Title: Transposable elements are vectors of recurrent transgenerational epigenetic inheritance in nature

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

Affiliations:

[1] Institut de Biologie de l'École Normale Supérieure, Université Paris Sciences et Lettres (PSL), UMR 8197 CNRS, U1024 INSERM ; Paris, 75005, France

[2] Université Paris-Saclay, INRAE, AgroParisTech, Institute Jean-Pierre Bourgin for Plant Sciences (IJPB) ; Versailles, 78000, France

[3] Gregor Mendel Institute of Molecular Plant Biology (GMI) of the Austrian Academy of Sciences; Vienna, 1030, Austria

[4] Evo-Eco-Paleo (EEP) Université de Lille, UMR 8198 CNRS ; Villeneuve d'Ascq, 59655, France.

[5] Department of Biology, Swiss Federal Institute of Technology (ETH), Zurich, CH-8092, Switzerland.

[6] Department of Genetics, University of Cambridge; Cambridge, CB2 3EH, United Kingdom

[7] Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles ; Paris, 75005, 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, 91190, France

*Corresponding authors. Email: pierre.baduel@ens.psl.eu; vincent.colot@ens.psl.eu

Abstract: Loss of DNA methylation over 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 at thousands of TE loci across the Arabidopsis thaliana genome, that experimentally-induced hypomethylation can be transmitted transgenerationally and reveal the role of small RNAs derived from related copies in counteracting this transmission. Using data from >700 strains collected worldwide, we uncover natural hypomethylation at hundreds of the same TE loci, often situated near stress-responsive genes. Like their experimental counterparts, 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. Crucially, we demonstrate that TE-mediated epiallelic variation associated with differential gene expression is generally causal and may be target of selection in specific environments, thus establishing its importance in nature.

Main Text:

Transposable elements (TEs) are ubiquitous components of eukaryotic genomes and most TEs are ancestral copies that have long lost their ability to transpose(*1*). In plants and mammals, TEs are typically targeted by DNA methylation and loss of this epigenetic modification can result not only in increased TE mobilization, but also, whether or not TEs are transpositionally active, in chromosome rearrangements as well as altered expression of neighboring genes(*2*, *3*). Moreover, plants, in contrast to mammals(*4*, *5*), do not extensively reset DNA methylation across generations(*6–8*), opening up the possibility for the accidental loss of this modification to be inherited as epialleles, i.e. independently of DNA sequence changes. Indeed, a number of plant epialleles have been reported(*9–12*), some of which with striking consequences on fitness-related traits such as fruit development in oil palm(*13*) or the onset of flowering in *A. thaliana(14–16)*. Crucially, most known epialleles involve ancestral, fixed or nearly-fixed TE insertions, thus raising the possibility that TE-mediated epiallelic variation has broad phenotypic impact in nature.

While undisputed in plants(*10*), TE-mediated transgenerational epigenetic inheritance (TEI) and its significance in nature remain poorly understood(*11*, *12*, *17*), in large part because of the difficulty to establish the independence of DNA methylation variation from DNA sequence polymorphisms(*18*). *Arabidopsis thaliana* offers a way forward however, thanks notably to several key resources and the deep understanding of the molecular players involved in DNA methylation of TEs developed in this model species. In particular, the Snf2-like chromatin remodeler DDM1 (DECREASE IN DNA METHYLATION 1), ortholog to human LSH/HELLS, 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*), and crucially, *ddm1*-induced loss of DNA methylation can be stably inherited in the absence of the mutation(*23*). On the basis of this property, which most likely results from the disruption of the interplay between DDM1 and histone variants(*24*), a population of isogenic *ddm1*-derived epigenetic recombinant inbred lines (epiRILs) was generated that uncovered a large reservoir of potential epialleles affecting complex traits(*25–27*).

Here, we used this experimental epiRIL population to investigate the inheritance of DNA hypomethylation at thousands of TE loci across the genome and to characterize the molecular underpinnings of the wide differences observed. We then combined the knowledge gained in the epiRILs on TE-mediated epiallelic variation 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 extent, dynamics, and implications of TE-mediated TEI in nature.

Inheritance patterns of *ddm1*-induced hypomethylation differ markedly 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. The two symmetrical sites CG and CWG (where W=A or T) offer the most robust estimates of methylation levels and we identified 25,462 regions with markedly reduced methylation at these sites in *ddm1* (Fig. S1-2, Dataset S1-2). 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 (Table S1), 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, S2e, S3) using as genetic markers an independent set of 1,092 DMRs that are regularly spaced along the five chromosomes and which segregate the two parental methylation states in a manner compatible with Mendelian inheritance (Dataset S3). In a last step, we attributed for each TE-DMR one of three possible DNA methylation states in the epiRILs: WT-like, *ddm1*-like or intermediate. Specifically, we compared the average CG methylation level in each epiRIL with that in the two parental lines, as CG sites exhibit the largest differences of methylation between WT and *ddm1* (Fig. S1c).

As previously reported using low-resolution DNA methylation data(*25*, *26*, *33*), while WT methylation is almost systematically inherited in the epiRILs (Fig. 1b, see Supplementary Note 1 for exceptions), this is not the case for *ddm1*-induced hypomethylation. Indeed, the vast majority of *ddm1*-derived TE-DMRs (N=6,311 or 89.1%) exhibit 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 Note 2).

Reversion is conspicuous but far from systematic at generation F9, except for a small number (395) of TE-DMRs and we defined five categories of TE-DMRs to account for the wide range of reversion frequency at this generation: r-0%, r-low, r-mid, r-high, and r-100% (Fig. 1d). However, the observation that some r-0% TE-DMRs have reverted by generation F21 (Fig. 1b-c, S1e) suggests that *ddm1*-induced hypomethylation may not be inherited indefinitely in the absence of selection even at TE-DMRs of the most stable category.

