Science

RESEARCH ARTICLES

Cite as: P. Baduel *et al.*, *Science* 10.1126/science.ady3475 (2025).

Transposable elements are vectors of recurrent transgenerational epigenetic inheritance

Pierre Baduel[1]*, Louna De Oliveira[1], Erwann Caillieux[1], Grégoire Bohl-Viallefond[1,2], Ciana Xu[1], Mounia El Messaoudi[1], Aurélien Petit[1], Maëva Draï[1], Matteo Barois[1,3], Vipin Singh[1], Alexis Sarazin[1,4], Felipe K. Teixeira[1,5], Martine Boccara[1,6], Elodie Gilbault[7], Antoine de France[1], Leandro Quadrana[1,8], Olivier Loudet[7], Vincent Colot[1]*

[1]Institut de Biologie de l'École Normale Supérieure, Université Paris Sciences et Lettres (PSL), UMR 8197 CNRS, U1024 INSERM, Paris, France. [2]Gregor Mendel Institute of Molecular Plant Biology (GMI) of the Austrian Academy of Sciences, Vienna, Austria. [3]Evo-Eco-Paleo (EEP) Université de Lille, UMR 8198 CNRS, Villeneuve d'Ascq, France. [4]Department of Biology, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland. [5]Department of Genetics, University of Cambridge, Cambridge, UK. [6]Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France. [7]Université Paris-Saclay, INRAE, AgroParisTech, Institute Jean-Pierre Bourgin for Plant Sciences (IJPB), Versailles, France. [8]Institute of Plant Sciences of Paris-Saclay (IPS2), Université Paris-Saclay, Université Paris Cité, Université Evry Val d'Essonne, U1403 INRAE, UMR 9113 CNRS, Gif-sur-Yvette, France.

***Corresponding author. Email: pierre.baduel@ens.psl.eu (P.B.); vincent.colot@ens.psl.eu (V.C.)**

DNA methylation loss at transposable elements (TEs) can affect neighboring genes and be epigenetically inherited in plants, yet the determinants and significance of this additional system of inheritance are unknown. Here, we demonstrate in *Arabidopsis thaliana* that transgenerational stability of experimentally-induced hypomethylation at TE loci is constrained by small RNAs derived from related copies. Using data from >700 strains collected worldwide, we uncover similar and recurrent hypomethylation at hundreds of these TE loci, often near genes. Most natural epivariants we tested can be inherited without DNA sequence changes and are therefore bona fide epialleles, although genetic factors modulate their recurrence or persistence. Epiallelic variants often cause gene expression changes and may be targets of selection, thus revealing their contribution to heritable phenotypic variation in nature.

Transposable elements (TEs) are typically targeted by cytosine DNA methylation in plants and mammals (*1*). Loss of this epigenetic modification can lead to increased transposition of mobile TEs, but it may also facilitate chromosome rearrangements or alter the expression of neighboring genes even in the case of decayed copies (*1–3*). Moreover, because plants do not extensively reset DNA methylation through reproduction (*4–6*), unlike mammals (*7*, *8*), stochastic or artificially induced loss of this modification at TEs in plants can be inherited across multiple generations as epialleles, i.e., independently of DNA sequence changes. Indeed, a number of such epialleles have been reported, albeit often as a result of artificial manipulations (*9–12*). Some have striking consequences on fitness-related traits, including fruit development in oil palm (*13*) or the onset of flowering in *A. thaliana* (*14–16*), thus raising the possibility that TE-mediated epiallelic variation has broad phenotypic consequences.

While undisputed in plants (*10*), TE-mediated transgenerational epigenetic inheritance (TEI) and its biological significance remain poorly understood (*11*, *12*, *17*). This is due in large part to the difficulty of establishing experimentally whether DNA methylation variation can be inherited independently of DNA sequence polymorphisms (*18*). *Arabidopsis thaliana* offers a way forward however, thanks to several key resources and the deep understanding developed in this model species of the molecular players involved in DNA methylation of TEs. In particular, the Snf2-like chromatin remodeler DDM1 (DECREASE IN DNA METHYLATION 1), was shown to be pivotal for the maintenance of methylation at the CG, CHG and CHH (where H=A,C,T) sites of TEs by allowing DNA methyltransferases access to heterochromatin (*16*, *19–22*). *ddm1*-induced loss of DNA methylation can be stably inherited in the absence of the mutation (*23*) most likely because of disruption to the interplay between DDM1 and histone variants (*24*).

Here, we used a previously generated population of isogenic *ddm1*-derived epigenetic recombinant inbred lines (epiRILs) (*25–27*) to investigate the inheritance of DNA hypomethylation at thousands of TE loci across the genome and to characterize the molecular underpinnings of the differences observed between TE loci as well as between epiRILs. We then combined the knowledge gained using the epiRILs with genome sequence, DNA methylome, transcriptome, phenotypic and geo-bioclimatic data available for >700 *A. thaliana* strains from across the world (*28–31*), to interrogate the scope, dynamics, and biological implications of TE-mediated TEI in nature.

Inheritance patterns of *ddm1*-induced hypomethylation differ between TEs

To elucidate the molecular underpinnings of TE-mediated TEI in *A. thaliana*, we first obtained DNA methylome data at

single nucleotide resolution for 120 *ddm1*-derived epiRILs (*25*) taken at generation F9 as well as for the *ddm1* and WT parental lines. Thanks to their symmetrical nature, CG and CWG (where W=A or T) sites offer the most robust estimates of methylation levels (fig. S1A). We identified 25,462 regions where methylation is reduced by at least 20% at these sites in *ddm1* and that exhibit similar methylation states across sites within each epiRIL (data S1 and S2). Using stringent criteria, including the presence of at least 5 differentially methylated CGs within a given differentially methylated region (DMR), we retained 7,023 DMRs (data S3), hereafter designated as TE-DMRs, that are fully contained within TE annotations. Importantly, TE-DMRs do not belong to TE copies that are mobile in the epiRILs (*32*), thus alleviating the risk of mismapping of sequence reads in this setting. To determine the parental origin of each TE-DMR in the epiRILs, we "epihaplotyped" them (Fig. 1A and fig. S1, G and H) using as genetic markers 1,092 of the 25,462 DMRs we defined initially. Markers were chosen among those that segregate the two parental methylation states in a manner compatible with Mendelian inheritance and so as to be regularly spaced along the five chromosomes (data S4). Finally, for each TE-DMR we compared its average CG methylation level in each epiRIL with that in the two parental lines, because CG sites exhibit the largest differences of methylation between WT and *ddm1*, and we used these comparisons to attribute one of three possible DNA methylation states of TE-DMRs in the epiRILs: WT-like, *ddm1*-like or intermediate (fig. S2A).

While inheritance of WT methylation is near systematic in the epiRILs (Fig. 1B; see supplementary text, note 1, for exceptions), this is not the case for *ddm1*-induced hypomethylation, consistent with low-resolution DNA methylation data (*25*, *26*, *33*). Specifically, 89.1% of *ddm1*-derived TE-DMRs exhibited at least one instance of WT-like or intermediate DNA methylation in the epiRILs, indicative of full or partial reversion, respectively (Fig. 1B). Moreover, DNA methylome data produced from a random subset of 12 epiRILs propagated to generation F21 revealed that once restored in F9, WT-like DNA methylation is almost always (99%) stably transmitted to F21 (Fig. 1C) and that intermediate DNA methylation at F9 usually translates into WT-like levels by F21 (Fig. 1C), in agreement with the stability of WT methylation and the known progressivity of reversion (*33*), respectively (see supplementary text, note 2).

Reversion is conspicuous but far from systematic at generation F9, except for a small number (395) of TE-DMRs. To account for the wide range of reversion frequency at this generation, we defined five categories of TE-DMRs: two extreme categories for TE-DMRs that never (r-0%) or always (r-100%) revert as well as three categories of non-systematically reverting TE-DMRs, r-low, r-mid, and r-high (Fig. 1D). However, because some r-0% TE-DMRs have reverted by generation F21 (Fig. 1, B and C, and fig. S2C) *ddm1*-induced hypomethylation is unlikely to be inherited indefinitely in the absence of selection even at TE-DMRs of this category.

RdDM-mediated reversion is driven by non-allelic interactions between related TEs

RNA-directed DNA methylation (RdDM) is required for reversion (*33*) and involves 23-24nt small RNAs (sRNAs) (*19*). To determine their origin, we sequenced sRNAs from the two parental lines and 10 epiRILs. In keeping with earlier findings (*33*), the abundance of 23-24nt sRNAs perfectly matching TE-DMRs of the r-0% or r-low categories is drastically reduced in *ddm1* compared to WT (Fig. 2A and fig. S3A). Moreover, we find that the positive correlation between reversion in the epiRILs and 23-24nt sRNAs abundance in *ddm1* is strongest for multiple-mappers (Fig. 2A and fig. S3B), whose levels also correlate strongly with mCHH levels in this mutant background (Fig. 2B and fig. S3C). As residual CHH methylation in *ddm1* depends almost entirely on RdDM (*21*, *33*), these findings reveal a predominant role for multiple-mapping 23-24nt sRNAs in guiding RdDM activity and thus reversion. Consistent with this conclusion, reversion in the epiRILs is associated with a higher abundance of multi- rather than single-mapping 23-24nt sRNAs (Fig. 2C) and is more frequent in epiRILs where related TE-DMRs have also reverted (Fig. 2D). In fact, reversion occurs in 20% of cases without any regain or even in the absence of single-mapping 23-24nt sRNAs (fig. S3, D and E).

To identify the potential sources of multi-mapping 23-24nt sRNAs guiding reversion, and because RdDM can be driven by 23-24nt sRNAs with imperfect homology (*34*), we examined all 23-24nt sRNAs that can be aligned with up to one mismatch to target TE-DMRs. As expected, most (92%) of these sRNAs map perfectly to TEs that belong to the same TE family or superfamily as the target TE-DMR (fig. S4A) and their number also correlates positively with reversion (Fig. 2, E and F, and fig. S4, B and C). Moreover, the proportion of TE-DMRs that belong to the same TE family and have reverted in a given epiRIL correlates positively with the number of related TE copies inherited from WT in that epiRIL (Fig. 2G). These results are in line with observations made using transgenes that efficient RdDM-targeting requires large amounts of 23-24nt sRNAs (*35–39*) and demonstrate that 23-24nt sRNAs derived from ectopic, WT-inherited sources are the main drivers of reversion.

To determine if specific TE copies, when inherited from WT, play predominant roles in driving family-wide reversion of TE-DMRs, we performed QTL mapping for each TE family using as trait the average CG methylation level of all non-systematically reverting TE-DMRs (i.e., r-low, r-mid, and r-high) of that family that are inherited from *ddm1*. To maximize the number of informative epiRILs, we considered the 105 TE

families with at least ten non-systematically reverting TE-DMRs. In most (99) cases, we detected either one or two major QTL^{epi}, which explain on average 13% of variation in family-wide reversion and in a manner consistent with the WT-derived genotype driving reversion (Fig. 2H and fig. S4, D and E). These QTL^{epi} tend to overlap with TE-rich pericentromeres and almost all of them (92%) encompass one or several of the potential sources of *trans*-acting sRNAs we identified (fig. S4D), a strong indication of their role in reversion. Conversely, the *ddm1*-derived QTL^{epi} do not affect WT-derived TE-DMRs, except in rare cases that mainly concern TE families with known *trans*-demethylation activity (*VANDAL* and *ATENSPM* families; fig. S4, D to F, and supplementary text, note 1).

The QTL mapping results imply that the hypomethylated epiallelic state at non-systematically reverting TE-DMRs should be stable in epiRILs where the main sources of *trans*-acting 23-24nt sRNAs are also inherited from *ddm1*. To verify that this is the case, we grew >75 siblings of two randomly chosen F9 epiRILs (#92 and #216) that still harbor the hypomethylated epiallele at some non-systematically reverting TE-DMRs. We selected five TE-DMRs that belong to different TE families and have distinct reversion frequencies for analysis. Importantly, three TE-DMRs belong to TE families for which we identified at least one QTL^{epi} (fig. S4D) and hypomethylation co-segregates with the *ddm1*-derived top QTL^{epi} in each case (fig. S4G). Using targeted DNA methylation analysis, we detected either zero or one reversion events among F9 siblings, even for the two r-high TE-DMRs (fig. S4H). These results demonstrate that hypomethylation can be stably transmitted when key genomic intervals are *ddm1*-derived and thus do not produce small RNAs in sufficient numbers to guide reversion of matching targets.

ddm1-induced transcription favors the inheritance and RdDM limits the occurrence of TE hypomethylation

Although reversion is usually coordinated at the TE-family level (Fig. 2C), 369 TE-DMRs remained in the hypomethylated epiallelic state in epiRILs where >75% of the related TE-DMRs have reverted (fig. S5A). These persistently hypomethylated or "stray" TE-DMRs tend to have higher levels of sequence divergence with their closest relative and fewer close relatives altogether (fig. S5, B and C), in agreement with the known reduction of RdDM efficiency when source and target diverge in sequence (*40*). Furthermore, stray TE-DMRs are enriched at or near RNA Pol II transcription units that are activated in *ddm1* (EPigenetically Induced Consensus tAg clusTers or EPICATs (*41*); fig. S5, D to G, and supplementary text, note 3), in keeping with the known inhibitory effect of Pol II transcription on RdDM (*35*, *42*, *43*). However, whether the absence of reversion at stray TE-DMRs is due primarily to gains of transcriptional activity, to reduced RdDM efficiency because of sequence divergence, or to these TE-DMRs being specifically targeted for active demethylation remains to be determined.

To explore if RdDM targeting may also affect the spontaneous occurrence of the hypomethylated state of TE-DMRs, we considered the 106 TE-DMRs with *ddm1*-like loss of parental WT methylation in one or two epiRILs only and that do not belong to TE families with known *trans*-demethylation activity (data S8). To increase our sample size, we exploited methylome data independently obtained at generation F9 for a minimally overlapping set of 169 epiRILs (*44*) and identified in this set another 114 TE-DMRs with similar spontaneous *ddm1*-like hypomethylation (data S8). The majority (65%) of the 207 TE-DMRs we identified in total belong to the r-0% and r-low categories (fig. S5H) and spontaneous hypomethylation is far less frequent (2.3-fold reduction) when all of the potential sources of *trans*-acting 23-24nt sRNAs are WT-derived (fig. S5I). Thus, we conclude that RdDM not only limits the inheritance, but also the spontaneous occurrence of epiallelic variation at TE-DMRs. In contrast, the presence of EPICATs does not appear to have any detectable impact on spontaneous hypomethylation (fig. S5J).

TE epivariants are common in nature and most are recurrent, bona fide epialleles

Having determined the potential for epiallelic variation at thousands of TEs across the genome using the epiRILs, we then asked how this potential translates in nature. To this end, we leveraged WGBS-seq data available from leaf tissues for 720 natural strains (*29*) and calculated for each of the 7023 experimentally-defined TE-DMRs its average CG methylation level in each of the strains where it is present (*45*, *46*). TE-DMRs are generally part of TE copies that are fixed or nearly so at the species level (fig. S6, A and B). However, because of the stringent filtering we applied to ensure robust measures of DNA methylation, we retained for analysis only about half of the TE-DMRs present in any given strain (fig. S6C and data S9; see supplementary text, note 4, for discussion of these and other biases).

In keeping with heavy and co-extensive DNA methylation being the default state for most TEs in *A. thaliana* (*29*, *47*), the vast majority (85%) of TE-DMRs have high average CG methylation across all informative strains (Fig. 3A). Nevertheless, 1,068 TE-DMRs exhibit severe, *ddm1*-like CG hypomethylation in at least one strain (data S10). Almost all of these natural epivariants likely result from rare and rapid loss of DNA methylation, given that intermediate levels of average CG methylation are seldom observed and that they generally have low population frequencies (Fig. 3A and fig. S6D). DNA methylation loss extends to CHG and CHH sites, as in *ddm1* (fig. S6, E and F), although the exact boundaries of

hypomethylation may differ (e.g., Fig. 3B). Genome assembly of 20 genetically diverse strains using long-read sequencing (data S11) confirmed that, thanks to our stringent filtering, most (>87%) of natural epivariants we called in each strain are at the same location as in the reference genome and are not associated with significant local rearrangements (e.g., Fig. 3B and supplementary text, note 5, for 100 other examples and detailed discussion). Moreover, remapping of WGBS-seq reads on the ONT genome assemblies validates *ddm1*-like hypomethylation in 97% of cases (Fig. 3C) and so does also direct DNA methylation calls on these assemblies (e.g., Fig. 3B and supplementary text, note 5), which together exclude major mismapping issues (*48*).

