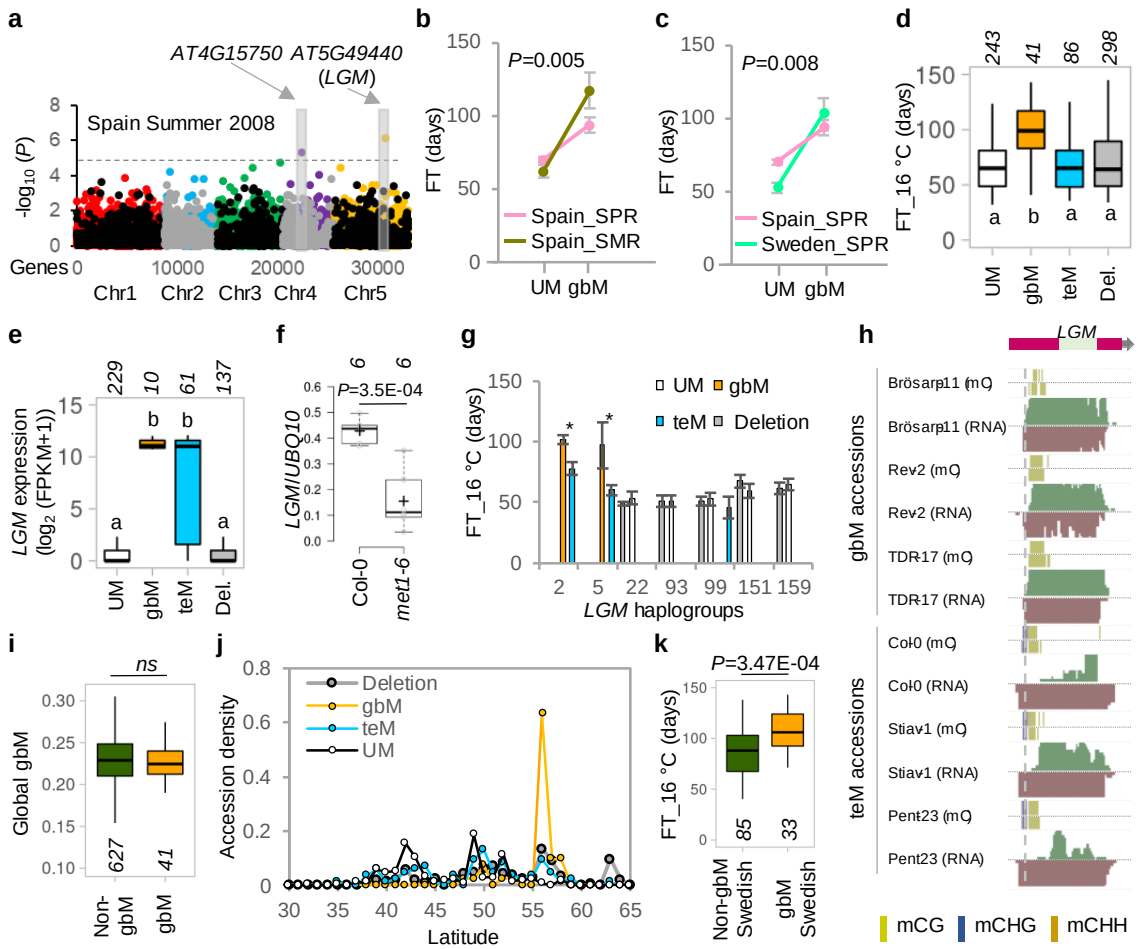

Fig. 4 | GbM in *LGM* delays flowering. a, Manhattan plot of MLM epiGWA mapping for flowering time in the 2008 Spain summer planting. GbM markers are depicted with coloured dots and teM markers are shown using grey and black dots. Horizontal dashed line shows 0.05 FDR. **b, c**, Sensitivity of AT5G49440 (*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**, **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 snapshot for representative *LGM* gbM and teM accessions along with sense (green) and antisense (brown) transcripts. 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.

to teM, UM or deleted accessions (Fig. 4d), 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. 4e). Col-0 is a teM accession, and *LGM* is strongly downregulated in *met1* (which partially reduces teM; Fig. 4f and Supplementary Fig. 19a), indicating that intragenic methylation is required for robust *LGM* expression.