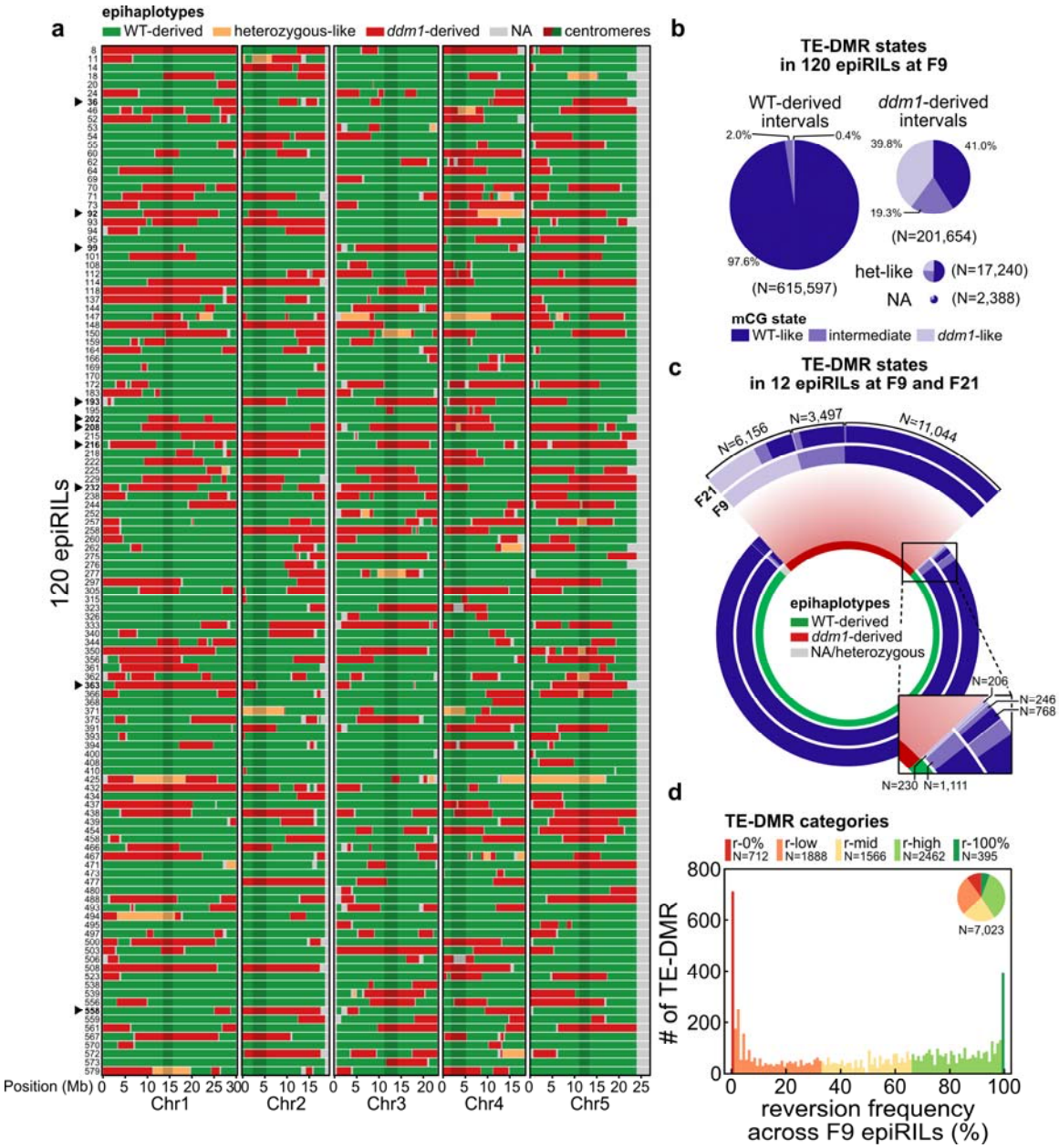

Figure 1. Inheritance patterns of *ddm1*-induced hypomethylation

(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.

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

Previous work established that the RNA-directed DNA methylation (RdDM) pathway, which involves 23-24nt small RNAs(*19*), is required for reversion(*33*). To investigate further the

molecular underpinnings of reversion, we sequenced small RNAs from the two parental lines and 10 epiRILs. Consistent with earlier findings(*33*), the abundance of 23-24nt sRNAs in *ddm1* correlates positively with reversion (Fig. 2a, S4a), as a result of the drastic loss of 23-24nt sRNAs matching TE-DMRs that belong to the r-0% or r-low categories in this mutant background (Fig. 2a, S4b). Moreover, our new analysis reveals that this positive correlation is strongest for multiple-mapping 23-24nt sRNAs, which correlate also strongly with mCHH levels in *ddm1* (Fig. 2b, S4c). Given that the residual CHH methylation present in *ddm1* depends almost entirely on RdDM(*22*, *33*), the last two findings suggest a predominant role of multiple-mapping 23-24nt sRNAs in guiding RdDM activity and thus reversion. Consistent with this interpretation, reversion often affects related TE-DMRs together in the epiRILs (Fig. 2c) and it is more strongly associated with a higher abundance of multiple- than of single-mapping 23-24nt sRNAs not only in *ddm1* (Fig. 2a, S4a) but also in the epiRILs (Fig. 2d). Indeed, reversion is not accompanied by the restoration of WT levels of single-mapping 23-24nt sRNAs in some instances (Fig. S4d-e).

To identify the potential sources of *trans*-acting 23-24nt sRNAs that guide 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 match perfectly to TEs that belong to the same TE family or superfamily as the target TE-DMR (Fig. S5a) and unsurprisingly, their number also correlates positively with reversion (Fig. 2e-f, S5b-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). This last observation indicates that WT-derived sources are the main drivers of reversion, in agreement with the fact that most 23-24nt sRNAs matching TE-DMRs of the r-0% to r-mid categories are lost in *ddm1*.

To determine if some TEs play predominant roles in reversion, we performed QTL mapping for each of the 105 TE families with a sufficient number of (≥10) non-systematically reverting TE-DMRs, taking as a trait the average CG methylation of the TE-DMRs inherited from *ddm1*. Using this approach, we identified the presence of one or at most two major QTL^{epi} for almost all (99) TE families. QTL^{epi} intervals tend to overlap with TE-rich pericentromeres (Fig. S5d) and the WT-derived epigenotype in each case explains on average 13% of variation in family-wide reversion (Fig. 2h-i, S5e-f). Moreover, almost all (92%) QTL^{epi} encompass one or several of the potential sources of *trans*-acting sRNAs (Fig. S5d), which suggest that they are causal. Conversely, the *ddm1*-derived QTL^{epi} do not impact WT-derived TE-DMRs, except in rare cases that mainly concern TE families with known *trans*-demethylation activity (*VANDAL* and *ATENSPM* families, Fig. S5d-e, see Supplementary Note 1).