Naturally epivariable TE-DMRs are strongly skewed toward the r-0% and r-low categories (Fig. 3D), suggesting that the inheritance properties of *ddm1*-induced hypomethylated TE-DMRs extend to their natural counterparts. To confirm that this is the case and that natural epivariants are therefore bona fide epialleles, we first carried out genome-wide association studies (GWAS) of natural epivariation at each of the 490 TE-DMRs that have sufficient epivariant counts. Most significant associations with single nucleotide polymorphisms (SNPs) are in *cis* (within 20kb; fig. S7, A and B), yet none of these are systematic (fig. S7C), thus ruling out a strict dependence of natural epivariation on neighboring SNPs.

To obtain direct evidence that sequence polymorphisms in *trans* do not dictate epivariation either, we performed linkage mapping using the population of recombinant inbred lines (RILs) previously built between Cvi-0 and Col-0 (*49*). This RIL population is particularly informative in this respect, as Cvi-0 harbors the second largest number of epivariants (113 in total; fig. S6C) among natural strains and also because almost all (110) of the corresponding TE-DMRs are hypomethylated in at least one other strain (data S13). Targeted (Fig. 3E) as well as whole-genome (fig. S7E) segregation analyses indicate that the two parental DNA methylation states almost systematically segregate in the RILs together with flanking sequence markers, and thus independently from sequence variants in *trans* (see supplementary text, note 6, for the one exception). In addition, many of the Cvi-0-derived TE-DMRs have regained dense, high level DNA methylation in at least one RIL (Fig. 3E and fig. S7E), which confirms that they are also not strictly conditioned by polymorphisms in *cis*.

Epivariants present in two or more strains could reflect identity by descent or recurrent occurrence. To distinguish between these two possibilities, we considered the 780 TE-DMRs with several instances of epivariation in nature and calculated the genetic relatedness (kinship) of the corresponding strains. While almost all (96%) TE-DMRs show hypomethylation in at least two strains that are highly divergent (genome-wide kinship <0.25), only half of them do so in two or more closely related strains (genome-wide kinship >0.75; fig. S8A). Furthermore, local haplotyping indicates that for the majority (66%) of TE-DMRs, epivariants are carried by at least two distinct alleles (fig. S8, B to D). Thus, the population frequencies of natural epivariants are determined in large part by recurrence rather than long-term inheritance, consistent with their experimental epiallelic counterparts being not indefinitely stable.

Natural epivariants are prevalent near genes

Unlike in the epiRILs, natural hypomethylated epivariants are usually few in any given strain (49 on average; fig. S6C) and tend to be either tightly clustered (within 10kb) and correlated between strains, or else dispersed and uncorrelated (Fig. 4A and fig. S8, E and F). In addition, naturally epivariable TE-DMRs are usually less pericentromeric (Fig. 4B), hence closer to genes than the complete set of experimental TE-DMRs (Fig. 4C and fig. S8G). Moreover, the population frequency of natural epivariants correlate positively with gene proximity, suggesting either that it facilitates their occurrence or inheritance, or else that epivariation is counterselected away from genes. Supporting a positive effect of gene proximity, we identified through GWAS the gene *EARLY FLOWERING 8* (*ELF8*) as the single major genetic determinant of the prevalence of natural epivariation within strains (Fig. 4, D to F). *ELF8* encodes a component of the highly conserved RNA polymerase-associated factor 1 (Paf-1) complex, and previous work showed that when marked by the histone H3 modification H3K4me2/me3 (*50*), which in plants requires Paf1 (*51*), genes hinder RdDM targeting of neighboring TEs. Consistent with these findings, strains with the non-reference, overexpressing allele of *ELF8* (hereafter designated *ELF8'*; see fig. S9, A to H, and supplementary text, note 7, for functional analysis) tend to harbor a higher number of natural epivariants (Fig. 4, E and G) and this trend is most pronounced for TE-DMRs that are located within 2kb of genes (fig. S9I). However, examination of the Cvi-0 x Col-0 RILs, which segregate the two *ELF8* alleles, indicate that increased ELF8 activity is insufficient on its own to promote the occurrence of epivariants or to stabilize their inheritance (Fig. 3E and fig. S9, J and K; see supplementary text, note 8). Thus, additional factors are likely involved in the association observed in nature.

To explore what these additional factors might be, we used a generalized linear model (GLM) and performed a stepwise selection of ten parameters, including allelic variation at *ELF8* and distance to genes, on the basis of their added explanatory power. Results of the GLM indicate that TE copy-number and allelic variation at *ELF8* are the two main parameters, which explain 5.5% and 3.5% of the variance in the proportion of TE-DMRs that are hypomethylated in a given strain, respectively (Fig. 4H). The negative correlation with

TE copy-number is consistent with RdDM strength increasing with the number of 23-24nt sRNA sources (*52*, *53*) and the role of RdDM in limiting both the inheritance and spontaneous occurrence of hypomethylation at TE-DMRs in the epiRILs (Figs. 2 and 3 and figs. S4 and S5). Indeed, our GLM revealed that the weak allele of the RdDM gene *NRPE1* (*46*, *54*) mitigates the negative effect of TE copy-number on epivariation. It also confirmed that the impact of gene proximity is stronger in strains carrying the *ELF8'* rather than the $ELF8^{ref}$ allele (fig. S9I). Together, these findings highlight the pivotal and antagonistic roles of RdDM and Paf1 activity in shaping the prevalence of TE-mediated epiallelic variation in nature.

Natural epialleles impact gene expression

To investigate the functional impact of natural epialleles, we first analyzed RNA-seq data available for >700 strains (*29*) and compared for each naturally epivariable TE-DMR the expression levels of the closest genes between strains with or without epivariation. Although we probed just one organ (leaf) and one condition (standard growth), we observed major (≥2-fold) gene expression changes in *cis* for 25% of TE-DMRs (fig. S10A). These changes were skewed toward up-regulation, particularly when the TE-DMRs overlapped with genes or were located immediately upstream (≤500bp) of them (fig. S10, A and B). To distinguish between cause and consequence, without the confounding effects of DNA sequence variants, we produced RNA-seq data from the 10 epiRILs that we used for the sRNA studies and asked whether the experimental counterparts of natural epivariants associated with changes in gene expression cause similar expression changes in the epiRILs. Comparisons could be made for 99 TE-DMRs in total. For 58 of them the expression changes between epiRILs with contrasted epiallelic states were congruent, both in direction and amplitude, with those in nature (Fig. 5A and fig. S10C) thus explaining 24% on average of the variation in expression levels observed between natural strains (data S16). In addition, congruence concerns 30 of the 32 TE-DMRs located within or upstream (≤500bp) of genes (Fig. 5A and fig. S10C) and for 14 of the 17 TE-DMRs with a reversion event in the 10 epiRILs sampled, WT expression is regained too (Fig. 5B and fig. S10D). Thus, the majority of natural epiallelic variants associated with gene expression differences in *cis* are causal and causality is almost systematic for TE-DMRs within or upstream of genes.

Many of the 60 genes for which we could prove a causal effect of TE-epiallelic variation on their expression have no known function and the phenotypic consequences of their misregulation remain to be determined. Among the few genes with known function (data S16), two were already reported to be affected by the epigenetic state of the TE sequences they contain, *QQS* (*55*) and *RPP7* (*56*). While the phenotypic impact of the epiallelic up-regulation of *QQS* (fig. S10E) remains to be determined, the inhibition of splicing caused by the loss of heterochromatin at the intronic TE present in *RPP7* (fig. S10F), was shown to considerably reduce RPP7-mediated defense against pathogens (*57*). Thus, epiallelic variation at *RPP7* may be deleterious in natural settings. Further supporting this interpretation, we only observed *RPP7* epivariation once among informative strains (data S9), despite the TE-DMR being of the r-low category. A third gene, *RAD3* (*AT2G05635*; fig. S11A), which encodes a DEAD helicase, also stands out because it lies within the major QTL^{epi} interval identified for primary root length in the epiRILs (*27*) and it has been implicated in natural variation for root growth under conditions of acid mine drainage (*58*). Epivariants of *RAD3* are relatively frequent in nature (95 strains) and biogeoclimatic data (*31*) indicate that they are underrepresented in locations with high NO_2 emission levels in summer (Fig. 5B and fig. S11B). Because NO_2 emissions correlate positively with levels of soil acidification (*59*), it is tempting to speculate that hypomethylation of the *RAD3* TE-DMR, which results in more abundant and longer transcripts (Fig. 5C and fig. S11A) and is associated with reduced growth of the primary root (fig. S11C), is selected against in habitats with acidic soils.

Natural epivariation may be selected to facilitate stress responses

Given that naturally epivariable TE-DMRs are particularly enriched near stress-responsive genes (fig. S11D and data S17), we asked if epivariation could influence gene expression specifically under stress conditions. For this purpose, we focused on the r-mid TE-DMR that is located ~900bp downstream of the cold- and drought-stress responsive gene *HVA22E* (*AT5G50720*; fig. S12A). In all of the epiRILs that we investigated, *HVA22E* expression is similarly low in the absence of stress, irrespective of the DNA methylation status of the TE-DMR (Fig. 5, D and E). In contrast, *HVA22E* induction by cold is on average 5-fold higher in the epiRILs where the TE-DMR is inherited in the hypomethylated state from *ddm1* rather than in those where it is inherited from WT, and this higher level of induction is lost upon reversion (Fig. 5E). Furthermore, cold treatment does not result in a loss of DNA methylation in the case of the WT-derived TE-DMR (fig. S12B). Altogether, these results indicate that the hypomethylated epiallele at *HVA22E* enables plants to enhance their response to cold stress, without affecting their basal expression.

Despite frequent (66%) reversion in the epiRILs, hypomethylation at the TE-DMR downstream of *HVA22E* is relatively common (24%) in nature, which suggests the presence of stabilizing genetic or genomic modifiers, or positive selection in this setting. Although discriminating between these possibilities can be challenging, we exploited the fact that the

TE-DMR is in the hypomethylated state in Cvi-0 to assess its stability in the Cvi-0 x Col-0 RIL population. We analyzed 88 RILs in total and found that the Cvi-0 allele remained hypomethylated in 72% of the 47 RILs that harbor it (Fig. 5F). However, this allele regained high levels of DNA methylation in 17% of cases and reversion co-segregates in seven of eight of these cases with the Col-0 haplotype at the QTL^{epi} interval we identified for this TE family in the epiRILs (fig. S12C). In addition the Col-0 allele of *HVA22E* loses DNA methylation over the TE-DMR in 29% of the 41 RILs that harbor it. Taken together, these results indicate that the Cvi-0 genetic background stabilizes as well as favors hypomethylation of the *HVA22E* TE-DMR and that conversely, the Col-0 QTL^{epi} promotes reversion.

Using biogeoclimatic data (*31*), we observed that hypomethylation at *HVA22E* is preferentially found in habitats with high summer frost and low summer precipitation (Fig. 5G and fig. S12, D and E). Furthermore, analysis of phenotypic data compiled in the AraPheno database indicates that strains with hypomethylation at the locus tend to have smaller stomata (fig. S12, F and G), a cell-type in which *HVA22E* is strongly induced by drought (*60*). These observations prompted us to determine if epiallelic variation at *HVA22E* could impact the ability of plants to respond to drought. Specifically, we used a high-throughput phenotyping platform (*61*) to monitor the growth of 112 epiRILs subjected to mild drought or grown in well-watered conditions. These included three lines with a *ddm1*-derived epiallele still in the hypomethylated state and 14 lines where the epiallele had reverted. While all epiRILs were equally affected by drought during the first seven days of growth under this condition, the epiRILs with the hypomethylated epiallele resumed growth faster than all of the other epiRILs (Fig. 5H, fig. S12H, and data S19), indicating that it could act as a selectable engine of enhanced response to stress.

Discussion

Our study provides a comprehensive assessment of the determinants and biological significance of heritable DNA methylation variation at TEs in plants.

The discovery that *trans*-acting sRNAs are key for reversion to the methylated state reveals the existence of a broad epigenetic surveillance mechanism in plants, that is fed by and targeted toward TEs, and which modulates their ability to serve as vectors of TEI. That pericentromeric TEs are the main source of *trans*-acting sRNAs is consistent with methylome data obtained in epigenetic F1 hybrids (*62*) and directly supports the proposition that these sRNAs, which have little role in *cis*, mainly serve to guide DNA methylation of related copies located on chromosome arms (*63–66*). "Classical" paramutation, which involves epigenetic interactions between alleles or epialleles of the same locus (*67*), would be an extreme manifestation of this surveillance mechanism and therefore the exception rather than the rule, consistent with previous considerations (*35*, *68–70*) as well as results of inter-strain crosses (this study and (*71*)). Thus, we propose to view the genome as an "epigenetic biome" where the epiallelic potential of each TE locus is shaped via sRNA-mediated interactions involving multiple related TE loci. This view has important practical implications, notably for crop improvement using epigenome editing (*72*) as heritable loss of DNA methylation at a given locus may be difficult to achieve when matching 23-24nt sRNAs produced in *trans* are abundant.

Despite this surveillance mechanism, epiallelic variation at TEs is relatively abundant in nature, in large part because it recurrently affects fixed or nearly fixed TE copies and we have identified several genetic factors that modulate its prevalence in strains. Specifically, while RdDM reduces the possibility of TE-mediated TEI, gene proximity and increased expression of *ELF8*, which encodes a component of the evolutionary conserved Paf1 complex, are the two main genetic factors that favor it. These findings are in line with results in *Schizosaccharomyces pombe* (*73*, *74*) and *A. thaliana* (*50*) indicating respectively that Paf1 and the active histone marks it helps deposit inhibit small RNA-mediated epigenetic silencing. Thus, in addition to its many roles in facilitating Pol-II transcription (*51*), Paf1 may also help safeguard genes from the silencing of neighboring TEs.

Natural epialleles can alter the expression of neighboring genes, notably in response to stress, and their presence correlates with specific environments in some cases. These findings suggest that despite their lower stability compared to DNA sequence variants, natural epialleles can be targets of selection, consistent with previous evidence (*27*, *75*). Given their recurrence and singular inheritance properties, they may contribute uniquely to local adaptation, although teasing apart the contribution of epialleles and of DNA sequence variants in natural populations remains challenging. In addition, a key question is whether epiallelic variation may be facilitated or stabilized in specific environments. Indeed, evidence supporting such effects is still lacking (this study and (*76*)), despite well-documented effects of biotic or abiotic stresses on the heritable loss of methylation at individual cytosines (*77–80*).

Material and methods

Genomic annotations

All genomic analyses were performed using the TAIR10 Col-0 reference genome sequence assembly. The reference TE and gene annotations were used after removal from the latter of the 3,903 genes that belong to TEs ('transposable_element_gene'; https://www.arabidopsis.org/download/).

Whole-genome bisulfite sequencing

To analyze DNA methylation specifically over TEs and separately in the three possible sequence contexts CG, CHG and CHH, which could not be done with the low resolution methylome data previously obtained for >120 epiRILs (*25*, *26*, *33*), we produced single-cytosine resolution methylomes using Whole-Genome Bisulfite sequencing (WGBS-seq, Illumina HiSeq X Ten). For each epiRIL, genomic DNA was extracted from rosette leaves collected at bolting from pools of six individuals grown in long days, 16h:8h light:dark, at 23°C. The 120 *ddm1*-derived epiRILs largely overlap with those previously epihaplotyped using low-resolution methods (*26*) but genomic DNA was extracted from individuals taken one generation later (F9) than before and also at generation F21 for 12 epiRILs. Single-cytosine resolution methylome data for the Col-0 WT and fourth generation *ddm1* parental lines were obtained using three biological replicates to provide robust measures of parental methylation levels at individual cytosines. Numbers of sequencing reads reporting methylation or no methylation over each cytosine were determined using Bismark (*81*) and pooled for Cs on opposite strands of CG and CWG (W=A or T) sites, which are symmetrical (in contrast to CCG and CGG sites) and indeed exhibit concordant levels of methylation on the two strands (fig. S1A).