Curiously, although *LGM* is highly expressed in teM accessions (Fig. 4e), we did not detect a significant association of teM with flowering time (Fig. 4a), and teM, UM and deleted accessions flower similarly early (Fig. 4d). Analysis within haplogroups (Supplementary Fig. 19a, Supplementary Table 8) revealed that gbM accessions flower significantly later in the two *LGM* haplotypes in which they are present (Fig. 4g). In contrast, teM is not associated with delayed flowering in any haplotype (Fig. 4g). 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[67]. 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. 4h and Supplementary Fig. 19b). TeM extends further 5´ than gbM and covers the apparent transcriptional start site (TSS) of the long sense *LGM* mRNA (Fig. 4h), which would be expected to silence the TSS[5]. These 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. 4i), suggesting that methylation at *LGM* is evolving independently of overall gbM. GbM accessions mainly grow in southern Sweden, around 55° latitude, where UM accessions are rare (Fig. 4j and Supplementary Fig. 19c, d). EpiGWA analysis using only southern Swedish accessions found a significant association between gbM and flowering time (Fig. 4k), 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. 3h-j). Considering the much greater relative flowering delay of gbM accessions grown in Sweden compared to Spain (Fig. 4c), the prevalence of *LGM* gbM in southern Sweden (Fig. 4j) 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[68]. Sensitivities of the *AT4G15750* and *LGM* fQTLs^{gbM} to climate (Fig. 4b, c and Supplementary Fig. 18) 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[69] (latitude, longitude, temperature, precipitation, atmospheric composition, and soil properties; Supplementary Table 15). We detected 571 significant associations at FDR 0.05 between 232 genes and 115 of these variables (Supplementary Table 15). SNP-based GWA analyses showed that 76.9% of epigenetic associations do not colocalize with genetic associations (Supplementary Fig. 20a and Supplementary Table 15). GbM variation in 57 genes is associated with at least three environments (Supplementary Table 15). EpiGWA *P* values for these genes are strongly correlated for associated environments (Fig. 5a, b and Supplementary Tables 16 and 17), 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. 5c-e, Supplementary Fig. 20b-d) that do not overlap genetic associations (Supplementary Table 15). GbM in *CCS*, which mediates heat stress responses[70], is associated with summer solar insolation, with gbM epialleles much more common in high insolation environments (Fig. 5c). GbM in *CHY1*, which is involved in cold signalling and promotes freezing tolerance[71], is associated with spring minimum temperature, with gbM epialleles rare in environments where temperature drops below -4°C (Fig. 5d). GbM in *HUP9*, a regulator of flooding stress response[72], is associated with annual precipitation (Supplementary Fig. 20b). *PYR1* gbM variation is associated with soil excess salts, with high salt soils almost exclusively featuring gbM epialleles (Supplementary Fig 20c). *PYR1* is an abscisic acid (ABA) receptor[73], and ABA is a central regulator of plant salt stress responses[74]. GbM variation in the calcium sensor *SOS3*[75] associates with soil salinity and sodicity (Supplementary Fig. 20d), 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 (Supplementary Fig. 20d). These associations suggest that natural gbM variation facilitates local adaptation in native habitats.