An important implication from the QTL mapping results is that the hypomethylated epiallelic state of non-systematically reverting TE-DMRs should be stable in epiRILs where the main sources of *trans*-acting 23-24nt sRNAs also derive 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 distinct TE families and have distinct reversion frequencies for analysis. Importantly, three TE-DMRs are members of TE families for which we identified at least one QTL^{epi} (Fig. S5d) and hypomethylation co-segregates with the *ddm1*-derived top QTL^{epi} in each case (Fig. S5g). Using targeted DNA methylation analysis, we detected zero or one reversion events among the F9 siblings, even for the two r-high TE-DMRs (Fig. S5f). These results indicate that hypomethylation can be stably transmitted in the absence of the sources of matching

small RNAs that guide reversion. In other words, the wide differences in reversion frequency observed between TE-DMRs at generation F9 are determined primarily in *trans* rather than in *cis*.

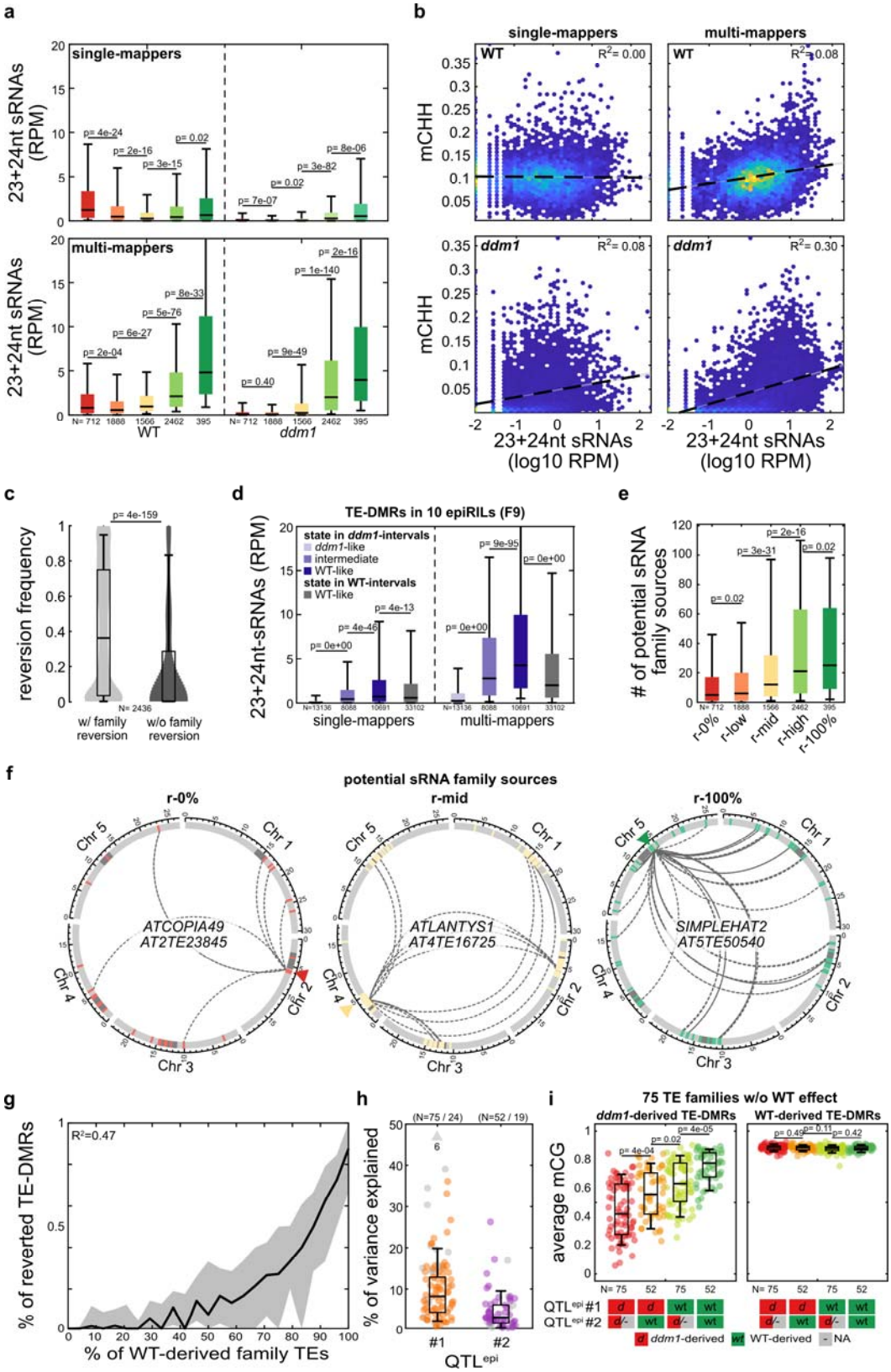

Figure 2. Reversion is driven mainly by *trans*-acting sRNAs

(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- (left) and multi-mapping (right) sRNAs in WT and *ddm1* (log10 RPM). The coefficient of correlation (R^2) of the linear regression (dotted line) is indicated. (c) 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. (d) 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. (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) Percentage of variance explained iteratively by the first and the second QTL^{epi} in the average mCG levels of *ddm1*-derived, non-systematically reverting TE-DMRs within a given TE family. Values for the 24 TE families for which the QTL^{epi} affects the methylation status of WT-derived TE-DMRs are shown in gray, with outliers and their number indicated with a triangle. (i) 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.

Transcriptional activity favors the inheritance but not the spontaneous occurrence of TE hypomethylation

Although reversion is usually concerted at the TE-family level (Fig. 2d), 369 TE-DMRs remained in the hypomethylated epiallelic state in epiRILs where over >75% of related TE-DMRs have reverted (Fig. S6a). Such “stray” TE-DMRs tend to have higher levels of sequence divergence with their closest relative and fewer close relatives altogether (Fig. 3a-b), in agreement with the known reduction of RdDM efficiency when source and target diverge in sequence(*35*). 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(*36*); Fig. 3c-d, S6b-c, Supplementary Note 3), in keeping with the known inhibitory effect of Pol II transcription on RdDM(*37–39*).

To determine whether RdDM targeting may also affect the spontaneous occurrence of the hypomethylated state of TE-DMRs, we considered the 106 TE-DMRs with severe loss of parental WT methylation in one or two epiRILs only and that do not belong to TE families with known *trans*-demethylation activity (Table S4). To increase our sample size, we exploited methylome data independently obtained at generation F9 for a minimally overlapping set of 169 epiRILs(*40*) and identified in this set another 114 TE-DMRs with similar spontaneous *ddm1*-like hypomethylation (Table S4). The majority (65%) of the 207 TE-DMRs we identified in total belong to the r-0% and r-low categories (Fig. 3e) and remarkably, spontaneous hypomethylation is far less frequent (2.3-fold reduction) when all rather than some or none of the potential sources of *trans*-acting 23-24nt sRNAs are WT-derived (Fig. 3f). 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 (Fig. S6d).