Identification of Mendelian and reverting differentially methylated sites (DMSs) in the epiRILs

To identify differentially methylated CG and CWG sites (DMSs) we used a generalized linear model (GLM) approach to distinguish in the Bismark output of each site across all three WT and three *ddm1* replicates a significant genotype effect from variation between replicates. To this end, we built for each of the 2.35 millions of CG and CWG sites that are methylated in WT and that have sufficient coverage (≥5 reads per site) in at least 20 epiRILs, a GLM with a negative binomial family and an identity link function using the R package MASS and the following formula: MethylatedReads~0+TotalReads+TotalReads:Genotype, where MethylatedReads represents the number of sequencing reads in each sample reporting methylation at the site (pooled between both strands), TotalReads is the sequencing depth at the site (total number of reads supporting both methylation and non-methylation pooled between strands), and Genotype indicates whether the sample is a WT or a *ddm1* replicate. Using this method we identified 807,450 sites with significant p-values for the interaction term (adjusted $p<0.05$ after Benjamini-Hochberg correction for multiple testing) and that are hypomethylated by at least 20% in *ddm1*.

Consistent with previous observations (*21*), most (72.8%) of these DMSs are within annotated TEs or their immediate flanks (300bp on either side; fig. S1A). Using the same GLM, we then determined the methylation status (i.e., WT-like, *ddm1*-like or intermediate) of these hypomethylated DMSs across epiRILs by comparing the levels calculated in each epiRIL with the predictions from the GLM for each genotype (R function predict.glm): all values below the upper 95% confidence interval (1.96 standard error) for *ddm1* were considered *ddm1*-like, those above the lower 95% confidence interval for WT were considered WT-like, and those that fall between these two thresholds were considered intermediate. For each DMS, we then compared the ratio of WT-like epiRILs across all informative epiRILs to that expected for Mendelian inheritance (73%, given that the F1 individual used to generate the epiRIL population was backcrossed to WT, with 8% of residual selfing (*25*)). To take into account the number of informative epiRILs, we used a poisson-binomial drawing from the selfing (50-50) or the backcrossing (75-25) distribution: for 20 informative epiRILs the 95% confidence interval of Mendelian segregation is 45%-95% while it narrows down to 64%-82% when the full set of 120 epiRILs is informative (fig. S1C). Following this approach, we identified 202,721 DMSs (25,1% of total) with WT-like methylation in a percentage of epiRILs consistent with Mendelian inheritance and 587,197 DMSs (72,7% of total) for which this percentage is indicative of either partially or fully penetrant reversion instead. Finally, 17,532 DMSs (2.2% of total) exhibit an excess of *ddm1*-like methylation, which may reflect either *trans* demethylation or else inevitable skewing when the number of informative epiRILs is low (fig. S1C).

Identification of differentially methylated regions (DMRs)

As expected from previous analyses (*26*), parental DMSs next to each other tend to show coherent methylation states within and between the epiRILs (fig. S1D). We used this property to iteratively merge coherently-segregating DMSs into differentially methylated regions (DMRs) as long as they were of the same class (Mendelian or reverting), separated by no more than one DMS of another class, and within 207bp of each other. This threshold distance was chosen as the distance from which the correlation between the methylation status of two adjacent DMSs across epiRILs starts to significantly increase in variability (fig. S1E). Out of these DMRs we removed those based on less than 5 DMSs and thus obtained 9,422 Mendelian DMRs and 16,040 reverting (data S1).

We then calculated the correlation between the methylation states of each pair of the 7,151 Mendelian DMRs that contain at least 5 DMSs covered in all 120 epiRILs (data S2) to deduce the rate of linkage decay across chromosome arms and pericentromeres (no decay observed in pericentromeres given the scarcity of crossing-overs in these regions). Using a non-linear regression model with variable variance to estimate the expectations of linkage decay as a function of genomic distance - linear in pericentromeres where there is

almost no decay (R function nlme::gls) and exponential in chromosome arms (nlme::gnls) - we filtered out 1729 outlier DMRs for which methylation states are either too highly or too lowly correlated (outside of 95% confidence intervals) with other DMRs across the epiRILs (fig. S1F). High correlation at a distance is observed notably for DMRs overlapping *VANDAL* and *ENSPM* TE annotations, in line with the ability of TEs from these TE families to demethylate cognate TEs in *trans* (*82*) and we therefore removed all the DMRs corresponding to these TE families. Redundant DMRs which are close by (<5kb) and highly correlated (R^2>0.6) were also trimmed down to one representative DMR. For pericentromeric regions, given the high density of DMRs and the scarcity of crossing-overs, we selected between 80 and 100 evenly distributed DMRs for each pericentromere. In total, we obtained a subset of 1,092 evenly distributed and non-redundant Mendelian DMRs (data S4).

Building the epihaplotypic map of 120 F9 epiRILs

To infer the epihaplotypes from the 1,092 DMRs thus defined, we denoised the data using a continuous-time Hidden Markov Model (HMM, R package MSM), with methylation status of DMRs as observed states and the real parental haplotype as hidden states. The initial state occupancy probabilities are the same as the Mendelian expectation (27% *ddm1* vs 73% WT). Using recombination rates measured in female backcrosses (*83*) we calculated the expected recombination rate in our F9 epiRIL population (assuming complete homozygosity by this generation) (*84*). From these recombination rates we calculated the expected transition matrix (WT-derived to *ddm1*-derived and the opposite).

As pericentromeric regions rarely recombine, we used *ddm1*-like and WT-like pericentromeres to calculate the distribution of methylation states of DMRs within *ddm1*-derived and WT-derived intervals, respectively. We used these distributions as emission probability functions to optimize the given transition matrix to adjust for the lack of recombination events in the pericentromeres. Together with the initial state occupancy probabilities and the transition matrix, we used these emission probability functions as parameters in our HMM to derive epihaplotypic intervals delimited by DMRs. Short intervals (<1.5 Mb) were removed from further analysis as well as the small fraction (2%) of larger intervals with DMRs mainly in the intermediate rather than in the *ddm1*- or WT-like methylation state (see e.g., in fig. S2E), which may correspond to regions of residual heterozygosity. Finally, we noted five epiRILs (#62, #172, #323, #394 and #506) with a high density of *ddm1*-like DMRs in the wild-type derived pericentromeric region of chromosome 4. Remarkably, *VANDAL* and *ENSPM* are also hypomethylated in these intervals, despite the absence of a *trans*-demethylating copy present in a *ddm1*-derived interval. The most parsimonious explanation is that these correspond to *ddm1*-derived intervals with high level or reversion. These intervals were also removed from further analysis.

Overall, the new epihaplotypes (data S4) are more resolutive and extend further than the initial ones (Fig. 1A and fig. S2E). Indeed, despite excluding intervals that are poorly resolved or that could reflect residual heterozygosity, we can nonetheless now infer unambiguously the WT or *ddm1* origin ~93% of the genome sequence of each epiRIL on average, compared to 82% before (fig. S2E).

Analysis of DNA methylation levels at TE-DMRs in the epiRILs

To analyze the inheritance of DNA methylation variation at TEs in the epiRILs, we focused on the 7,023 parental DMRs that are ≥100bp-long, comprise at least 10 DMSs (at least 75% of which must be of the same class), overlap annotated TEs, and are located within the newly defined WT and *ddm1*-derived epihaplotypic intervals. Average DNA methylation levels for each C context (CG, CHG, or CHH) within each TE-DMR was calculated within each of the parents and each of the 120 epiRILs as the total proportion of methylated reads summed over all the sites of the given context within the TE-DMR that are covered by at least 5 reads.

To define lower and upper thresholds of WT-like and *ddm1*-like mCG levels, respectively, we took the first percentile of average mCG levels over these TE-DMRs in the three WT replicates and the last percentile in the three *ddm1* replicates (fig. S2A). The methylation state of a TE-DMR in an epiRIL was considered *ddm1*-like when its mCG level was strictly lower than the former threshold (0.3944), WT-like when it was strictly higher than the latter (0.7586), and intermediate otherwise (data S5). Reversion frequency per TE-DMR was calculated as the proportion of F9 epiRILs that inherited the TE-DMR from the *ddm1* parent and present WT-like mCG levels (counted as 1 event of reversion) or intermediate mCG levels (counted as 0.5 event of reversion). The thresholds between r-low, r-mid, and r-high were set arbitrarily at reversion frequencies of 33 and 66%, respectively. Persistently hypomethylated or "stray" TE-DMRs were identified for non-*VANDAL* and non-*ENSPM* TE families with at least 4 TE-DMRs, as those that remain *ddm1*-like in epiRILs where 75% or more of the TE-DMRs from the same TE family have fully reverted.

Small RNA mapping and analysis

Total RNA was extracted with Trizol according to manufacturer's instructions (Invitrogen) from pools of seedlings of *ddm1* and *ddm1rdr2*, *ddm1rdr6*, and *ddm1dcl2dcl3* compound mutants as well as from inflorescences collected from pools of 5-10 siblings of 10 F9 epiRILs selected for having contrasted epihaplotypic combinations (Fig. 1A and data S4).

sRNAs were sequenced by Fasteris with Illumina for the mutants and by the BGI with DNBseq for the 10 F9 epiRILs.

Reads were mapped using bowtie2 (*85*) using the --very-sensitive mode (-L 15 -N 1 -p 4) from which unmapped reads and secondary alignments were removed using samtools view (-F 260) (*86*). The total number of reads ranging in length from 15 to 30nt and mapping on nuclear chromosomes was extracted using samtools view to obtain the sRNA library size of each sample. Multi-mapping reads (attributed randomly to any of its mapping locations by bowtie2 in the default mode) were distinguished from unique-mapping reads based on the presence of an XS:i field in the BAM output of bowtie2 and no-mismatch reads were selected based on the XM:i field. For each category of mapping reads (multi- or single-mapping) the coverage over TE-DMRs in reads of a given length (19 to 26nt-long) was normalized by the sRNA library size of each sample in millions of reads (reads per million, RPM). TE-DMRs were considered to lose all or almost all of their single-mapping no-mismatch 23-24nt sRNAs in *ddm1* when read counts were >0 RPM in WT and in *ddm1* either 0 or at least 99% lower (fig. S3C). For representation purposes a pseudo-count of 0.001 was added to all read-counts.

To identify all potential sources of multi-mapping sRNAs, multi-mapping reads were aligned, either allowing mismatches or not, at up to 50 locations using the --very-sensitive mode of bowtie2 (-L 15 -N 1 -p 4 -k 50). 23 or 24nt-long reads mapping with or without mismatches over each TE-DMR were then identified using BedTools intersect and their (up to 50) no-mismatch alignments were counted across TE or gene annotations using htseq-count in the -m intersection-strict mode.

QTL^{epi} mapping of reversion at the TE family level

Out of 1092 DMRs used for epihaplotyping, we selected as genetic markers for interval mapping a subset of 139 that are non-redundant, have no missing information, and are separated by >250kb when located within chromosome arms and by >1Mb when within pericentromeric regions (within 5Mb of centromeres) to take into account the lower recombination rates in these regions. These 139 markers (data S4), which cover ~84.4% of the *A. thaliana* genome and are on average separated by 749.7kb, were then used as input for classical interval mapping as implemented in the mqmscan function in R/qtl. For each of the 105 TE families with at least 10 non-systematically reverting TE-DMRs (all but r-0% and r-100%), we used as a trait the average CG methylation level in each epiRIL of the subset of those TE-DMRs that are *ddm1*-derived. The mapping was performed using a step size of 2, and significance was determined for each TE family using 1000 permutations of the data, with LOD significance thresholds corresponding to 5% genome-wide false positive rates. For representation purposes, LOD scores were then normalized by the significance threshold (fig. S4D). The marker with the highest LOD score was used as top ^{epi}QTL and the second marker with the highest LOD score that is not within a continuous interval of LOD scores above the significance threshold minus 1 with the top marker was used as second ^{epi}QTL.

The percentage of variance explained by the epiallelic segregation at each of the ^{epi}QTLs was estimated by adding iteratively their epihaplotypes as parameters in a logistic GLM. We also used the epihaplotype at the top ^{epi}QTL in a logistic GLM of average CG methylation levels over the same TE-DMRs but within WT-derived intervals to identify the 24 TE families where it is significantly associated with the epihaplotypic segregation at this locus (p-value<0.05) and thus where effects within *ddm1*-derived intervals may be the result of *trans*-demethylating effects (fig. S4F).

MRSE-based identification of TE-DMR revertants in large siblings panels

To determine the methylation states of different categories of TE-DMRs in individual siblings of given F9 epiRILs, we obtained from NEB two methylation-sensitive restriction enzymes (MSREs): MspI and AvaII, whose digestion is blocked by DNA methylation and whose restriction sites are typically found once or twice per TE-DMR (MspI | NEB, AvaII | NEB). We then selected TE-DMRs that represented a diversity of non-systematic reversion categories (r-low to r-high) and i) for which we could design primers (data S6) that specifically amplify a 100 to 200nt-long region containing a digest site of MspI and/or AvaII, ii) were present in contrasted methylation states in the two epiRILs chosen (#92 and #216) and iii) for which we could confirm their differential methylation between 4 Col-WT and 4 Col-*ddm1* plants by comparing digested and non-digested samples using qPCR (Roche LightCycler 480). We then measured methylation states using this method on DNA extracted using Macherey-Nagel 96-well plate extraction kit (Macherey-Nagel, NucleoSpin 96 Plant II) from rosette leaf tissue from 96 F9 2-weeks old siblings of the two epiRILs. All qPCR measures were done in duplicates and calculations were performed on the mean of the two duplicates unless inconsistent (standard deviation >1 Ct), in which case they were discarded. Expected numbers of reversion events were calculated on the basis of the frequency of reversion among *ddm1*-derived F9 epiRILs and assuming a reversion rate per generation that is constant over generations and equal across lines (frequency of non-reversion in F9 = (1-reversion rate)^8).

Phylogeny and divergence of TE sequences

Divergence between related TE-DMRs was calculated by building a phylogeny of all the TE-DMRs within each TE family using phyml (v3.3.3) (*87*) with parameters -s SPR -d nt -a e -q --no_memory_check after trimming the MAFFT (v7.487)

(*88*) alignment of fasta sequences using trimal (*89*) in -phylip3.2. For each TE-DMR the distance to the closest TE-DMR was retrieved from the phylogeny of the corresponding TE family using the DendroPy-4.4.0 python library (*90*).

Identification of spontaneous hypomethylation events in the epiRILs

In a first step, we identified in the set of 120 F9 epiRILs all the cases where a TE-DMR, that do not belong to any of the *VANDAL* or *ENSPM* TE families, present one or two cases of *ddm1*-like hypomethylation within WT-derived epihaplotypic intervals. In order to avoid potentially confounding spontaneous hypomethylation with epihaplotyping errors at the edge of epihaplotypic interval, we then filtered out the 14 such TE-DMRs where hypomethylation is observed at a TE-DMR that is separated from the edge of their WT-derived interval by less than two TE-DMRs that are in a WT-like methylation state.

Epihaplotyping of the 169 epiRILs that minimally overlaps with ours (37 epiRILs are in common) for which single-cytosine resolution methylome data were obtained independently by others (*44*) was performed following the method described by the authors: i.e., based on the WT-like (>0.5) or *ddm1*-like (<0.5) mCG levels of 140 of the 144 DMRs they identified as stably segregating (4 were removed as they deviate by more than 7% from the 75% of WT-like expected). Using this epihaplotypic map, we identified 393 TE-DMRs with one or two events of *ddm1*-like mCG levels within WT-derived intervals that do not belong to the *VANDAL* and *ENSPM* superfamilies. However, we found events of hypomethylation within WT-derived intervals to occur almost systematically on the same epiRILs (Pearson correlation > 0.8) for two blocks of 55 and 196 TE-DMRs on Chr3 and Chr5, respectively (fig. S5E), suggestive of an epihaplotyping error over the corresponding intervals that are most likely *ddm1*-derived. 129 TE-DMRs remained after filtering out the 264 TE-DMRs with Pearson correlation>0.8.