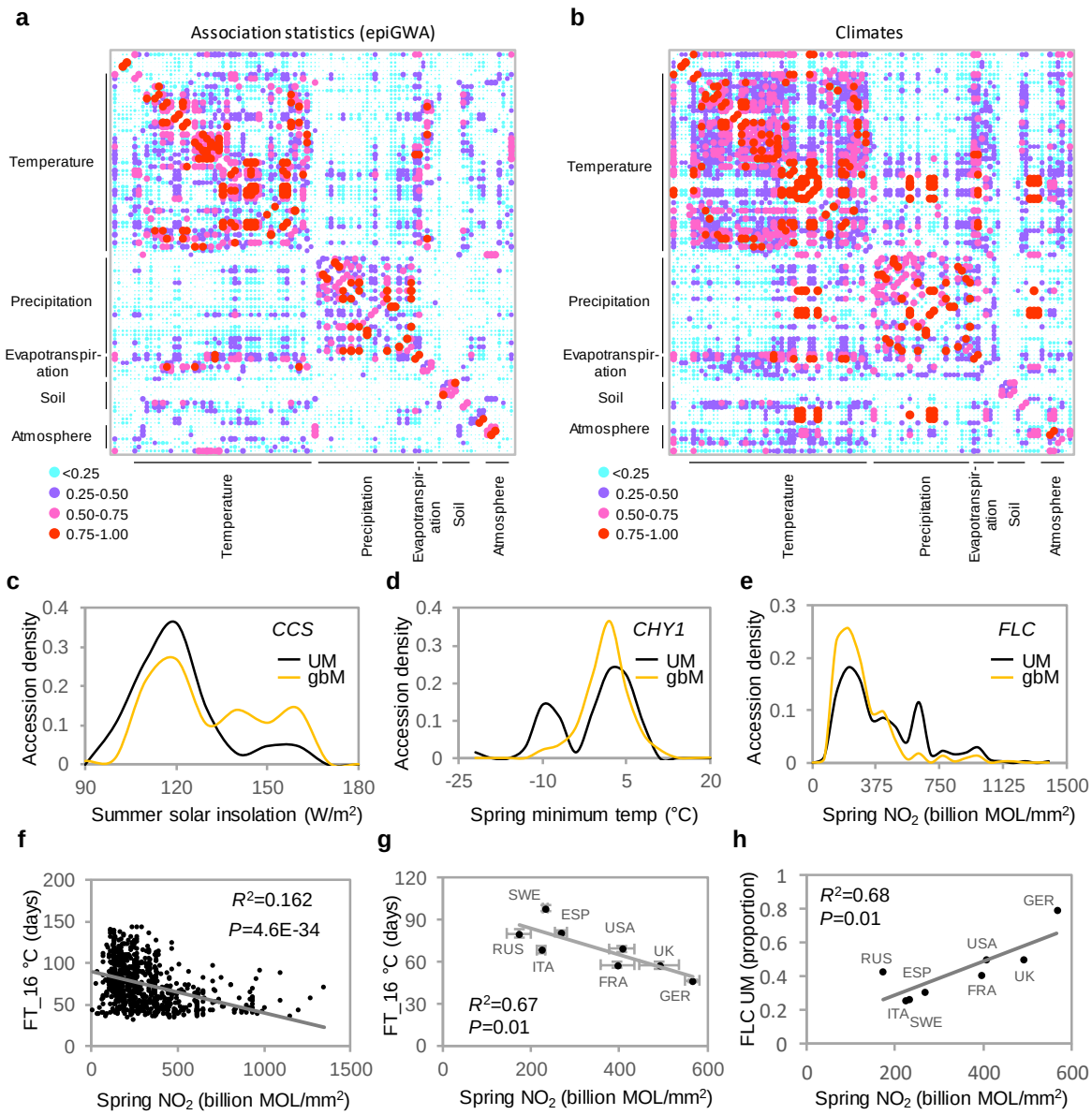

Fig. 5 | GbM variation is associated with geoclimatic variables. a, **b**, Correlation (R^2) matrices of epiGWA *P* values (**a**) and environmental variables (**b**) for 57 genes identified in at least three epiGWA analyses. See Supplementary Tables 16 and 17 for individual environment labels in order. **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.

The most striking association we discovered is between gbM in *FLC*, flowering time, and springtime concentration of NO_2, an anthropogenic atmospheric pollutant[76] (Fig. 5e-h and Supplementary Table 18). We find that 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. 5f, Supplementary Table 19). Moreover, average concentrations of NO_2 across countries show a remarkable linear correlation (R^2=0.67) with flowering time in the laboratory (Fig. 5g), indicating that earlier flowering is advantageous in higher NO_2 environments. The *FLC* UM epiallele linked with reduced expression and accelerated flowering (Fig. 3d-g) is nearly ubiquitous in environments with high NO_2 (Fig. 5e). Prevalence of the *FLC* UM epiallele in countries is also strongly correlated with NO_2 (R^2=0.68; Fig. 5h), and enrichment of the UM allele at 48-52° latitude (Fig. 3i) corresponds to Germany, which has especially high NO_2 (Fig. 5g, h). These results suggest *FLC* gbM variation is selected to adapt flowering time to atmospheric 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 trait variance as genetic polymorphism (Fig. 2e and 3b). Furthermore, the epigenetic and genetic effects on 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*) 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 like rice, wheat and maize.