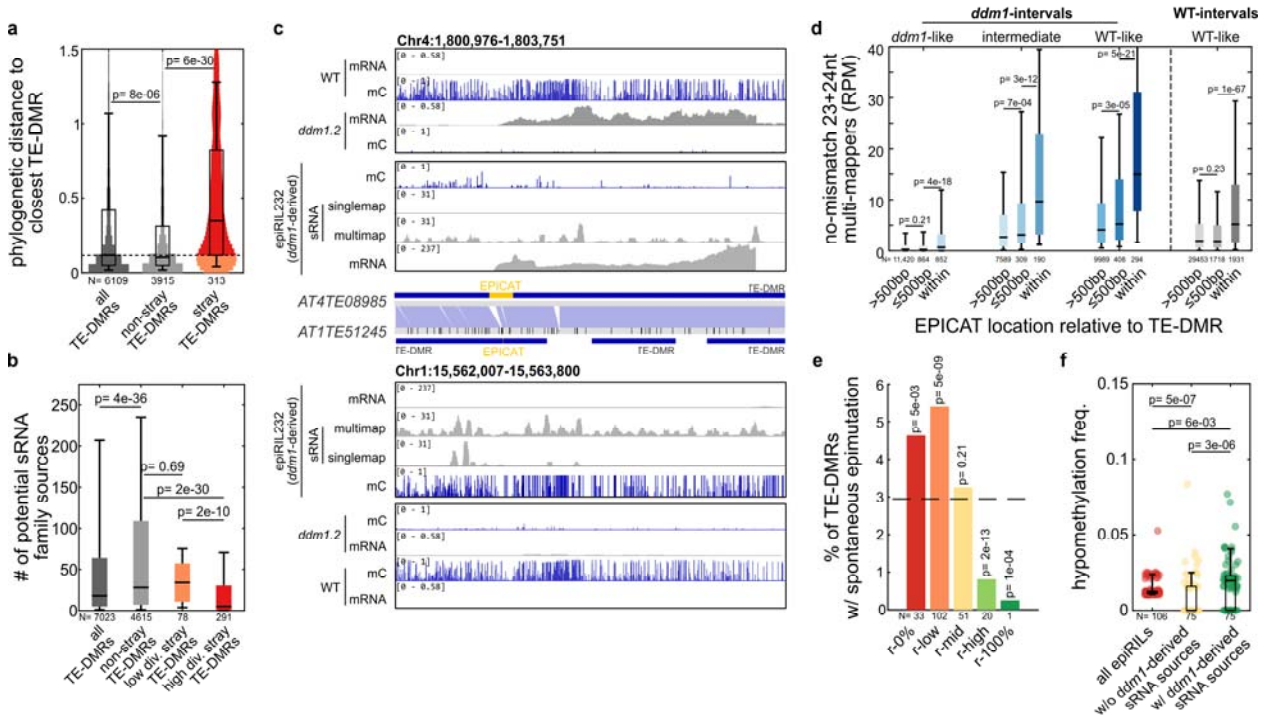

Figure 3. Inheritance and spontaneous occurrence of TE hypomethylation is modulated by two main antagonistic features

(a) Phylogenetic distance to closest TE-DMR for all, stray, and non-stray TE-DMRs. Stray TE-DMRs with distance lower and higher than the median are highlighted in orange ("low div.") and red ("high div."), respectively . (b) Number of potential 23-24 nt sRNA sources belonging to the same TE family for all, non-stray, "low div." stray, and "high div." stray TE-DMRs. (c) Levels of mRNAs(*36*), DNA methylation and 23-24 nt sRNAs, over a stray TE-DMR (*AT4TE08985*) that contains a long and strong *ddm1* EPICAT and which does not revert in epiRIL232, despite high levels of multi-mapping sRNAs and the presence in the same epiRIL of a closely related TE-DMR (*AT1TTE51245*) that has reverted. (d) Levels of multi-mapping 23-24nt sRNAs over TE-DMRs in 10 sequenced epiRILs (F9), in relation to their methylation state when they are WT- or *ddm1*-derived and as a function of their distance relative to the nearest EPICAT. (e) Percentage across the five reversion categories of TE-DMRs with spontaneous *ddm1*-like hypomethylation in the 120 epiRILs and mostly non-overlapping 169 epiRILs from Zhang et al.(*40*). The p-values indicate the cumulative binomial probability of observing similar or more extreme representation from random sampling (dashed line). (f) Frequency of spontaneous *ddm1*-like hypomethylation among all 120 epiRILs, only those where no related TE-DMR is *ddm1*-derived, or only those where at least one related TE-DMR is *ddm1*-derived. The presence of outliers and their numbers are indicated by a gray triangle.

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(*41*, *42*). Although TE-DMRs are generally part of TEs that are molecular fossils(*42–44*) and are thus fixed or nearly so at the species level (Fig. S7a-b), we

retained for analysis only about half of them in any given strain because of the stringent filtering we applied to ensure robust measures of DNA methylation (Fig. S7c, Dataset S5).

In keeping with heavy and co-extensive DNA methylation being generally the default state for TEs in *A. thaliana(29, 45)*, the vast majority (85%) of TE-DMRs have high average CG methylation across all informative strains (Fig. 4a). Nevertheless, 1,068 TE-DMRs exhibit severe, *ddm1*-like CG hypomethylation in at least one strain (Table S5). Most of these natural epivariants likely result from rare and rapid loss of DNA methylation, given that they usually have low population frequencies (Fig. 4a; Fig. S7d) and that intermediate levels of average CG methylation are seldom observed. Importantly, DNA methylation loss extends to CHG and CHH sites, as in *ddm1* (Fig. S7e,f), although the exact boundaries of this loss may differ (e.g. Fig. 4b). In addition, long-read genome sequencing of 20 genetically diverse strains (Table S6) indicates that the mismapping issues typically generated by the presence of related TEs in non-reference strains(*46*) affects only a small fraction (<15%) of the natural epivariants we called in each strain, thanks to our stringent filtering (Fig. 4c, see Supplementary Note 4).