Epivariation over TE-DMRs in natural strains

Processed bisulfite sequencing data (BS-seq) and paired-end short reads whole-genome sequencing (WGS) data from leaves of 720 genomes of *A. thaliana* were obtained from the 1,001 Genomes Project (1001genomes.org and signal.salk.edu/1001.php). For each strain, we calculated the average CG methylation weighted by the sequencing depth at each site over TE-DMRs within TEs that are present in the strain based on previous work (*46*). In addition, WGS coverage was calculated using the tool bam-readcount (*91*) in order to detect potential structural variation unaccounted for by sole TE presence/absence polymorphisms. Specifically, for each TE-DMR in each genome, we filtered out regions that have not enough WG-seq coverage (>5% of TE-DMR with 0 coverage) or an excess of reads (median TE-DMR coverage > 2x genome wide median coverage). In addition, we filtered out TE-DMRs with a number of CG sites overly divergent from the reference genome (>50%) or with less than 5 BS-seq reads per CG on average.

Analysis of DNA methylation over TE-DMRs in 20 de novo assembled genomes

Long-read Oxford Nanopore Technology (ONT) sequencing data were produced for 20 strains (chosen to represent each genetic group; data S11) out of the 720 for which BS-seq data are available. High-molecular-weight genomic DNA was extracted from grinded leaves (*92*) and sequenced at ~30X (N50=37kb; data S11) following recommended guidelines on ONT MinION flow cell (R9.4.1).

Basecalling was performed using Guppy v6.4.6 with the dna_r9.4.1_450bps_sup_plant.cfg model. Long-read fastq files were used for de novo assembly using two different assembly tools: Flye (*93*) and NextDenovo (*94*), from which the assembly with the best N50 was selected. To fix base errors in the assemblies, we polished the genomes using medaka with long-read data (https://github.com/nanoporetech/medaka), and two rounds of NextPolish (*95*) with short-read data. Finally, the contigs of assemblies are scaffolded using Ragtag (*96*) based on alignment on the reference genome (TAIR10). The location of reference genes and TEs were identified within each genome assembly using Liftoff (*97*), then TE-DMRs were located within their TE by blast. TE-DMRs were considered to be located at the same region in the de novo assembly when surrounded by the same reference genes in the genome assembly as in TAIR10. For this analysis we required that at least one the three nearest genes (restricted to TAIR10 IDs ending in 0 to avoid mis-annotated TE-genes that are themselves likely to be translocated between strains) on each side of the TE-DMRs in the de novo assemblies were also found surrounding the TE-DMRs in the reference genome. When located in synteny, TE-DMRs were considered full-length when their length in the de novo assembly was greater than 80% of that in TAIR10. BSseq data was then remapped on each de novo assembly using Bismark (*81*) to calculate mCG levels over the TE-DMRs. For visual inspection (see supplementary text, note 5), DNA methylation was also called directly from ONT-sequencing data using DeepSignal-plant (*98*) and reciprocal alignments of each assembly to the TAIR10 reference genome were produced using Minimap2 (*99*) with the -ax asm5 option.

Kinship, local haplotyping, and GWAS

Genome wide IBS and aBN kinship matrix were calculated using the emmax-kin-intel64 function of EMMAX (https://genome.sph.umich.edu/wiki/EMMAX) with

parameters -v -s -d 10 and -v -d 10, respectively, on the PLINK-transposed (v1.90p) (*100*) vcf of the short variants identified in the 1001 Genomes (1001genomes_snp-short-indel_with_tair10_only_ACGTN.vcf from https://1001genomes.org/data/GMI-MPI/releases/v3.1/) out of which we only kept bi-allelic SNPs with a minimum MAF of 0.05 and a minimum missing genotyping rate of 0.1 and found within the 720 strains with BS-seq data (SNP720.MAF005.g01 vcf).

Local haplotyping of each TE-DMRs was calculated using BEAGLE (*101*) on the SNP720.MAF005.g01 vcf in windows of 1 kb (500 bp on each side of the most central SNP within each TE-DMR). Local PCAs (fig. S8, C and D) were performed on the SNPs from SNP720.MAF005.g01 located within 1 kb around the TE-DMR.

GWASs were performed using emmax-intel64 with parameters -v -d 10 using the aBN kinship matrix and the SNP720.MAF005.g01 vcf. For epivariation prevalence at each TE-DMR, we only analyzed the 490 naturally epivariable TE-DMRs with at least 5 *ddm1*-like epivariants and with low genomic inflation (λ_{GC}<1.1). For heatmap representation of the results (fig. S7A) we subsampled to the SNPs with the highest p-values within 20 kb windows across the genome. For genome-wide epivariation prevalence, we used as phenotype the proportion of natural *ddm1*-like epivariant per strain among all naturally epivariable TE-DMRs present in the strain.

McrBC-qPCR analysis of TE-DMR epivariation in Cvi-0 x Col-0 RILs

We selected for an extended segregation analysis 40 of the 115 naturally hypomethylated TE-DMRs identified in Cvi-0 to represent a variety of superfamilies and reversion categories (Fig. 3E). Genomic DNA was extracted from frozen leaf tissue using Qiagen DNeasy Plant Mini Kit, and 30 to 100ng was used for McrBC (NEB) overnight digestion (15hr). Levels of unmethylated DNA over a target region were calculated as the fraction of DNA concentration measured by qPCR remaining following McrBC digestion compared to the same sample where McrBC is replaced by water (undigested). Efficiency of digestion was confirmed using the *ddm1*-independent DNA methylation at the 3′ end of *AT5G13440* as a positive control, and the unmethylated *AT5G13440* as a negative control (primers listed in data S14).

Using primers designed with primer3 (https://github.com/primer3-org/primer3) to specifically amplify each region (data S14), we first confirmed for 34 out of the 38 TE-DMRs that they are differentially methylated between Col-0 and Cvi-0 in freshly harvested tissue of two sibling plants. The remaining 4 TE-DMRs either displayed WT-like methylation in Cvi-0 (1 TE-DMR), hypomethylation in Col-0 (1 TE-DMR), or did not amplify in Cvi-0 (2 TE-DMRs), and we therefore excluded them from further analysis. Using McrBC-qPCR, we then measured the methylation levels of the 34 differentially methylated TE-DMRs across a diverse set of 36 F8 Cvi x Col RILs (data S15), that include the 20 of the minimal set identified by Simon *et al.* (*49*) as well as 16 additional with maximal genetic differences to the initial set of 20. TE-DMRs were considered hypomethylated when McrBC-qPCR methylation levels were below 40%, WT-like methylated when above 60%, and intermediate otherwise. TE-DMRs and ELF8 alleles were considered as Cvi-derived when both bordering markers are of the Cvi-genotype (B), Col-derived when both are A, and NA otherwise (C or D genotypes or non-agreeing bordering markers). Reversion events were identified as Cvi-derived TE-DMRs with WT-like methylation levels.

Genetic associations between TE-DMR methylation levels and the genetic maps of these RILs (*49*) were computed using the mqmscan function of the qtl R package with a window.size of 10 and a step.size of 2. The significance threshold at alpha = 0.05 was calculated using mqmpermutation with the same parameters as well as n.perm = 10 and batchsize = 25. LOD scores were normalized by the significance threshold for heatmap visualization (Fig. 3E).

For the 10 RILs for which BS-seq data was publicly available (*102*), mCG levels were calculated over all naturally epivariable TE-DMRs with *ddm1*-like methylation in Cvi-0 (N=104; figs. S7E and S9K) or with WT-like methylation in Cvi-0 (N=465; fig. S9J). Average mCG levels were calculated over both replicates of each RIL and compared to each replicates of the Col-0 and Cvi-0 parental lineages sequenced as part of the same study (fig. S7E and fig. S9, J and K). Hypomethylation and WT-like methylation were classified using the thresholds defined in the epiRILs (<0.3944 for *ddm1*-like and >0.7586 for WT-like). Genotypes were again determined using concordant bordering markers, with reversion events identified as Cvi-derived TE-DMRs with WT-like methylation levels, while spontaneous hypomethylation events as Col-derived TE-DMRs with *ddm1*-like methylation levels. Correlations between methylation levels and genotypes were calculated as a Pearson correlation coefficient.

Generalized linear modeling of genetic determinants

Association between the presence of *ddm1*-like epivariants over the 1068 naturally epivariable TE-DMRs and a set of selected genetic parameters was calculated using a binomial Generalized linear model (logit link) using R glm function. In order to limit the risk of overfitting, we restricted our analysis to the following set of the genetic parameters which we deemed most biologically relevant: copy number of each TE family in each genome, genotype at *ELF8* (using the top SNP at 2:2431793 identified in our GWAS), genotype at *NRPE1* (using the top SNP at 2:16719082 identified in previous studies (*46*, *54*)), presence of an EPICAT as defined previously (*41*)

within, under 500bp, or >500bp from the TE-DMR, distance to the nearest gene, density of CG sites, and length of the TE-DMR. TE family copy-numbers were estimated using average read-depth after remapping short-read sequencing to a library of consensus TEs as described previously (*103*). For all 10 parameters, the variables were tested as fixed effects in the glm, and for *NRPE1* and *ELF8* genotypes we also included interaction effects with copy-number and gene distance respectively. Discrete parameters such as genotypes or associations with EPICATs were considered as categorical variables, while quantitative parameters were Z-scored (R scale function) to avoid range effects. For each fixed or interaction effect, an individual GLM was built to assess their significance in explaining the prevalence of *ddm1*-like epivariation over the 1068 naturally epivariable TE-DMRs after including the three first components of the PCA of the aBN kinship matrix calculated from the SNP720.MAF005.g01 vcf to take into account population structure (together the three PCs explain ~55% of the variance). If significant ($p<0.05$), the variable with the largest R^2 (McFadden adjusted) was included in the GLM, to then iteratively test the effect of the remaining variables. For significant interaction effects, the variable was included along with its fixed effect.

Environmental associations with AraClim biovariables

196 quantitative AraClim biovariables were downloaded from https://gramene.org/CLIMtools/arabidopsis_v2.0/AraCLIM-V2/ and Z-scored (R scale function). To detect potential associations between these biovariables and the occurrence of epivariation at a TE-DMR of interest (among the 490 TE-DMRs with sufficient epivariant counts, see GWAS methods), we tested each biovariable individually as an additional explanatory variable in a binomial GLM (logit link) which included the first three PCs of the kinship matrix and the two main *trans* modifiers of the number of *ddm1*-like epivariation per strain, namely TE copy-number and *ELF8*. P-values were then corrected using the Benjamini and Hochberg correction for multiple testing using the R p.adjust function. When several AraClim variables were found to be significant (adjusted $p<0.05$), the one explaining the largest fraction of variance (McFadden adjusted R^2) was selected.

RNA-seq analyses in natural strains and in the epiRILs

For analysis of gene expression in natural strains, RNA-seq data of the 728 strains analyzed previously (*29*) was downloaded from GEO (Accession GSE80744). For analysis of gene expression in the epiRILs, total RNA was extracted with Trizol according to manufacturer's instructions (Invitrogen) from rosette leaves collected from pools of 5-10 siblings of the same set of 10 F9 epiRILs selected for sRNA analysis and sequenced by the BGI. Reads were mapped using STAR (*104*) with options --outFilterMultimapNmax 1 --alignIntronMax 10000. Coverage over exons was calculated using bedtools coverage and read counts were then normalized over entire transcripts using R package DEseq2. A pseudo count of 0.001 was added to all normalized expression levels to compute log ratios.

For each naturally epivariable TE-DMRs, expression levels of the two closest genes were compared between strains carrying the TE-DMR with *ddm1*-like or with WT-like methylation by computing the log10 ratio between the median expression in the first set over that in the second set of strains (hypo/hyper log ratios) as well as the p-value of a Wilcoxon rank-sum test between the two sets. To obtain random expectations of hypo/hyper log ratios (gray lines in fig. S10A) we shuffled randomly the strain labels among all strains with RNA-seq data and repeated the reshuffling for a total of 10 times. For each of the two closest genes, the percentage of variance in expression level that is explained by natural epivariation at the TE-DMR was calculated as the fraction of sum of squares between hypo and hyper strains over the total sum of squares.

For TE-DMRs where epivariation in nature is associated with ≥2-fold expression changes at one of the two nearest genes, we searched in the set of 10 epiRILs sequenced for those that were stably inherited hypomethylated from the *ddm1* parent in at least one of the epiRILs. To limit the possibility of indirect epihaplotypic effects that could be due to the co-occurrence of multiple hypomethylated TE-DMRs in the epiRILs, we restricted our analysis to naturally epivariable TE-DMRs that are within 5kb of the nearest gene and which are not separated from it by another naturally epivariable TE-DMR. For the 99 TE-DMRs matching these criteria we then calculated hypo/hyper log10 ratios of gene expression averages between the epiRILs where they are *ddm1*-derived and hypomethylated and the epiRILs where they are WT-derived and WT-like methylated. Gene expression changes were considered to be congruent between epiRILs and nature (fig. S10C) if both log ratios were of the same sign and the fold-change was also ≥2-fold in the epiRILs or the p-value in the epiRILs was <=0.05. For the TE-DMRs that have reverted to WT-like methylation in at least one but not all of the epiRILs where it is *ddm1*-derived, we also calculated the log10 ratio of gene expression averages between the epiRILs where they are *ddm1*-derived and hypomethylated and the epiRILs where they are *ddm1*-derived and WT-like methylated (rev/hypo log ratios; fig. S10D).

Phenotypic associations with AraPheno

417 quantitative phenotypic variables were downloaded from the AraPheno database (https://arapheno.1001genomes.org/) and Z-scored (R scale function). For each of them we retrieved from the SNP720.MAF005.g01 vcf the genotypes in

the 720 strains with BS-seq data at up to 5 of the top significant SNPs identified in araGWAS (https://aragwas.1001genomes.org/#/). To account first for population structure and the effect of these top SNPs, we built for each Z-scored AraPheno variable a Gaussian GLM (identity link) with R glm function using as explanatory variables the first three principal components of the kinship matrix as well as the genotypes at the top significant SNPs (if still significant in the subset of 720 strains). To detect residual associations between the occurrence of epivariation at a TE-DMR of interest (among the 490 TE-DMRs with sufficient epivariant counts, see GWAS methods) and variation at one of these AraPheno, we tested the epivariation as an additional categorical predictor in each GLM. P values were then corrected using the Benjamini and Hochberg correction for false discovery rate using the R p.adjust function and significant associations (adjusted p <0.05) were ranked by the fraction of phenotypic variance explained (McFadden adjusted R^2).

ELF8 *sequence analysis*

RNA extractions were performed on flash-frozen ground tissue (shoots) from Col-0 WT using Macherey-Nagel Nucleospin RNA plant kit following manufacturer's instructions. Reverse-transcriptase was performed using SuperScript IV (ThermoFisher Scientific) with oligo $d(T)_{20}$ primers and cDNA was amplified by PCR using primers listed in data S18. The PCR product was gel extracted using Macherey-Nagel Nucleospin Gel and PCR Cleanup kit and sequenced from both primers by Sanger sequencing by Eurofins Genomics. The resulting cDNA sequences and the *ELF8'* of Cvi-0 obtained by blast were analyzed and aligned to the *ELF8* TAIR10 sequence using Geneious 10.2.6 (www.geneious.com).

Analysis of HVA22E *cold-stress induction in epiRILs with contrasted epiallelic states at TE-DMR*

In order to establish that the TE-DMR is stably inherited in the *ddm1*-like or WT-like methylation states independently of the cold-stress, F10 offsprings were first obtained from F9 siblings of epiRILs #150 and #229 (*ddm1*-like at *AT5TE74320*) as well as of epiRILs #558 and #193 (WT-like at *AT5TE74320*) by bulk propagation in standard growth conditions. F10 progeny of each epiRIL was grown on MS-media in Petri dishes in standard growth conditions (23°C) for 2 weeks. Cold-stress (CS) was performed by transferring plates on ice at 4°C for 24 hours. Shoots of seedlings exposed (CS) or not (CTL) to the cold-stress were simultaneously collected at the end of the 24h cold-stress and flash-frozen individually. Methylation levels of *AT5TE74320* were measured in individual F10 seedlings using McrBC-qPCR on flash-frozen shoot tissue as described above (primers listed in data S14).