DNA methylation is mutagenic[77], and its presence in coding sequences likely incurs a fitness cost[21]. 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[78–80]. The direct relationship between gbM and mRNA levels (Fig. 1e and Supplementary Fig. 2) also indicates that gbM changes should quantitatively alter gene expression, whereas most mutations are expected to be neutral or detrimental to gene function[81]. 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 species, epigenetic gbM

variation is as important as genetic variation for driving phenotypic diversity, and may account for much of the 'missing heritability'[82,83] within GWA studies.

The association between atmospheric NO_2, flowering time and *FLC* gbM (Fig. 5e-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 accessions' native environments and flowering time under laboratory conditions (Fig. 5f, g) indicates that accelerated flowering is adaptive under elevated NO_2. Natural genetic variation at *FLC* is a major determinant of flowering time[49,57,58], 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[69]. The majority of atmospheric NO_2 (>75%) is produced by recent human activities, especially the burning of fossil fuel[76]. Therefore, *Arabidopsis* populations have had to adapt to NO_2 concentrations 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 through GWA analysis. However, epigenetic gbM variation at *FLC* is significantly associated with atmospheric NO_2 (Fig. 5e, h), and therefore has likely facilitated adaptation to anthropogenic NO_2 increases. This epigenetic mechanism may be a generally important component of environmental adaptation.

Although this study investigates plant epigenetic inheritance, gbM variation is likely to influence population dynamics in many animal lineages. Direct comparisons of plant gbM with vertebrate methylation are challenging because vertebrates have broadly methylated genomes in which methylation is the default state[84]. Furthermore, methylation is extensively reprogrammed during mammalian embryogenesis and germline development[85], which precludes methylation inheritance at most loci[86–88]. However, many invertebrates, from basal cnidarians to chordates, display gbM patterns highly similar to those of plants[3–5,21]. 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[89], as would be expected if methylation is epigenetically inherited. Moreover, methylation fidelity is high in honeybee sperm compared to somatic tissues[90], and methylation patterns can be stably transmitted from fathers to daughters[91]. These features of invertebrate methylation suggest that epigenetically heritable gbM may facilitate environmental adaptation in animals. This may be especially important for species like aphids[92], 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 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 Supplementary Figures 1-20.

Supplementary Tables: This file contains Supplementary Tables 1-22.

References

1. Heard, E. & Martienssen, R. A. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* **157**, 95–109 (2014).
2. Quadrana, L. & Colot, V. Plant transgenerational epigenetics. *Annu. Rev. Genet.* **50**, 467–491 (2016).
3. Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**, 916–919 (2010).
4. Feng, S. *et al.* Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci.* **107**, 8689–8694 (2010).
5. Bewick, A. J. & Schmitz, R. J. Gene body DNA methylation in plants. *Curr. Opin. Plant Biol.* **36**, 103–110 (2017).
6. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492 (2012).
7. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220 (2010).
8. Zhang, H., Lang, Z. & Zhu, J.-K. Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* **19**, 489–506 (2018).
9. Du, J., Johnson, L. M., Jacobsen, S. E. & Patel, D. J. DNA methylation pathways and their crosstalk with histone methylation. *Nat. Rev. Mol. Cell Biol.* **16**, 519–532 (2015).
10. Qiu, Y. & Köhler, C. Mobility connects: transposable elements wire new transcriptional networks by transferring transcription factor binding motifs. *Biochem. Soc. Trans.* **48**, 1005–1017 (2020).
11. Shahzad, Z., Eaglesfield, R., Carr, C. & Amtmann, A. Cryptic variation in RNA-directed DNA-methylation controls lateral root development when auxin signalling is perturbed. *Nat. Commun.* **11**, 218 (2020).
12. Hollister, J. D. & Gaut, B. S. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* **19**, 1419–1428 (2009).