Naturally epivariable TE-DMRs are strongly skewed towards the r-0% and r-low categories (Fig. 4d), 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. S8a-b), yet none of these are systematic (Fig. S8c), thus ruling out a strict dependence of natural epivariation on neighboring SNPs.

To obtain direct evidence that 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(*47*). This RIL population is particularly informative in this respect, as Cvi-0 harbors the second largest number of epivariants (113 in total, Fig. S7c) among natural strains and also because almost all (110) of the corresponding TE-DMRs are hypomethylated in at least one other strain (Table S7). Targeted (Fig. 4e) as well as whole-genome (Fig. S8e) 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 Note 5 for the one exception). In addition, many of the Cvi-0-derived TE-DMRs have regained dense DNA methylation in at least one RIL (Fig. 4e, S8e), 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. S9a). Furthermore, local haplotyping indicates that for the majority (66%) of TE-DMRs, epivariants are carried by at least two distinct alleles (Fig. S9b-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 heritable indefinitely in the absence of selection.

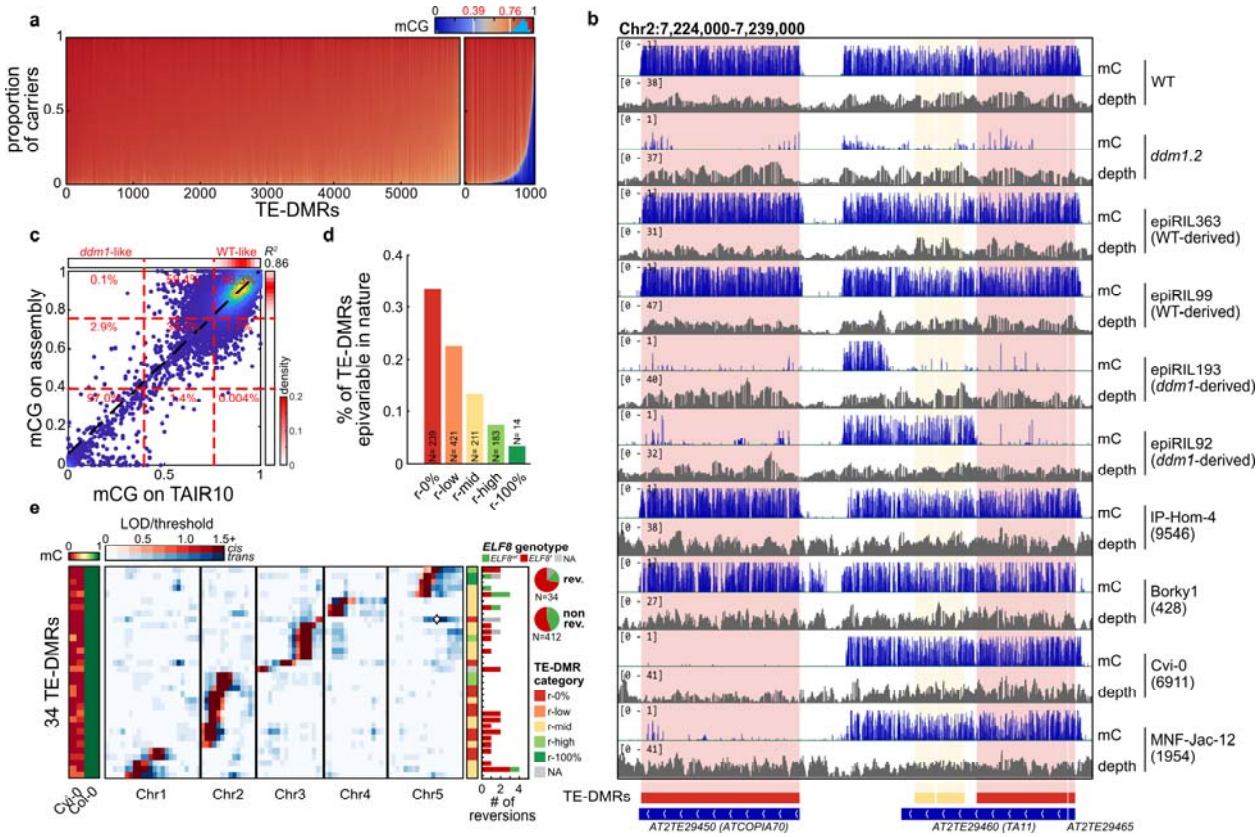

Figure 4. TE epivariants are recurrent in nature

(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 colorbar). (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 in which two exhibit epivariation over one (*ATCOPIA70)* of the three TE-DMRs. (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 (*42*) on Chr5 that is closely related to that present methylated on Chr3 in Col-0 (Fig. S8d), 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.

Natural epivariants are prevalent near genes

Unlike in the epiRILs, natural hypomethylated epivariants are generally few in any given strain (49 on average, Fig. 4b, S7c) and tend to be either tightly clustered (within 10kb) and correlated between strains, or else dispersed and uncorrelated (Fig. 5a, S9e-f). In addition, naturally epivariable TE-DMRs are usually less pericentromeric (Fig. 5b), hence closer to genes than the complete set of experimental TE-DMRs (Fig. 5c). Importantly, the population frequency of natural epivariants correlate positively with gene proximity, suggesting that it facilitates either

their occurrence or inheritance, or both. Further supporting this interpretation, 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. 5d-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(*48*), which in plants requires Paf1(*49*), 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. S10a-h and Supplementary Note 6 for functional analysis) tend to harbor a higher number of natural epivariants (Fig. 5e,g) and this trend is most pronounced for TE-DMRs that are located within 2kb of genes (Fig. S10i). 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. 4e, S10j-k, see Supplementary Note 7). 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. 5h). The negative correlation with TE-copy number is consistent with RdDM strength increasing with TE copy-number(*50, 51*) and the role of RdDM in limiting both the inheritance and spontaneous occurrence of hypomethylation at TE-DMRs in the epiRILs (Fig. 2-3, S5-6). Indeed, our GLM revealed that the weak allele of the RdDM gene *NRPE1(42, 52)* mitigates the negative effect of TE copy-number on epivariation. Moreover, it also confirmed that the impact of gene proximity is stronger in strains carrying the *ELF8'* rather than the $ELF8^{ref}$ allele (Fig. S10i). Together, these findings highlight the pivotal and antagonistic roles of RdDM and Paf1 activity in shaping the prevalence of epivariation in nature.