For *HVA22E* expression analysis, F10 offsprings of the same epiRILs as above (#150 and #229, #558 and #193) as well as that of #114 (*ddm1*-like at *AT5TE74320*), and of #36, and #232 (*ddm1*-derived but reverted at *AT5TE74320*) were obtained from F9 siblings propagated in a manner as described above. F10 progeny of each epiRIL was grown in liquid MS-media in 6-well plates (~15 seedlings per well) in standard growth conditions (23°C) for 2 weeks. Cold-stress (CS) was performed by transferring plates on ice at 4°C for 24 hours. Entire seedlings were collected in pools (one for each well, i.e., ~15 seedlings per pool), liquid media was removed by rinsing in 4°C sterile water, extra water was removed by padding on paper towels, and pools were flash-frozen in 1.5ml Eppendorf tubes.

RNA extractions were performed on flash-frozen ground tissue from three pools of F10 seedlings using Macherey-Nagel Nucleospin RNA plant kit following manufacturer's instructions. Reverse-transcriptase was performed using SuperScript IV (ThermoFisher Scientific) with oligo $d(T)_{20}$ primers and cDNA was cleaned up for qPCR using Macherey-Nagel Nucleospin PCR Cleanup. Expression levels of *HVA22E* were normalized by the average of two reference genes *ACTIN2* (*AT3G18780*) and *PP2AA3* (*AT1G13320*). Primers used for qPCR are listed in data S18.

Phenotyping of epiRILs growth rates in mild-drought or well-watered conditions

Growth and phenotyping of epiRILs on the Phenoscope (https://phenoscope.versailles.inrae.fr/) were performed as described previously (*61*). Briefly, using seeds obtained from the Versailles *Arabidopsis* Stock Center (https://publiclines.versailles.inrae.fr/), six plants per epiRIL were grown in individual pots, three of which were maintained at 60% (control) of the maximum soil water content (SWC) and the other three at 30% (mild-drought) of this maximum. First, seeds were germinated at soil saturation (= 100% of the maximum SWC) and seedlings were set up on the robot (= day 0 on the Phenoscope) at 8 days after sowing (DAS). Control and mild-drought SWC were stably reached and maintained from 12 and 16 DAS, respectively. Instant Relative Expansion Rates (RER instant) were calculated as the relative growth rate of the projected rosette area, integrated over +/−3 days windows. The projected rosette area was extracted by segmentation from daily zenithal images.

Statistical analyses

Unless specified otherwise, p-values are obtained by two-sided Wilcoxon rank sum test and statistical analyses and graphics were performed and obtained in MATLAB.

REFERENCES AND NOTES

1. J. N. Wells, C. Feschotte, A Field Guide to Eukaryotic Transposable Elements. *Annu. Rev. Genet.* **54**, 539–561 (2020). doi:10.1146/annurev-genet-040620-022145 Medline
2. C. D. Hirsch, N. M. Springer, Transposable element influences on gene expression

in plants. *Biochim. Biophys. Acta. Gene Regul. Mech.* **1860**, 157–165 (2017). doi:10.1016/j.bbagrm.2016.05.010 Medline
3. J. Y. Choi, Y. C. G. Lee, Double-edged sword: The evolutionary consequences of the epigenetic silencing of transposable elements. *PLOS Genet.* **16**, e1008872 (2020). doi:10.1371/journal.pgen.1008872 Medline
4. M. Gehring, Epigenetic dynamics during flowering plant reproduction: Evidence for reprogramming? *New Phytol.* **224**, 91–96 (2019). doi:10.1111/nph.15856 Medline
5. S. He, X. Feng, DNA methylation dynamics during germline development. *J. Integr. Plant Biol.* **64**, 2240–2251 (2022). doi:10.1111/jipb.13422 Medline
6. H. T. Chow, R. A. Mosher, Small RNA-mediated DNA methylation during plant reproduction. *Plant Cell* **35**, 1787–1800 (2023). doi:10.1093/plcell/koad010 Medline
7. S. Seisenberger, J. R. Peat, T. A. Hore, F. Santos, W. Dean, W. Reik, Reprogramming DNA methylation in the mammalian life cycle: Building and breaking epigenetic barriers. *Phil. Trans. R. Soc. B* **368**, 20110330 (2013). doi:10.1098/rstb.2011.0330 Medline
8. M. V. C. Greenberg, D. Bourc'his, The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019). doi:10.1038/s41580-019-0159-6 Medline
9. E. J. Richards, Inherited epigenetic variation—Revisiting soft inheritance. *Nat. Rev. Genet.* **7**, 395–401 (2006). doi:10.1038/nrg1834 Medline
10. E. Heard, R. A. Martienssen, Transgenerational epigenetic inheritance: Myths and mechanisms. *Cell* **157**, 95–109 (2014). doi:10.1016/j.cell.2014.02.045 Medline
11. L. Quadrana, V. Colot, Plant Transgenerational Epigenetics. *Annu. Rev. Genet.* **50**, 467–491 (2016). doi:10.1146/annurev-genet-120215-035254 Medline
12. P. Baduel, V. Colot, The epiallelic potential of transposable elements and its evolutionary significance in plants. *Phil. Trans. R. Soc. B* **376**, 20200123 (2021). doi:10.1098/rstb.2020.0123 Medline
13. M. Ong-Abdullah, J. M. Ordway, N. Jiang, S.-E. Ooi, S.-Y. Kok, N. Sarpan, N. Azimi, A. T. Hashim, Z. Ishak, S. K. Rosli, F. A. Malike, N. A. A. Bakar, M. Marjuni, N. Abdullah, Z. Yaakub, M. D. Amiruddin, R. Nookiah, R. Singh, E.-T. L. Low, K.-L. Chan, N. Azizi, S. W. Smith, B. Bacher, M. A. Budiman, A. Van Brunt, C. Wischmeyer, M. Beil, M. Hogan, N. Lakey, C.-C. Lim, X. Arulandoo, C.-K. Wong, C.-N. Choo, W.-C. Wong, Y.-Y. Kwan, S. S. R. S. Alwee, R. Sambanthamurthi, R. A. Martienssen, Loss of *Karma* transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature* **525**, 533–537 (2015). doi:10.1038/nature15365 Medline
14. W. J. Soppe, S. E. Jacobsen, C. Alonso-Blanco, J. P. Jackson, T. Kakutani, M. Koornneef, A. J. Peeters, The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* **6**, 791–802 (2000). doi:10.1016/S1097-2765(05)00090-0 Medline
15. Y. Kinoshita, H. Saze, T. Kinoshita, A. Miura, W. J. J. Soppe, M. Koornneef, T. Kakutani, Control of *FWA* gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *Plant J.* **49**, 38–45 (2007). doi:10.1111/j.1365-313X.2006.02936.x Medline
16. Z. Lippman, A.-V. Gendrel, M. Black, M. W. Vaughn, N. Dedhia, W. R. McCombie, K. Lavine, V. Mittal, B. May, K. D. Kasschau, J. C. Carrington, R. W. Doerge, V. Colot, R. Martienssen, Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**, 471–476 (2004). doi:10.1038/nature02651 Medline
17. D. Weigel, V. Colot, Epialleles in plant evolution. *Genome Biol.* **13**, 249 (2012). doi:10.1186/gb-2012-13-10-249 Medline
18. F. Johannes, V. Colot, R. C. Jansen, Epigenome dynamics: A quantitative genetics perspective. *Nat. Rev. Genet.* **9**, 883–890 (2008). doi:10.1038/nrg2467 Medline
19. G. Xu, J. A. Law, Loops, crosstalk, and compartmentalization: It takes many layers to regulate DNA methylation. *Curr. Opin. Genet. Dev.* **84**, 102147 (2024). doi:10.1016/j.gde.2023.102147 Medline
20. H. Zhang, Z. Lang, J.-K. Zhu, Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* **19**, 489–506 (2018). doi:10.1038/s41580-018-0016-z Medline
21. A. Zemach, M. Y. Kim, P.-H. Hsieh, D. Coleman-Derr, L. Eshed-Williams, K. Thao, S. L. Harmer, D. Zilberman, The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**, 193–205 (2013). doi:10.1016/j.cell.2013.02.033 Medline
22. J. A. Jeddeloh, T. L. Stokes, E. J. Richards, Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat. Genet.* **22**, 94–97 (1999). doi:10.1038/8803 Medline
23. A. Vongs, T. Kakutani, R. A. Martienssen, E. J. Richards, *Arabidopsis thaliana* DNA methylation mutants. *Science* **260**, 1926–1928 (1993). doi:10.1126/science.8316832 Medline
24. S. C. Lee, D. W. Adams, J. J. Ipsaro, J. Cahn, J. Lynn, H.-S. Kim, B. Berube, V. Major, J. P. Calarco, C. LeBlanc, S. Bhattacharjee, U. Ramu, D. Grimanelli, Y. Jacob, P. Voigt, L. Joshua-Tor, R. A. Martienssen, Chromatin remodeling of histone H3 variants by DDM1 underlies epigenetic inheritance of DNA methylation. *Cell* **186**, 4100–4116.e15 (2023). doi:10.1016/j.cell.2023.08.001 Medline
25. F. Johannes, E. Porcher, F. K. Teixeira, V. Saliba-Colombani, M. Simon, N. Agier, A. Bulski, J. Albuisson, F. Heredia, P. Audigier, D. Bouchez, C. Dillmann, P. Guerche, F. Hospital, V. Colot, Assessing the impact of transgenerational epigenetic variation on complex traits. *PLOS Genet.* **5**, e1000530 (2009). doi:10.1371/journal.pgen.1000530 Medline
26. M. Colomé-Tatché, S. Cortijo, R. Wardenaar, L. Morgado, B. Lahouze, A. Sarazin, M. Etcheverry, A. Martin, S. Feng, E. Duvernois-Berthet, K. Labadie, P. Wincker, S. E. Jacobsen, R. C. Jansen, V. Colot, F. Johannes, Features of the *Arabidopsis* recombination landscape resulting from the combined loss of sequence variation and DNA methylation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16240–16245 (2012). doi:10.1073/pnas.1212955109 Medline
27. S. Cortijo, R. Wardenaar, M. Colomé-Tatché, A. Gilly, M. Etcheverry, K. Labadie, E. Caillieux, F. Hospital, J.-M. Aury, P. Wincker, F. Roudier, R. C. Jansen, V. Colot, F. Johannes, Mapping the epigenetic basis of complex traits. *Science* **343**, 1145–1148 (2014). doi:10.1126/science.1248127 Medline
28. 1001 Genomes Consortium, 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*. *Cell* **166**, 481–491 (2016). doi:10.1016/j.cell.2016.05.063 Medline
29. T. Kawakatsu, S.-S. C. Huang, F. Jupe, E. Sasaki, R. J. Schmitz, M. A. Urich, R. Castanon, J. R. Nery, C. Barragan, Y. He, H. Chen, M. Dubin, C.-R. Lee, C. Wang, F. Bemm, C. Becker, R. O'Neil, R. C. O'Malley, D. X. Quarless, 1001 Genomes Consortium, N. J. Schork, D. Weigel, M. Nordborg, J. R. Ecker, Epigenomic Diversity in a Global Collection of *Arabidopsis thaliana* Accessions. *Cell* **166**, 492–505 (2016). doi:10.1016/j.cell.2016.06.044 Medline
30. M. Togninalli, Ü. Seren, D. Meng, J. Fitz, M. Nordborg, D. Weigel, K. Borgwardt, A. Korte, D. G. Grimm, The AraGWAS Catalog: A curated and standardized *Arabidopsis thaliana* GWAS catalog. *Nucleic Acids Res.* **46**, D1150–D1156 (2018). doi:10.1093/nar/gkx954 Medline
31. Á. Ferrero-Serrano, S. M. Assmann, Phenotypic and genome-wide association with the local environment of *Arabidopsis*. *Nat. Ecol. Evol.* **3**, 274–285 (2019). doi:10.1038/s41559-018-0754-5 Medline
32. L. Quadrana, M. Etcheverry, A. Gilly, E. Caillieux, M.-A. Madoui, J. Guy, A. Bortolini Silveira, S. Engelen, V. Baillet, P. Wincker, J.-M. Aury, V. Colot, Transposition favors the generation of large effect mutations that may facilitate rapid adaption. *Nat. Commun.* **10**, 3421 (2019). doi:10.1038/s41467-019-11385-5 Medline
33. F. K. Teixeira, F. Heredia, A. Sarazin, F. Roudier, M. Boccara, C. Ciaudo, C. Cruaud, J. Poulain, M. Berdasco, M. F. Fraga, O. Voinnet, P. Wincker, M. Esteller, V. Colot, A role for RNAi in the selective correction of DNA methylation defects. *Science* **323**, 1600–1604 (2009). doi:10.1126/science.1165313 Medline
34. L. M. Martins, J. A. Law, Moving targets: Mechanisms regulating siRNA production and DNA methylation during plant development. *Curr. Opin. Plant Biol.* **75**, 102435 (2023). doi:10.1016/j.pbi.2023.102435 Medline
35. F. K. Teixeira, V. Colot, Repeat elements and the Arabidopsis DNA methylation landscape. *Heredity* **105**, 14–23 (2010). doi:10.1038/hdy.2010.52 Medline
36. D. M. Bond, D. C. Baulcombe, Small RNAs and heritable epigenetic variation in plants. *Trends Cell Biol.* **24**, 100–107 (2014). doi:10.1016/j.tcb.2013.08.001 Medline
37. M. A. Matzke, R. A. Mosher, RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15**, 394–408 (2014). doi:10.1038/nrg3683 Medline
38. J. M. Wendte, C. S. Pikaard, The RNAs of RNA-directed DNA methylation. *Biochim. Biophys. Acta. Gene Regul. Mech.* **1860**, 140–148 (2017). doi:10.1016/j.bbagrm.2016.08.004 Medline
39. K. Osabe, Y. Harukawa, S. Miura, H. Saze, Epigenetic Regulation of Intronic Transgenes in *Arabidopsis*. *Sci. Rep.* **7**, 45166 (2017). doi:10.1038/srep45166 Medline