13. Hollister, J. D. *et al.* Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata. *Proc. Natl. Acad. Sci.* **108**, 2322–2327 (2011).
14. Quadrana, L. *et al.* The Arabidopsis thaliana mobilome and its impact at the species level. *eLife* **5**, e15716 (2016).
15. Stuart, T. *et al.* Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. *eLife* **5**, e20777 (2016).
16. Yan, H. *et al.* DNA methylation in social insects: how epigenetics can control behavior and longevity. *Annu. Rev. Entomol.* **60**, 435–452 (2015).
17. Kawakatsu, T. *et al.* Epigenomic diversity in a global collection of Arabidopsis thaliana accessions. *Cell* **166**, 492–505 (2016).
18. Schmitz, R. J. *et al.* Patterns of population epigenomic diversity. *Nature* **495**, 193–198 (2013).
19. Niederhuth, C. E. *et al.* Widespread natural variation of DNA methylation within angiosperms. *Genome Biol.* **17**, 194 (2016).
20. Takuno, S., Ran, J.-H. & Gaut, B. S. Evolutionary patterns of genic DNA methylation vary across land plants. *Nat. Plants* **2**, 1–7 (2016).
21. Zilberman, D. An evolutionary case for functional gene body methylation in plants and animals. *Genome Biol.* **18**, 87 (2017).
22. Lyons, D. B. & Zilberman, D. DDM1 and Lsh remodelers allow methylation of DNA wrapped in nucleosomes. *eLife* **6**, e30674 (2017).
23. Lewis, S. H. *et al.* Widespread conservation and lineage-specific diversification of genome-wide DNA methylation patterns across arthropods. *PLOS Genet.* **16**, e1008864 (2020).
24. Dixon, G. B., Bay, L. K. & Matz, M. V. Evolutionary consequences of DNA methylation in a basal metazoan. *Mol. Biol. Evol.* **33**, 2285–2293 (2016).
25. Bräutigam, K. & Cronk, Q. DNA methylation and the evolution of developmental complexity in plants. *Front. Plant Sci.* **9**, (2018).
26. Seymour, D. K. & Gaut, B. S. Phylogenetic shifts in gene body methylation correlate with gene expression and reflect trait conservation. *Mol. Biol. Evol.* **37**, 31–43 (2020).

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

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

29. Muyle, A., Ross-Ibarra, J., Seymour, D. K. & Gaut, B. S. Gene body methylation is under selection in Arabidopsis thaliana. *bioRxiv* 10.1101/2020.09.04.283333 (2020).

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

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

32. Wedd, L. & Maleszka, R. DNA methylation and gene regulation in honeybees: from genome-wide analyses to obligatory epialleles. in *DNA Methyltransferases - Role and Function* (eds. Jeltsch, A. & Jurkowska, R. Z.) 193–211 (Springer International Publishing, 2016). doi:10.1007/978-3-319-43624-1_9.

33. Bewick, A. J. *et al.* Dnmt1 is essential for egg production and embryo viability in the large milkweed bug, Oncopeltus fasciatus. *Epigenetics Chromatin* **12**, 6 (2019).

34. Bewick, A. J. *et al.* On the origin and evolutionary consequences of gene body DNA methylation. *Proc. Natl. Acad. Sci.* **113**, 9111–9116 (2016).

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

36. Cortijo, S. *et al.* Mapping the epigenetic basis of complex traits. *Science* **343**, 1145–1148 (2014).

37. Johannes, F. *et al.* Assessing the impact of transgenerational epigenetic variation on complex traits. *PLOS Genet.* **5**, e1000530 (2009).

38. Furci, L. *et al.* Identification and characterisation of hypomethylated DNA loci controlling quantitative resistance in Arabidopsis. *eLife* **8**, e40655 (2019).

39. Schmid, M. W. *et al.* Contribution of epigenetic variation to adaptation in Arabidopsis. *Nat. Commun.* **9**, 4446 (2018).

40. Catoni, M. *et al.* DNA sequence properties that predict susceptibility to epiallelic switching. *EMBO J.* **36**, 617–628 (2017).