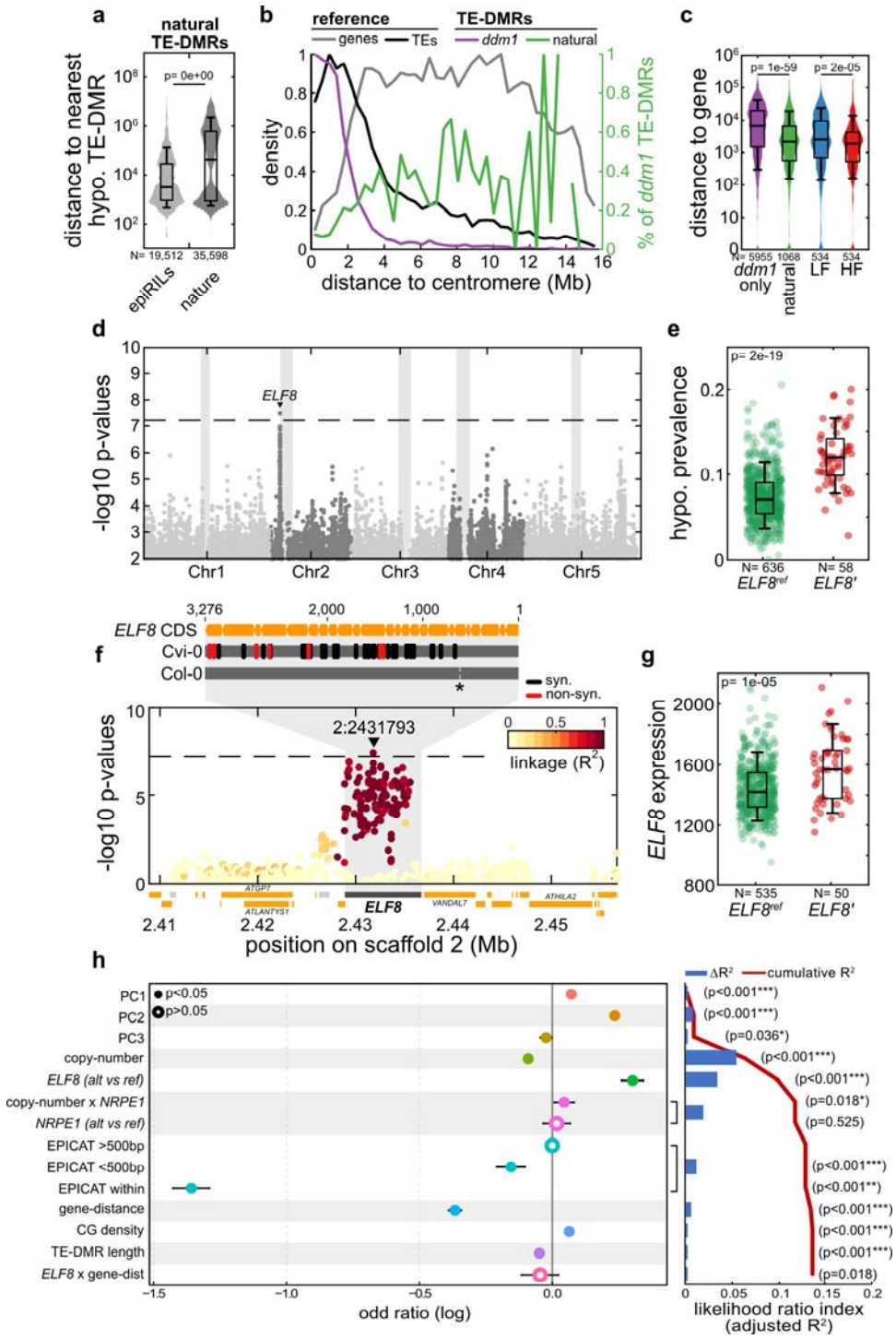

Figure 5. Natural epivariants are enriched near genes

(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 Note 6). (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.

Natural epialleles impact gene expression

To investigate the functional impact of natural epivariation, 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. S11a). These changes were strongly skewed towards upregulation, notably when the TE-DMRs overlap with genes or are located immediately upstream (≤500bp) of them (Fig. S11a-b). To distinguish between cause and consequence, without the possible confounding effects of DNA sequence variants, we produced RNA-seq data from the 10 epiRILs we used for the sRNA studies and asked whether the experimental counterparts of epivariants associated with changes in gene expression in nature cause similar 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. 6a, S11c) thus explaining 24% on average of the variation in expression levels observed between natural strains (Table S10). In addition, congruence concerns 30 of the 32 TE-DMRs located within or upstream (≤500bp) of genes (Fig. 6a, S11c) and for 14 of the 17 TE-DMRs with a reversion event in the 10 epiRILs sampled, WT expression is regained too (Fig. 6b, S11d). 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 causality have no known function and the phenotypic consequences of their misregulation remain to be determined. Among those with known function (Table S10), two were previously shown to be affected by the epigenetic state of the TE sequences they contain (*QQS(53)* and *RPP7(54)*). While the phenotypic impact of the epiallelic upregulation of *QQS* (Fig. S11e) remains to be determined, the inhibition of splicing caused by the loss of heterochromatin at the intronic TE present in *RPP7* (Fig. S11f), was shown to considerably reduce RPP7-mediated defence against pathogens(*55*). Thus, epiallelic variation at *RPP7* may be deleterious in natural settings, and consistent with this interpretation, we only observed it once among informative strains, despite the TE-DMR being of the r-low category. *RAD3* (*AT2G05635*), 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*). Moreover, *RAD3* has been implicated in natural variation for root growth under conditions of acid mine drainage(*56*). Our transcriptomic analysis also shows that loss of DNA methylation at *RAD3* results in more abundant and longer transcripts (Fig. 6b, S12a) and is associated in the epiRILs with reduced growth of the primary root (Fig. S12b). Importantly, epivariants of *RAD3* are relatively frequent in nature in nature (95 strains) and biogeoclimatic data(*57*) indicate that they are underrepresented in locations with high NO_2 emission levels in summer (Fig. 6c, S12c-d). As NO_2 emissions correlate positively with levels of soil acidification(*58*), it is therefore tempting to speculate that hypomethylation of the *RAD3* TE-DMR is selected against in habitats with acidic

soils. Thus, in addition to allelic variation, epiallelic variation at *RAD3* has phenotypic consequences that could be ecologically relevant.