40. A. Dalakouras, E. Dadami, M. Wassenegger, G. Krczal, M. Wassenegger, RNA-directed DNA methylation efficiency depends on trigger and target sequence identity. *Plant J.* **87**, 202–214 (2016). doi:10.1111/tpj.13193 Medline
41. N. T. Le, Y. Harukawa, S. Miura, D. Boer, A. Kawabe, H. Saze, Epigenetic regulation of spurious transcription initiation in *Arabidopsis*. *Nat. Commun.* **11**, 3224 (2020). doi:10.1038/s41467-020-16951-w Medline
42. J. Li, D.-L. Yang, H. Huang, G. Zhang, L. He, J. Pang, R. Lozano-Durán, Z. Lang, J.-K. Zhu, Epigenetic memory marks determine epiallele stability at loci targeted by de novo DNA methylation. *Nat. Plants* **6**, 661–674 (2020). doi:10.1038/s41477-020-0671-x Medline
43. H. Bente, A. M. Foerster, N. Lettner, O. Mittelsten Scheid, Polyploidy-associated paramutation in Arabidopsis is determined by small RNAs, temperature, and allele structure. *PLOS Genet.* **17**, e1009444 (2021). doi:10.1371/journal.pgen.1009444 Medline
44. Y. Zhang, H. Jang, R. Xiao, I. Kakoulidou, R. S. Piecyk, F. Johannes, R. J. Schmitz, Heterochromatin is a quantitative trait associated with spontaneous epiallele formation. *Nat. Commun.* **12**, 6958 (2021). doi:10.1038/s41467-021-27320-6 Medline
45. T. Stuart, S. R. Eichten, J. Cahn, Y. V. Karpievitch, J. O. Borevitz, R. Lister, Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. *eLife* **5**, e20777 (2016). doi:10.7554/eLife.20777 Medline
46. P. Baduel, B. Leduque, A. Ignace, I. Gy, J. Gil Jr., O. Loudet, V. Colot, L. Quadrana, Genetic and environmental modulation of transposition shapes the evolutionary potential of *Arabidopsis thaliana*. *Genome Biol.* **22**, 138 (2021). doi:10.1186/s13059-021-02348-5 Medline
47. M. W. Vaughn, M. Tanurdzić, Z. Lippman, H. Jiang, R. Carrasquillo, P. D. Rabinowicz, N. Dedhia, W. R. McCombie, N. Agier, A. Bulski, V. Colot, R. W. Doerge, R. A. Martienssen, Epigenetic natural variation in Arabidopsis thaliana. *PLOS Biol.* **5**, e174 (2007). doi:10.1371/journal.pbio.0050174 Medline
48. B. Jaegle, R. Pisupati, L. M. Soto-Jiménez, R. Burns, F. A. Rabanal, M. Nordborg, Extensive sequence duplication in *Arabidopsis* revealed by pseudo-heterozygosity. *Genome Biol.* **24**, 44 (2023). doi:10.1186/s13059-023-02875-3 Medline
49. M. Simon, O. Loudet, S. Durand, A. Bérard, D. Brunel, F.-X. Sennesal, M. Durand-Tardif, G. Pelletier, C. Camilleri, Quantitative trait loci mapping in five new large recombinant inbred line populations of *Arabidopsis thaliana* genotyped with consensus single-nucleotide polymorphism markers. *Genetics* **178**, 2253–2264 (2008). doi:10.1534/genetics.107.083899 Medline
50. M. V. C. Greenberg, A. Deleris, C. J. Hale, A. Liu, S. Feng, S. E. Jacobsen, Interplay between active chromatin marks and RNA-directed DNA methylation in *Arabidopsis thaliana*. *PLOS Genet.* **9**, e1003946 (2013). doi:10.1371/journal.pgen.1003946 Medline
51. S. Oh, S. Park, S. van Nocker, Genic and global functions for Paf1C in chromatin modification and gene expression in Arabidopsis. *PLOS Genet.* **4**, e1000077 (2008). doi:10.1371/journal.pgen.1000077 Medline
52. N. C. Deans, J. R. B. Talbot, M. Li, C. Sáez-González, I. Hövel, D. Heavens, M. Stam, J. B. Hollick, Paramutation at the maize *pl1* locus is associated with RdDM activity at distal tandem repeats. *PLOS Genet.* **20**, e1011296 (2024). doi:10.1371/journal.pgen.1011296 Medline
53. V. L. Chandler, W. B. Eggleston, J. E. Dorweiler, Paramutation in maize. *Plant Mol. Biol.* **43**, 121–145 (2000). doi:10.1023/A:1006499808317 Medline
54. E. Sasaki, T. Kawakatsu, J. R. Ecker, M. Nordborg, Common alleles of *CMT2* and *NRPE1* are major determinants of CHH methylation variation in *Arabidopsis thaliana*. *PLOS Genet.* **15**, e1008492 (2019). doi:10.1371/journal.pgen.1008492 Medline
55. A. B. Silveira, C. Trontin, S. Cortijo, J. Barau, L. E. V. Del Bem, O. Loudet, V. Colot, M. Vincentz, Extensive natural epigenetic variation at a *de novo* originated gene. *PLOS Genet.* **9**, e1003437 (2013). doi:10.1371/journal.pgen.1003437 Medline
56. T. N. Le, Y. Miyazaki, S. Takuno, H. Saze, Epigenetic regulation of intragenic transposable elements impacts gene transcription in *Arabidopsis thaliana*. *Nucleic Acids Res.* **43**, 3911–3921 (2015). doi:10.1093/nar/gkv258 Medline
57. T. Tsuchiya, T. Eulgem, An alternative polyadenylation mechanism coopted to the *Arabidopsis RPP7* gene through intronic retrotransposon domestication. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3535–E3543 (2013). doi:10.1073/pnas.1312545110 Medline
58. B. Ghimire, T. Saminathan, A. Bodunrin, V. L. Abburi, A. O. Kshetry, S. Shinde, P. Nimmakayala, U. K. Reddy, Genome-Wide Association Study of Natural Variation in *Arabidopsis* Exposed to Acid Mine Drainage Toxicity and Validation of Associated Genes with Reverse Genetics. *Plants* **10**, 191 (2021). doi:10.3390/plants10020191 Medline
59. J. Prakash, S. B. Agrawal, M. Agrawal, Global Trends of Acidity in Rainfall and Its Impact on Plants and Soil. *J. Soil Sci. Plant Nutr.* **23**, 398–419 (2023). doi:10.1007/s42729-022-01051-z Medline
60. H. Bauer, P. Ache, S. Lautner, J. Fromm, W. Hartung, K. A. S. Al-Rasheid, S. Sonnewald, U. Sonnewald, S. Kneitz, N. Lachmann, R. R. Mendel, F. Bittner, A. M. Hetherington, R. Hedrich, The stomatal response to reduced relative humidity requires guard cell-autonomous ABA synthesis. *Curr. Biol.* **23**, 53–57 (2013). doi:10.1016/j.cub.2012.11.022 Medline
61. S. Tisné, Y. Serrand, L. Bach, E. Gilbault, R. Ben Ameur, H. Balasse, R. Voisin, D. Bouchez, M. Durand-Tardif, P. Guerche, G. Chareyron, J. Da Rugna, C. Camilleri, O. Loudet, Phenoscope: An automated large-scale phenotyping platform offering high spatial homogeneity. *Plant J.* **74**, 534–544 (2013). doi:10.1111/tpj.12131 Medline
62. I. Kakoulidou, R. S. Piecyk, R. C. Meyer, M. Kuhlmann, C. Gutjahr, T. Altmann, F. Johannes, Mapping parental DMRs predictive of local and distal methylome remodeling in epigenetic F1 hybrids. *Life Sci. Alliance* **7**, e202402599 (2024). doi:10.26508/lsa.202402599 Medline
63. X. Zhong, C. J. Hale, J. A. Law, L. M. Johnson, S. Feng, A. Tu, S. E. Jacobsen, DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat. Struct. Mol. Biol.* **19**, 870–875 (2012). doi:10.1038/nsmb.2354 Medline
64. H. Stroud, T. Do, J. Du, X. Zhong, S. Feng, L. Johnson, D. J. Patel, S. E. Jacobsen, Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat. Struct. Mol. Biol.* **21**, 64–72 (2014). doi:10.1038/nsmb.2735 Medline
65. A. D. McCue, S. Nuthikattu, R. K. Slotkin, Genome-wide identification of genes regulated in *trans* by transposable element small interfering RNAs. *RNA Biol.* **10**, 1379–1395 (2013). doi:10.4161/rna.25555 Medline
66. S. Li, L. E. Vandivier, B. Tu, L. Gao, S. Y. Won, S. Li, B. Zheng, B. D. Gregory, X. Chen, Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in *Arabidopsis* reveals features and regulation of siRNA biogenesis. *Genome Res.* **25**, 235–245 (2015). doi:10.1101/gr.182238.114 Medline
67. J. B. Hollick, Paramutation and related phenomena in diverse species. *Nat. Rev. Genet.* **18**, 5–23 (2017). doi:10.1038/nrg.2016.115 Medline
68. R. Martienssen, Epigenetic phenomena: Paramutation and gene silencing in plants. *Curr. Biol.* **6**, 810–813 (1996). doi:10.1016/S0960-9822(02)00601-2 Medline
69. A. M. Settles, A. Baron, A. Barkan, R. A. Martienssen, Duplication and suppression of chloroplast protein translocation genes in maize. *Genetics* **157**, 349–360 (2001). doi:10.1093/genetics/157.1.349 Medline
70. J. B. Hollick, J. E. Dorweiler, V. L. Chandler, Paramutation and related allelic interactions. *Trends Genet.* **13**, 302–308 (1997). doi:10.1016/S0168-9525(97)01184-0 Medline
71. I. K. Greaves, M. Groszmann, H. Ying, J. M. Taylor, W. J. Peacock, E. S. Dennis, Trans chromosomal methylation in *Arabidopsis* hybrids. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3570–3575 (2012). doi:10.1073/pnas.1201043109 Medline
72. H. Zhang, J.-K. Zhu, Epigenetic gene regulation in plants and its potential applications in crop improvement. *Nat. Rev. Mol. Cell Biol.* **26**, 51–67 (2025). doi:10.1038/s41580-024-00769-1 Medline
73. K. M. Kowalik, Y. Shimada, V. Flury, M. B. Stadler, J. Batki, M. Bühler, The Paf1 complex represses small-RNA-mediated epigenetic gene silencing. *Nature* **520**, 248–252 (2015). doi:10.1038/nature14337 Medline
74. L. Duempelmann, F. Mohn, Y. Shimada, D. Oberti, A. Andriollo, S. Lochs, M. Bühler, Inheritance of a Phenotypically Neutral Epimutation Evokes Gene Silencing in Later Generations. *Mol. Cell* **74**, 534–541.e4 (2019). doi:10.1016/j.molcel.2019.02.009 Medline
75. M. W. Schmid, C. Heichinger, D. Coman Schmid, D. Guthörl, V. Gagliardini, R. Bruggmann, S. Aluri, C. Aquino, B. Schmid, L. A. Turnbull, U. Grossniklaus, Contribution of epigenetic variation to adaptation in *Arabidopsis*. *Nat. Commun.* **9**, 4446 (2018). doi:10.1038/s41467-018-06932-5 Medline

76. M. W. Schmid, K. Kropivšek, S. E. Wuest, B. Schmid, U. Grossniklaus, Weak evidence for heritable changes in response to selection by aphids in *Arabidopsis* accessions. bioRxiv 2025.01.21.634027 [Preprint] (2025). doi:10.1101/2025.01.21.634027
77. C. Jiang, A. Mithani, E. J. Belfield, R. Mott, L. D. Hurst, N. P. Harberd, Environmentally responsive genome-wide accumulation of de novo *Arabidopsis thaliana* mutations and epimutations. *Genome Res.* **24**, 1821–1829 (2014). doi:10.1101/gr.177659.114 Medline
78. N. Yao, Z. Zhang, L. Yu, R. Hazarika, C. Yu, H. Jang, L. M. Smith, J. Ton, L. Liu, J. J. Stachowicz, T. B. H. Reusch, R. J. Schmitz, F. Johannes, An evolutionary epigenetic clock in plants. *Science* **381**, 1440–1445 (2023). doi:10.1126/science.adh9443 Medline
79. D. R. Ganguly, P. A. Crisp, S. R. Eichten, B. J. Pogson, The Arabidopsis DNA Methylome Is Stable under Transgenerational Drought Stress. *Plant Physiol.* **175**, 1893–1912 (2017). doi:10.1104/pp.17.00744 Medline
80. A. Wibowo, C. Becker, G. Marconi, J. Durr, J. Price, J. Hagmann, R. Papareddy, H. Putra, J. Kageyama, J. Becker, D. Weigel, J. Gutierrez-Marcos, Hyperosmotic stress memory in Arabidopsis is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity. *eLife* **5**, e13546 (2016). doi:10.7554/eLife.13546 Medline
81. F. Krueger, S. R. Andrews, Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011). doi:10.1093/bioinformatics/btr167 Medline
82. T. Sasaki, K. Ro, E. Caillieux, R. Manabe, G. Bohl-Viallefond, P. Baduel, V. Colot, T. Kakutani, L. Quadrana, Fast co-evolution of anti-silencing systems shapes the invasiveness of *Mu*-like DNA transposons in eudicots. *EMBO J.* **41**, e110070 (2022). doi:10.15252/embj.2021110070 Medline
83. L. Giraut, M. Falque, J. Drouaud, L. Pereira, O. C. Martin, C. Mézard, Genome-wide crossover distribution in *Arabidopsis thaliana* meiosis reveals sex-specific patterns along chromosomes. *PLOS Genet.* **7**, e1002354 (2011). doi:10.1371/journal.pgen.1002354 Medline
84. Z. Sun, H. Li, L. Zhang, J. Wang, Estimation of recombination frequency in bi-parental genetic populations. *Genet. Res.* **94**, 163–177 (2012). doi:10.1017/S0016672312000353 Medline
85. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi:10.1038/nmeth.1923 Medline
86. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009). doi:10.1093/bioinformatics/btp352 Medline
87. S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010). doi:10.1093/sysbio/syq010 Medline
88. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002). doi:10.1093/nar/gkf436 Medline
89. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009). doi:10.1093/bioinformatics/btp348 Medline
90. J. Sukumaran, M. T. Holder, DendroPy: A Python library for phylogenetic computing. *Bioinformatics* **26**, 1569–1571 (2010). doi:10.1093/bioinformatics/btq228 Medline
91. A. Khanna, D. E. Larson, S. N. Srivatsan, M. Mosior, T. E. Abbott, S. Kiwala, T. J. Ley, E. J. Duncavage, M. J. Walter, J. R. Walker, O. L. Griffith, M. Griffith, C. A. Miller, Bam-readcount - Rapid generation of basepair-resolution sequence metrics. *J. Open Source Softw.* **7**, 3722 (2022). doi:10.21105/joss.03722
92. B. Mayjonade, J. Gouzy, C. Donnadieu, N. Pouilly, W. Marande, C. Callot, N. Langlade, S. Muños, Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *Biotechniques* **61**, 203–205 (2016). doi:10.2144/000114460 Medline
93. M. Kolmogorov, J. Yuan, Y. Lin, P. A. Pevzner, Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019). doi:10.1038/s41587-019-0072-8 Medline
94. J. Hu, Z. Wang, Z. Sun, B. Hu, A. O. Ayoola, F. Liang, J. Li, J. R. Sandoval, D. N. Cooper, K. Ye, J. Ruan, C.-L. Xiao, D. Wang, D.-D. Wu, S. Wang, NextDenovo: An efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol.* **25**, 107 (2024). doi:10.1186/s13059-024-03252-4 Medline
95. J. Hu, J. Fan, Z. Sun, S. Liu, NextPolish: A fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253–2255 (2020). doi:10.1093/bioinformatics/btz891 Medline
96. M. Alonge, L. Lebeigle, M. Kirsche, K. Jenike, S. Ou, S. Aganezov, X. Wang, Z. B. Lippman, M. C. Schatz, S. Soyk, Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* **23**, 258 (2022). doi:10.1186/s13059-022-02823-7 Medline
97. A. Shumate, S. L. Salzberg, Liftoff: Accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021). doi:10.1093/bioinformatics/btaa1016 Medline
98. P. Ni, N. Huang, F. Nie, J. Zhang, Z. Zhang, B. Wu, L. Bai, W. Liu, C.-L. Xiao, F. Luo, J. Wang, Genome-wide detection of cytosine methylations in plant from Nanopore data using deep learning. *Nat. Commun.* **12**, 5976 (2021). doi:10.1038/s41467-021-26278-9 Medline
99. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018). doi:10.1093/bioinformatics/bty191 Medline
100. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, s13742-015-0047-8 (2015). doi:10.1186/s13742-015-0047-8 Medline
101. S. R. Browning, B. L. Browning, Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007). doi:10.1086/521987 Medline
102. C. L. Picard, M. Gehring, Proximal methylation features associated with nonrandom changes in gene body methylation. *Genome Biol.* **18**, 73 (2017). doi:10.1186/s13059-017-1206-2 Medline
103. L. Quadrana, A. Bortolini Silveira, G. F. Mayhew, C. LeBlanc, R. A. Martienssen, J. A. Jeddeloh, V. Colot, The *Arabidopsis thaliana* mobilome and its impact at the species level. *eLife* **5**, e15716 (2016). doi:10.7554/eLife.15716 Medline
104. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). doi:10.1093/bioinformatics/bts635 Medline
105. A. Hosaka, R. Saito, K. Takashima, T. Sasaki, Y. Fu, A. Kawabe, T. Ito, A. Toyoda, A. Fujiyama, Y. Tarutani, T. Kakutani, Evolution of sequence-specific anti-silencing systems in *Arabidopsis*. *Nat. Commun.* **8**, 2161 (2017). doi:10.1038/s41467-017-02150-7 Medline
106. R. Nagar, B. Schwessinger, DNA size selection (>3-4kb) and purification of DNA using an improved homemade SPRI beads solution, protocols.io (2018); https://dx.doi.org/10.17504/protocols.io.n7hdhj6.
107. L. He, W. Wu, G. Zinta, L. Yang, D. Wang, R. Liu, H. Zhang, Z. Zheng, H. Huang, Q. Zhang, J.-K. Zhu, A naturally occurring epiallele associates with leaf senescence and local climate adaptation in *Arabidopsis* accessions. *Nat. Commun.* **9**, 460 (2018). doi:10.1038/s41467-018-02839-3 Medline
108. I. Ahmed, A. Sarazin, C. Bowler, V. Colot, H. Quesneville, Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. *Nucleic Acids Res.* **39**, 6919–6931 (2011). doi:10.1093/nar/gkr324 Medline
109. J. Liu, X. Zhong, Epiallelic variation of non-coding RNA genes and their phenotypic consequences. *Nat. Commun.* **15**, 1375 (2024). doi:10.1038/s41467-024-45771-5 Medline
110. A. E. Kornienko, V. Nizhynska, A. Molla Morales, R. Pisupati, M. Nordborg, Population-level annotation of lncRNAs in Arabidopsis reveals extensive expression variation associated with transposable element–like silencing. *Plant Cell* **36**, 85–111 (2023). doi:10.1093/plcell/koad233 Medline
111. Z. Shahzad, J. D. Moore, J. Choi, D. Zilberman, Epigenetic inheritance mediates phenotypic diversity in natural populations. bioRxiv 2021.03.15.435374 [Preprint] (2021). doi:10.1101/2021.03.15.435374
112. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska,