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

42. Atwell, S. *et al.* Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature* **465**, 627–631 (2010).

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

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

45. Stroud, H. *et al.* Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat. Struct. Mol. Biol.* **21**, 64–72 (2014).

46. Ågren, J., Oakley, C. G., Lundemo, S. & Schemske, D. W. Adaptive divergence in flowering time among natural populations of Arabidopsis thaliana: Estimates of selection and QTL mapping. *Evolution* **71**, 550–564 (2017).

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

48. Taylor, M. A. *et al.* Large-effect flowering time mutations reveal conditionally adaptive paths through fitness landscapes in Arabidopsis thaliana. *Proc. Natl. Acad. Sci.* **116**, 17890–17899 (2019).

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

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

51. Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).

52. Zhang, Z. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355–360 (2010).

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

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

55. Grimm, D. G. *et al.* EasyGWAS: a cloud-based platform for comparing the results of genome-wide association studies. *Plant Cell* **29**, 5–19 (2017).

56. Johanson, U. *et al.* Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science* **290**, 344–347 (2000).

57. Hepworth, J. *et al.* Natural variation in autumn expression is the major adaptive determinant distinguishing Arabidopsis FLC haplotypes. *eLife* **9**, e57671 (2020).

58. Li, P. *et al.* Multiple FLC haplotypes defined by independent cis-regulatory variation underpin life history diversity in Arabidopsis thaliana. *Genes Dev.* **28**, 1635–1640 (2014).

59. Shindo, C. *et al.* Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. *Plant Physiol.* **138**, 1163–1173 (2005).

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

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

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

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

64. Corbesier, L. *et al.* FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. *Science* **316**, 1030–1033 (2007).

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

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

67. Stolc, V. *et al.* Identification of transcribed sequences in Arabidopsis thaliana by using high-resolution genome tiling arrays. *Proc. Natl. Acad. Sci.* **102**, 4453–4458 (2005).

68. Ågren, J. & Schemske, D. W. Reciprocal transplants demonstrate strong adaptive differentiation of the model organism Arabidopsis thaliana in its native range. *New Phytol.* **194**, 1112–1122 (2012).

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

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

71. Dong, C.-H. *et al.* Disruption of Arabidopsis CHY1 reveals an important role of metabolic status in plant cold stress signaling. *Mol. Plant* **2**, 59–72 (2009).

72. Lee, S. C. *et al.* Molecular characterization of the submergence response of the Arabidopsis thaliana ecotype Columbia. *New Phytol.* **190**, 457–471 (2011).

73. Cutler, S. R., Rodriguez, P. L., Finkelstein, R. R. & Abrams, S. R. Abscisic Acid: Emergence of a Core Signaling Network. *Annu. Rev. Plant Biol.* **61**, 651–679 (2010).

74. Yu, Z. *et al.* How plant hormones mediate salt stress responses. *Trends Plant Sci.* **25**, 1117–1130 (2020).

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

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

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

78. Graaf, A. van der *et al.* Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proc. Natl. Acad. Sci.* **112**, 6676–6681 (2015).

79. Schmitz, R. J. *et al.* Transgenerational epigenetic instability Is a source of novel methylation variants. *Science* **334**, 369–373 (2011).

80. Becker, C. *et al.* Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. *Nature* **480**, 245–249 (2011).

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

82. Brachi, B., Morris, G. P. & Borevitz, J. O. Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* **12**, 232 (2011).

83. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).

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

85. Zeng, Y. & Chen, T. DNA Methylation Reprogramming during Mammalian Development. *Genes* **10**, (2019).

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

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

88. Kazachenka, A. *et al.* Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. *Cell* **175**, 1259-1271.e13 (2018).

89. Wang, X., Werren, J. H. & Clark, A. G. Allele-Specific Transcriptome and Methylome Analysis Reveals Stable Inheritance and Cis-Regulation of DNA Methylation in Nasonia. *PLOS Biol.* **14**, e1002500 (2016).

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

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

92. Mathers, T. C. *et al.* Sex-specific changes in the aphid DNA methylation landscape. *Mol. Ecol.* **28**, 4228–4241 (2019).