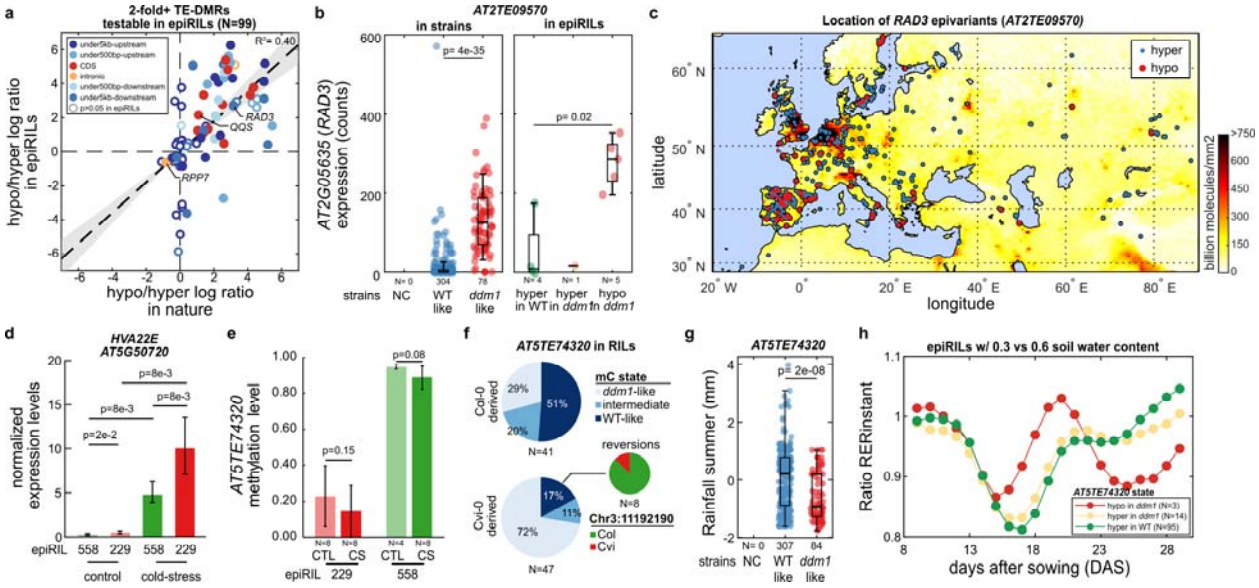

Figure 6. Natural epivariation has functional consequences

(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) 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). (c) 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 neo.gsfc.nasa.gov) (d) Expression level of *HVA22E* measured in three independent pools of 8 F10 seedlings of each of the epiRIL229 and epiRIL558 (with and without *ddm1*-induced hypomethylation of the downstream TE-DMR, respectively) grown under control conditions or exposed for 24h to cold-stress. (e) DNA methylation levels at *AT5TE74320* TE-DMR measured in 4-8 F10 individuals of the epiRIL229 and epiRIL558 grown under control conditions or exposed to cold-stress for 24h. (f) Methylation states of *AT5TE74320* within 88 Cvi-0 x Col-0 RILs (*47*) depending on genotypes at both bordering markers (left two pie charts). Genotype at top marker within 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 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.

Natural epivariation may be selected to facilitate stress responses

Given that naturally epivariable TE-DMRs are particularly enriched near stress-responsive genes (Fig. S12e), we asked if epivariation could influence gene expression 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. S13a). In each of the two pairs of epiRILs that we investigated with contrasted parental epiallelic states at the TE-DMR, *HVA22E* expression is similarly low in the absence of stress, but its induction by cold is markedly higher (1.6 and 2.4-fold) when the TE-DMR is hypomethylated (Fig. 6d, S13b). Furthermore, cold treatment does not result in a loss of DNA methylation in the case of the WT-derived TE-DMR

(Fig. 6e, S13c). Altogether, these results reveal that the hypomethylated epiallele at *HVA22E* enables plants to enhance their response to cold stress, without affecting their basal expression.

Hypomethylation at The TE-DMR downstream of *HVA22E* is quite common (24%) among natural strains, despite reversion being frequent (66%) in the epiRILs (Table S1), which suggest the presence of stabilizing genetic or genomic modifiers, or positive selection in nature. 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 88 Cvi-0 x Col-0 RILs. The Cvi-0 allele remained hypomethylated in 72% of the of the RILs that inherited it but regained high levels of DNA methylation in eight of them (Fig. 6f), seven of which harbor the Col-0 haplotype at the top marker within the QTL^{epi} interval we identified for this TE family in the epiRILs (Fig. S13d). These observations, together with the fact that the Col-0-derived allele loses DNA methylation in 29% of RILs indicate that the Cvi-0 genotype stabilizes hypomethylation at *HVA22E* and conversely, that the Col-0 QTL^{epi} has a strong destabilizing effect.

We observed that hypomethylation at *HVA22E* is preferentially found in habitats with high summer frost and low summer precipitation (Fig. 6g, S13e-f). Furthermore, analysis of phenotypic data compiled in the AraPheno database (https://arapheno.1001genomes.org/) indicates that strains with hypomethylation at the locus tend to have smaller stomata (Fig. S13f-g), a cell-type where *HVA22E* is strongly induced by drought(*59*). 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(*60*) to monitor the growth of 112 epiRILs subjected or not to mild drought, including three with a *ddm1*-derived epiallele and 14 where the *ddm1*-derived TE-DMR 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. 6h, S13h), indicating that it could act as a selectable engine of enhanced response to stress.

Discussion

By combining experimental and natural population epigenomics, we conducted a comprehensive study of the molecular underpinnings and biological significance of TE-mediated TEI in plants. Using the *ddm1*-derived epiRILs, we demonstrated that inheritance of hypomethylation at a given TE is determined in *trans*, because RdDM, which drives reversion, is guided by 23-24nt sRNAs that derive from related TE copies mainly. On the basis of this new knowledge obtained for thousands of TE loci across the genome, most of whom are ancestral, we then interrogated the presence and functional impact of natural epivariants with similar loss of DNA methylation over the same TE loci in nature. Despite stringent filtering and biases inherent to our study design (see Supplementary Note 8), we identified hundreds of TE loci with natural *ddm1*-like epivariation and showed that they tend to be located near stress-responsive genes. These natural epivariants reflect abrupt loss rather than gradual erosion of DNA methylation across generations and most of them are recurrent, *bona fide* epialleles that are relatively short-lived, like their experimental counterparts. Moreover, epiallelic variation at TE loci associated with changes in gene expression in *cis* is generally causal, especially when the TEs are located within or upstream of genes. Finally, we showed that effects can be more pronounced in response to stress or specific to it and that causal epialleles are not randomly distributed across habitats, suggesting selection is at play. Our study therefore provides a detailed understanding of the mechanisms responsible for TE-mediated TEI in plants and demonstrates the importance of this additional system of inheritance in nature.