T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). doi:10.1038/s41586-021-03819-2 Medline
113. M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: Making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022). doi:10.1038/s41592-022-01488-1 Medline
114. Y. He, M. R. Doyle, R. M. Amasino, PAF1-complex-mediated histone methylation of *FLOWERING LOCUS C* chromatin is required for the vernalization-responsive, winter-annual habit in *Arabidopsis*. *Genes Dev.* **18**, 2774–2784 (2004). doi:10.1101/gad.1244504 Medline
115. S. Oh, H. Zhang, P. Ludwig, S. van Nocker, A mechanism related to the yeast transcriptional regulator Paf1c is required for expression of the Arabidopsis *FLC/MAF* MADS box gene family. *Plant Cell* **16**, 2940–2953 (2004). doi:10.1105/tpc.104.026062 Medline
116. S. Singh, S. Kailasam, J.-C. Lo, K.-C. Yeh, Histone H3 lysine4 trimethylation-regulated GRF11 expression is essential for the iron-deficiency response in *Arabidopsis thaliana*. *New Phytol.* **230**, 244–258 (2021). doi:10.1111/nph.17130 Medline
117. J. Denkena, F. Johannes, M. Colomé-Tatché, Region-level epimutation rates in *Arabidopsis thaliana*. *Heredity* **127**, 190–202 (2021). doi:10.1038/s41437-021-00441-w Medline

ACKNOWLEDGMENTS

The work on natural epivariation was made possible thanks to the genome, methylome, and transcriptome resources provided by the *Arabidopsis* 1001 Genomes Project. We thank current and past members of the Colot group for discussions and feedback over the lifespan of the project. We thank Edith Heard for critical reading of the manuscript and suggestions for improvement. Computational analyses were performed on the IBENS bioclust computing cluster and we thank the bioinformatics platform team led by Pierre Vincens for their technical support. High-throughput phenotyping benefited from the support of the IJPB Plant Observatory facilities PO-PHENO and PO-VASC. **Funding:** European Union (EpiGeneSys FP7 Network of Excellence #257082, to VC); French Agence Nationale pour la Recherche (ANR-12-ADAP-00020-01 to VC and OL; ANR-21-CE45-0018, to VC) and by a grant from the PSL-QLife Institute (to VC); Fondation pour la Recherche Médicale (FRM SPF20170938626 postdoctoral fellowship to PB); Saclay Plant Sciences-SPS (ANR-17-EUR-0007, to OL). **Author contributions:** Study design: PB, VC; Supervision: PB, VC; Writing – original draft: PB, VC; Writing – review & editing: PB, VC; DNA, RNA, and sRNA extractions of epiRILs, ONT-sequencing of 20 strains: EC; Analysis of WGBS-seq data, GWAS, GLM analysis, de novo assembly of 20 genomes and DNA methylation calling using ONT-sequencing data: LDO; Epihaplotyping of F9 epiRILs: GBV, VS; Analysis of epivariation segregation over TE-DMRs in the epiRILs and RILs: PB; McrBC analysis of RILs: CX; Analysis of sRNA seq data: AS and PB; Expression and DNA methylation analysis at *HVA22E*: MEM, CX, and MD; Genome to assembly alignments for visual inspection: AP; Visual inspection: AP, PB, and VC; Single-read DNA methylation analysis using ONT: Mba; Generation of *ddm1* and *ddm1* compound mutants and extraction of sRNAs: FKT, AS and Mbo; Estimation of TE family copy-numbers in the genomes of the 1001 Genomes Project: ADF and LQ; Phenotyping of epiRILs using the Phenoscope: EG and OL. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** Raw BS-seq, RNA-seq, sRNA-seq, and ONT-seq read-data (fastq and fasta of 20 de novo genome assemblies, fast5 of epiRIL238) generated and analyzed for this study are available in the European Nucleotide Archive (ENA) under project PRJEB77668. WGS data was obtained in FASTQ format for the 1047 strains of the 1001genomes.org project from NCBI SRA Project PRJNA273563. Raw RNA-seq data was obtained from NCBI GEO Accession GSE80744. Processed BS-seq data was obtained from NCBI GEO Accession GSE43857. Raw DAP-seq data was obtained from NCBI GEO Accession GSE60143. Seeds of the *ddm1*-derived epiRIL population and the Cvi x Col RILs were obtained from the Versailles *Arabidopsis* Stock Center (https://publiclines.versailles.inrae.fr/).

This research was funded in whole or in part by the Agence Nationale pour la Recherche (ANR-12-ADAP-00020-01 and ANR-21-CE45-0018), a cOAlition S organization. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ady3475
Supplementary Text
Figs. S1 to S12
References (*105–117*)
MDAR Reproducibility Checklist
Data S1 to S20

Submitted 18 April 2025; accepted 11 August 2025
Published online 18 September 2025
10.1126/science.ady3475

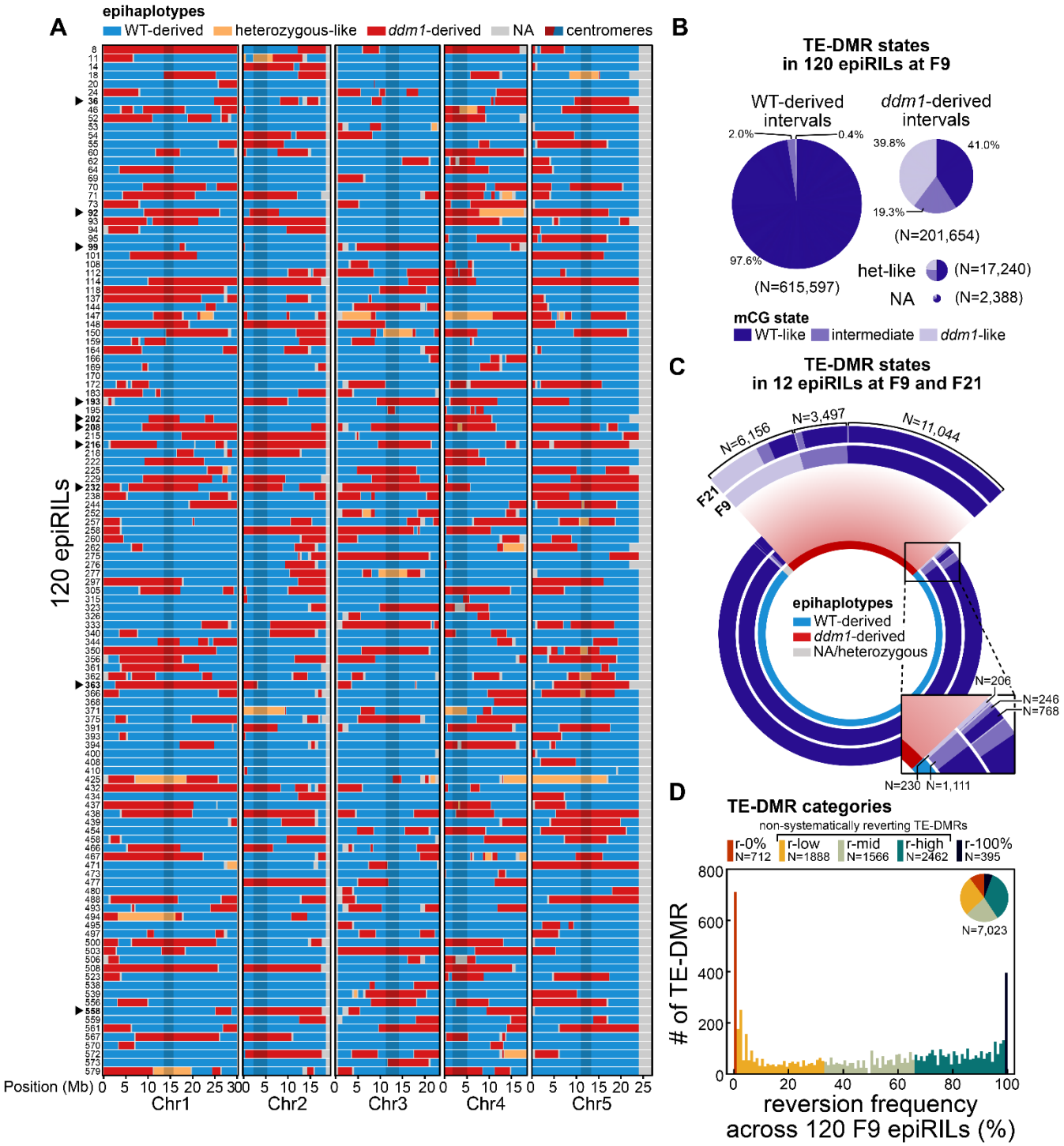

Fig. 1. Inheritance patterns of *ddm1*-induced hypomethylation.

Fig. 1. Inheritance patterns of *ddm1*-induced hypomethylation. (**A**) Epihaplotypic map of 120 F9 epiRILs. Centromeric regions are shaded. EpiRIL ID number is indicated on the left. (**B**) Distribution of methylation states (*ddm1*-like, intermediate, or WT-like mCG levels) of TE-DMRs within WT-derived or *ddm1*-derived intervals in the 120 F9 epiRILs. Black triangles point to the 10 contrasted epiRILs selected for RNA-seq. (**C**) Distribution of methylation states in the 12 epiRILs with F9 and F21 data. (**D**) Distribution of reversion frequencies measured at F9. Colors indicate the five categories of TE-DMRs we defined based on reversion frequency (r-0% to r-100%) and the pie chart indicates their relative proportion.

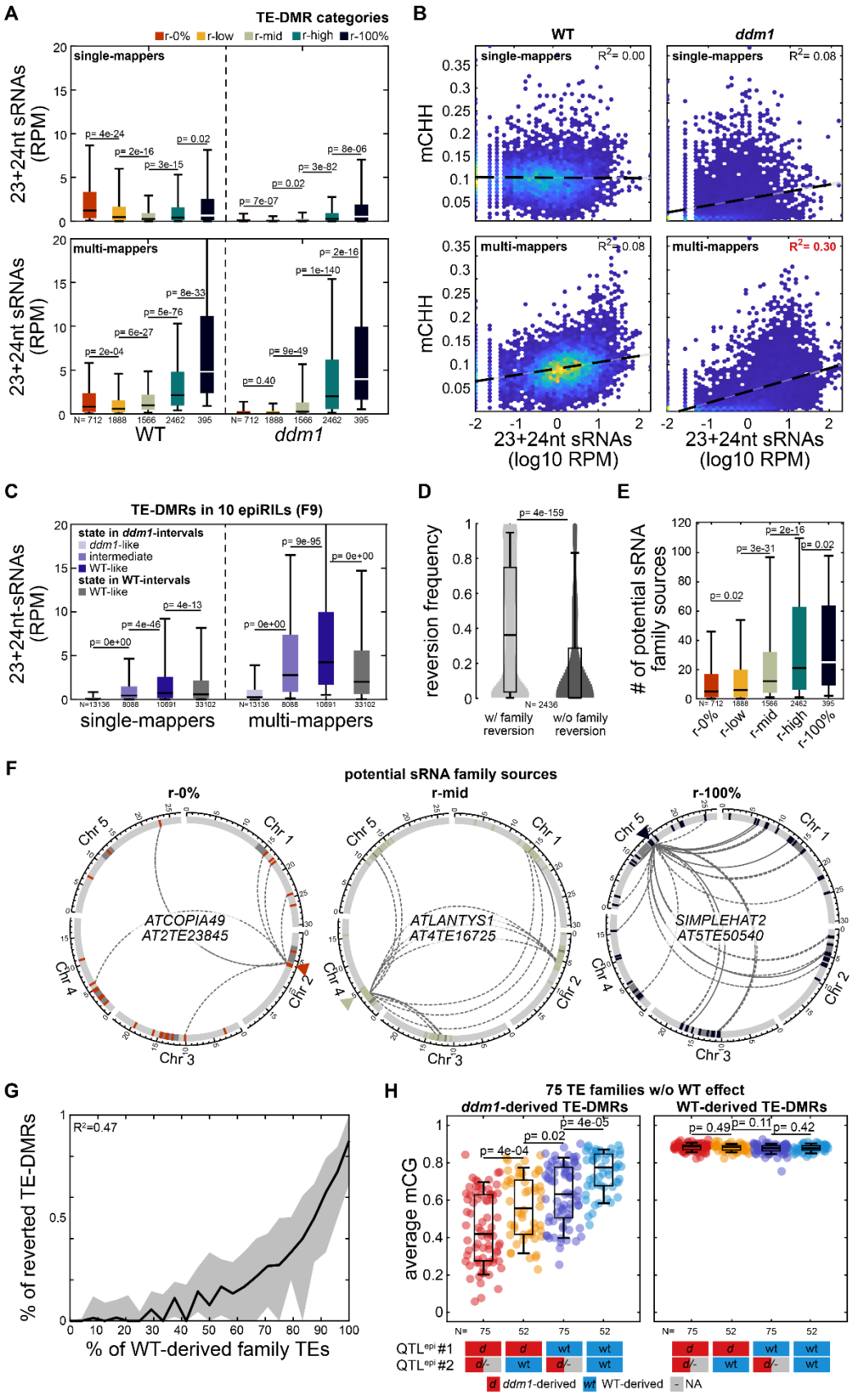

Fig. 2. Reversion is driven mainly by *trans*-acting sRNAs.

Fig. 2. Reversion is driven mainly by *trans*-acting sRNAs. (**A**) Levels of 23-24nt single- (top) and multi-mapping (bottom) sRNAs in WT and *ddm1* over the five TE-DMR reversion categories. (**B**) Levels of mCHH vs levels of 23-24nt single- (top) and multi-mapping (bottom) sRNAs in WT and *ddm1* (log10 RPM). The coefficient of correlation (R^2) of the linear regression (dotted line) is indicated. (**C**) Levels of 23-24nt single- and multi-mapping sRNAs over TE-DMRs in 10 sequenced F9 epiRILs, depending on their parental origin and DNA methylation state. Only TE-DMRs that reverted in at least one of the 10 epiRILs were considered. (**D**) Reversion frequencies of TE-DMRs in relation to whether or not the epiRILs exhibit reversion for at least one other TE-DMR of the same TE-family. (**E**) Number of potential 23-24nt sRNA sources belonging to the same TE family as the focal TE-DMRs, for each of the five reversion categories. (**F**) Genomic localization of three representative r-0%, r-mid, and r-100% TE-DMRs (indicated by triangles) and of related TE copies, including those that are potential sources of 23-24nt sRNAs (indicated by arcs). (**G**) Percentage of reverted TE-DMRs per TE-family relative to the proportion of related TEs that are WT-derived. (**H**) Average mCG levels of *ddm1*- or WT-derived TE-DMRs per TE family in relation to the epihaplotype at the single or two major QTL^{epi} #1 or #2. Relations are only shown for the 75 TE families for which the QTL^{epi} do not affect the methylation status of WT-derived TE-DMRs.

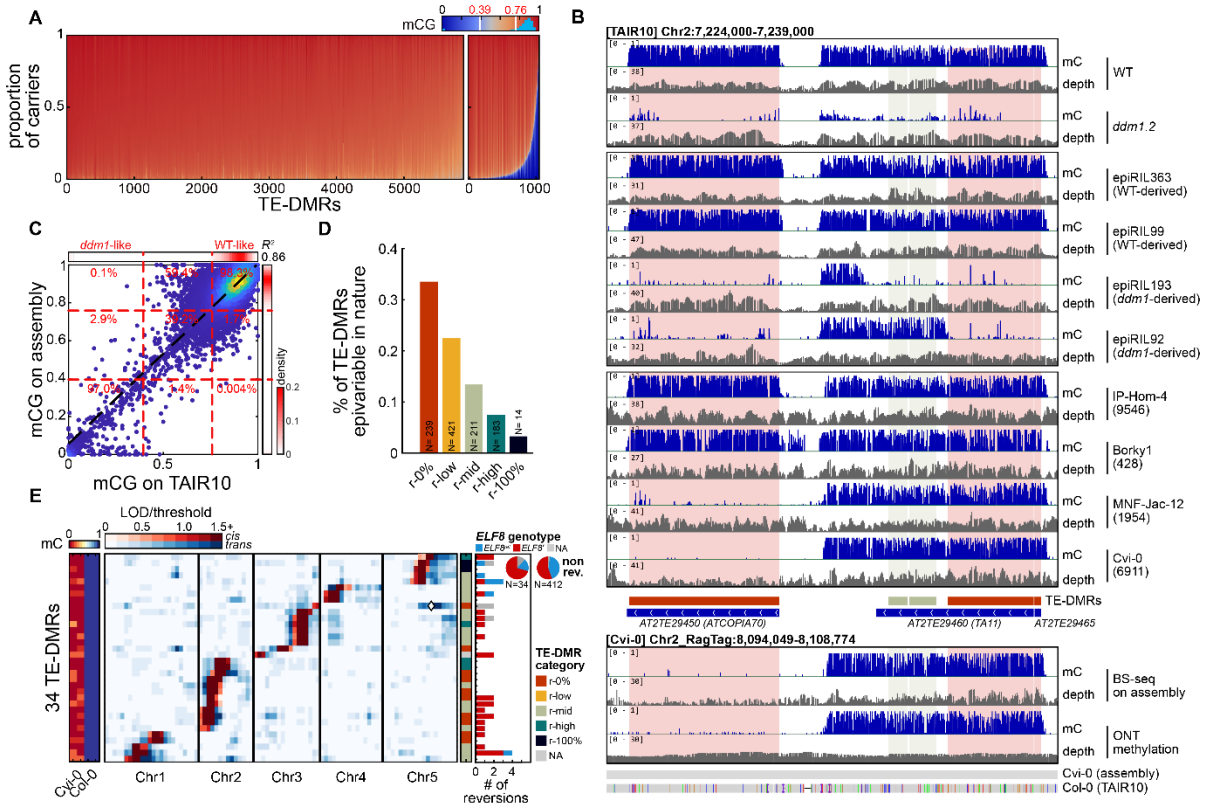

Fig. 3. TE epivariants are recurrent in nature.

Fig. 3. TE epivariants are recurrent in nature. (**A**) Distribution of average mCG levels over TE-DMRs in 720 natural strains. TE-DMRs with mCG under 0.39 were considered *ddm1*-like and over 0.76 as WT-like (density of average mCG levels is represented in cyan over the color bar). (**B**) BS-seq mC levels and coverage over three TE-DMRs in the WT and *ddm1* parents, four epihaplotypically contrasted epiRILs and four natural strains including two that exhibit epivariation over one (*ATCOPIA70*) of the three TE-DMRs. (lower panel) Alignment of TAIR10 against the Cvi-0 genome assembly in the corresponding region, with DNA methylation tracks from BS-seq data remapped on assembly and directly called from ONT sequencing data. (**C**) Comparison in 20 strains of average mCG levels measured by aligning BS-seq data on the de novo genome assembly vs on the TAIR10 reference genome assembly over TE-DMRs that are found at the reference location neither truncated nor missing in these 20 strains. The black dashed line represents linear regression. Red lines represent upper and lower thresholds for *ddm1*-like and WT-like methylation, respectively. Congruence is indicated by the percentage in each rectangle defined by these two thresholds. (**D**) Proportion of TE-DMRs with natural epivariation across the five reversion categories. (**E**) Heatmap of the mQTL mapping LOD scores in 36 Cvi-0 x Col-0 RILs for 34 TE-DMRs located across all five chromosomes. LOD score values are in shades of red or blue for associations in *cis* or trans, respectively. The unmethylated *ATCOPIA23* copy present in Cvi-0 on Chr5 that is closely related to that present methylated on Chr3 in Col-0 (fig. S7D), is indicated with a white diamond. The number of reversion events is represented for each TE-DMR as a bar colored by the *ELF8* genotype of the RILs in which they occurred. Frequency of reversion or stable inheritance by genotype are summarized in pie charts.

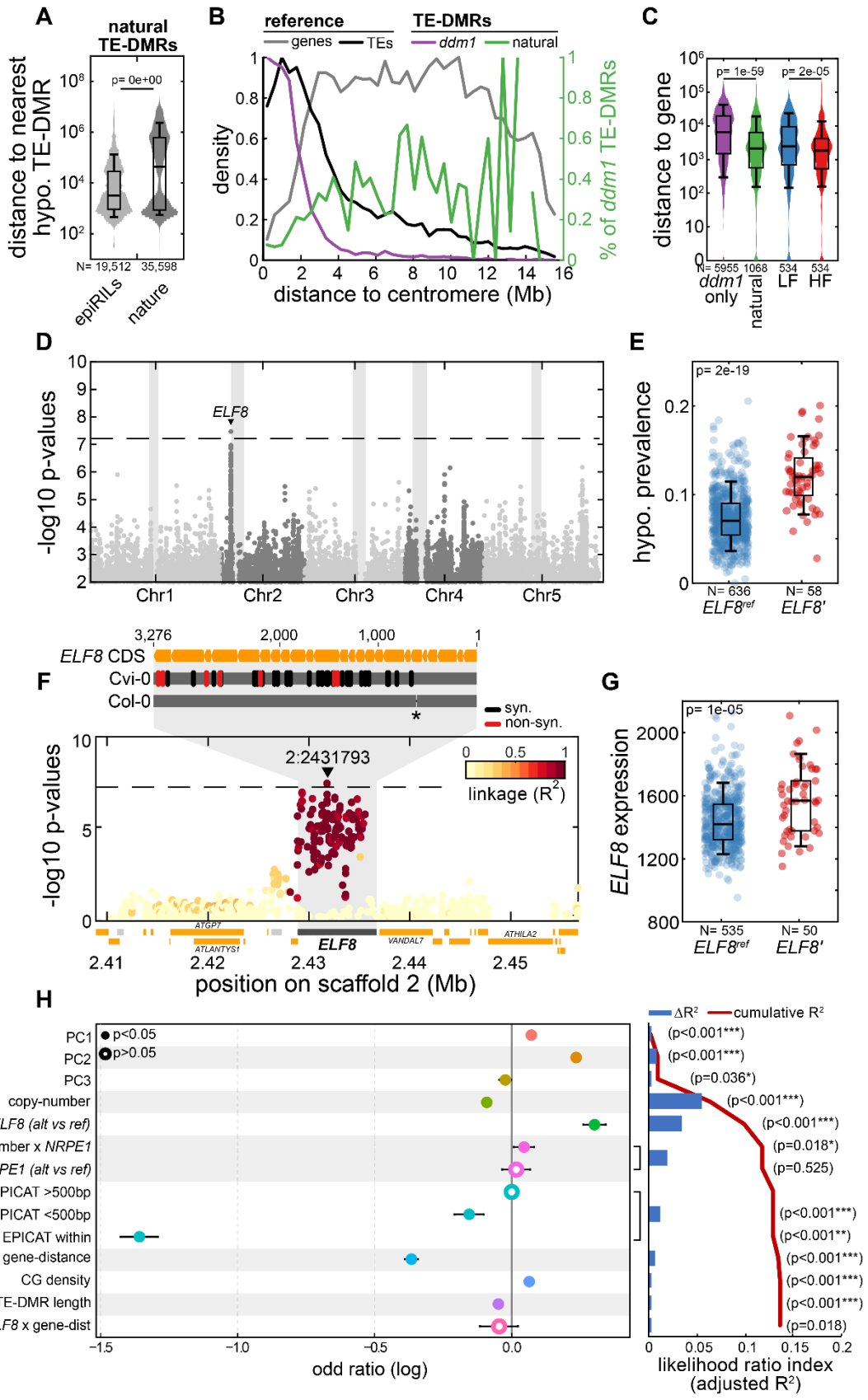

Fig. 4. Natural epivariants are enriched near genes.

Fig. 4. Natural epivariants are enriched near genes. (**A**) Distance between closest naturally epivariable TE-DMRs in the hypomethylated state in the epiRILs and in natural strains. (**B**) Metaplot of the density of TEs, genes, *ddm1*-induced TE-DMRs in relation to their distance from centromeres (left y-axis). The proportion of *ddm1*-induced TE-DMRs that are also epivariable in nature is represented in green (right y-axis). (**C**) Distance to the nearest gene (+1bp for log representation) of all *ddm1*-induced TE-DMRs, naturally epivariable TE-DMRs and those for which epivariation segregate at low or high frequency (LF, first two quartiles; HF, last two quartiles). (**D**) Manhattan-plot of p-values of GWAS of prevalence of natural epivariation by genome. (**E**) Prevalence of natural epivariation among strains carrying the reference *ELF8*^{ref} or non-reference *ELF8'* alleles. (**F**) Close-up Manhattan-plot of GWAS p-values around *ELF8*, colored by linkage with the leading SNP (2:2431793). Coding-sequence polymorphisms with the Cvi-0 *ELF8'* allele are indicated above. Asterisk marks a frame-shifting single-nucleotide deletion in the Col-0 TAIR10 sequence that is not detected in re-sequenced Col-0 *ELF8* cDNA (see supplementary text, note 7). (**G**) *ELF8* expression in strains carrying the *ELF8*^{ref} or *ELF8'* alleles. (**H**) Marginal effects (log odd ratio) and likelihood ratio index (McFadden adjusted R^2) of each explanatory variable in GLM of natural epivariation prevalence ranked by R^2.

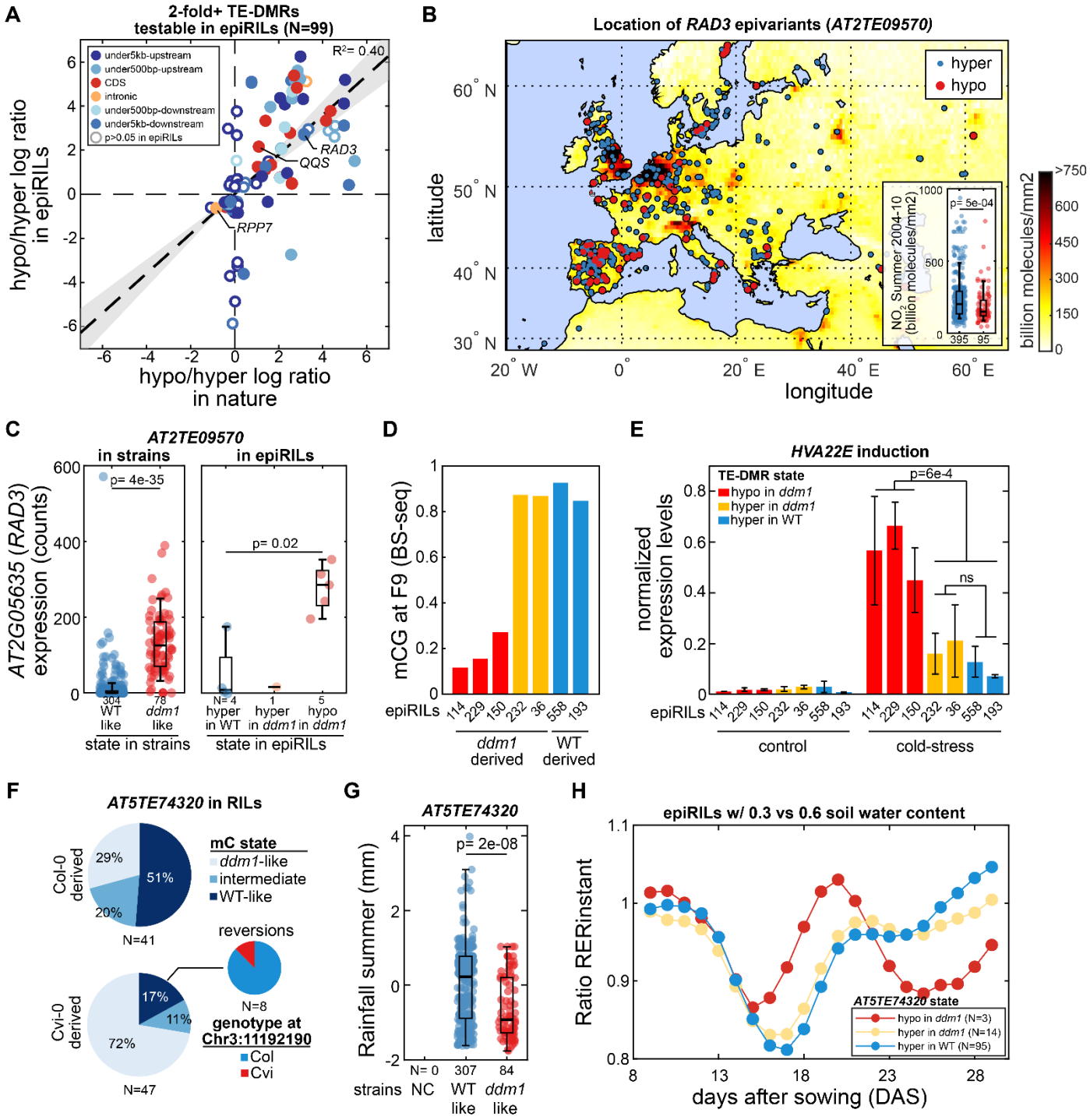

Fig. 5. Natural epivariation has functional consequences.

Fig. 5. Natural epivariation has functional consequences. (**A**) Comparison of the hypo/hyper log_{10} ratios measured in the epiRILs and in nature for each of the 99 naturally epivariable TE-DMR with ≥2-fold gene expression differences in nature and for which the 10 epiRILs are informative. Position of TE-DMRs relative to the affected genes is indicated by a color code. (**B**) Localization of Eurasian *A. thaliana* strains with epivariation at the *RAD3* locus against NO_2 atmospheric levels in June 2005 (AURA_NO2_M_2005-06-01 from https://neo.gsfc.nasa.gov/). Boxplot of average NO_2 atmospheric levels in summer (2004-2010) at the collection site of strains with or without epivariation at *RAD3* TE-DMR. (**C**) Comparison of *RAD3* expression level between strains with or without epivariation at the fixed *RAD3* TE-DMR (left panel) and between the 10 epiRILs according to the epiallelic state of the *RAD3* TE-DMR (right panel). (**D**) DNA methylation levels at the *AT5TE74320* TE-DMR located downstream of *HVA22E* in seven selected F9 epiRILs. (**E**) Expression levels of *HVA22E* measured in pools of F10 seedlings of the seven epiRILs grown under control conditions or exposed for 24h to cold-stress. (**F**) Methylation state of the *AT5TE74320* TE-DMR in 88 Cvi-0 x Col-0 RILs depending on its parental origin (left two pie charts). Genotype of top marker within the *ATHAT1* QTL^{epi} interval for 8 RILs with reversion (right pie chart). (**G**) Comparison of rainfall in summer (2001-2010) at the collection site of strains with or without epivariation at the fixed *AT5TE74320* TE-DMR (NC=0) downstream of *HVA22E*. The presence of outliers and their numbers are indicated by a red triangle. (**H**) Comparison of the difference in daily relative growth rate (RER ratio) of epiRILs grown on the Phenoscope in mild-drought vs well-watered conditions depending on the epiallelic state of the *HVA22E* TE-DMR.