The demonstration that sRNAs derived from related TEs play a prominent role in limiting TE-mediated TEI is in line with 23-24nt sRNAs being required in large amounts for efficient RdDM-targeting of transgenes(*37*, *61–64*), indicates that the determinants of epigenetic stability at a given TE locus are intrinsically polygenic. Thus, the genome could be viewed as an epigenetic biome in which the epiallelic properties of a given locus relies on interactions, additive or not, between multiple loci. Furthermore, the observation that pericentromeric TEs are the main source of *trans*-acting sRNAs and have little impact in *cis* is consistent with methylome remodeling data obtained in epigenetic F1 hybrids(*65*) and lends direct support to the proposition that these sRNAs mainly serve to guide DNA methylation of related copies located on chromosome arms(*66–69*). More generally, the finding that almost all TEs we analysed can sustain TEI in at least one epiRIL indicate that a one-to-one ratio of sources of 23-24nt sRNAs to targets is usually insufficient to direct RdDM to these, unlike in "classical" paramutation, which involves allelic or epiallelic interactions, such as at the *b* locus in maize(*70*). As a matter of fact, classical paramutation is rare in the F1 progeny of crosses between *A. thaliana* strains(*71*), which our analysis of the Cvi-0 x Col-0 RILs also indicate, and it may be the exception rather than the rule, as previously proposed based on more limited data (*37*, *72–74*).

Our analysis of natural epivariation revealed that 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 (75, 76)* and *A. thaliana(48)* 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(*49*), Paf1 may also help safeguard genes from the silencing of neighboring TEs.

We showed that natural epialleles can alter the expression of neighboring genes and that their effects may be exacerbated in response to stress. Furthermore, their presence does correlate with specific environments in some cases, in line with evidence indicating that despite their lower stability compared to DNA sequence variants, natural epialleles can be targets of selection (*27*, *77*). Given their recurrence and singular inheritance properties, these natural epialleles therefore may contribute uniquely to local adaptation, especially in cases where their effects on gene expression are minimal in the absence of stress. However, determining the extent of this contribution remains a major challenge, because of the difficulty in teasing apart the effects of epialleles and DNA sequence variants in natural populations. Moreover, a key open question is whether epiallelic variation may be facilitated or stabilized in specific environments, as evidence in support of such effects is still lacking (this study and(*78*)), despite well-documented effects of biotic or abiotic stresses on the heritable loss of methylation at individual cytosines(*79–82*). Whatever the case might be, our findings clearly establish that ancestral TEs are an important and unique source of recurrent heritable variation in plants, with broad implications for conservation efforts and breeding programs.

References and Notes

1. J. N. Wells, C. Feschotte, A Field Guide to Eukaryotic Transposable Elements. Annu. Rev. Genet. 54, 539–561 (2020).

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).

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).

4. 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. Philos. Trans. R. Soc. Lond. B Biol. Sci. 368, 20110330 (2013).

5. 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).

6. M. Gehring, Epigenetic dynamics during flowering plant reproduction: evidence for reprogramming? New Phytol. 224, 91–96 (2019).

7. S. He, X. Feng, DNA methylation dynamics during germline development. J. Integr. Plant Biol. 64, 2240–2251 (2022).

8. H. T. Chow, R. A. Mosher, Small RNA-mediated DNA methylation during plant reproduction. Plant Cell 35, 1787–1800 (2023).

9. E. J. Richards, Inherited epigenetic variation — revisiting soft inheritance. Nat. Rev. Genet. 7, 395–401 (2006).

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

11. L. Quadrana, V. Colot, Plant Transgenerational Epigenetics. Annu. Rev. Genet. 50, 467–491 (2016).

12. P. Baduel, V. Colot, The epiallelic potential of transposable elements and its evolutionary significance in plants. Philos. Trans. R. Soc. Lond. B Biol. Sci. 376, 20200123 (2021).

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).

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).

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).

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).

17. D. Weigel, V. Colot, Epialleles in plant evolution. Genome Biol. 13, 249 (2012).

18. F. Johannes, V. Colot, R. C. Jansen, Epigenome dynamics: a quantitative genetics perspective. Nat. Rev. Genet. 9, 883–890 (2008).

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).

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).

21. 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).

22. 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).

23. A. Vongs, T. Kakutani, R. A. Martienssen, E. J. Richards, Arabidopsis thaliana DNA methylation mutants. Science 260, 1926–1928 (1993).

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).

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).

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).

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).

28. 1001 Genomes Consortium. Electronic address: magnus.nordborg@gmi.oeaw.ac.at, 1001 Genomes Consortium, 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell 166, 481–491 (2016).

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).

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).

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).

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).

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. [Preprint] (2009). https://doi.org/10.1126/science.1165313.

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).

35. 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).

36. 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).

37. F. K. Teixeira, V. Colot, Repeat elements and the Arabidopsis DNA methylation landscape. [Preprint] (2010). https://doi.org/10.1038/hdy.2010.52.

38. 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).

39. 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).

40. 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).

41. 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 (2016).

42. 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).

43. F. Maumus, H. Quesneville, Ancestral repeats have shaped epigenome and genome composition for millions of years in Arabidopsis thaliana. Nat. Commun. 5, 1–9 (2014).

44. 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 (2016).

45. 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).

46. 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, 1–19 (2023).

47. 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).

48. 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).

49. 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).

50. N. C. Deans, J.-E. 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).

51. V. L. Chandler, W. B. Eggleston, J. E. Dorweiler, Paramutation in maize. Plant Mol. Biol. 43, 121–145 (2000).

52. 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).

53. 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).

54. 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).

55. 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, E3535–43 (2013).

56. 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).

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

58. 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).

59. 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).

60. 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).

61. D. M. Bond, D. C. Baulcombe, Small RNAs and heritable epigenetic variation in plants. [Preprint] (2014). https://doi.org/10.1016/j.tcb.2013.08.001.

62. M. A. Matzke, R. A. Mosher, RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. [Preprint] (2014). https://doi.org/10.1038/nrg3683.

63. J. M. Wendte, C. S. Pikaard, The RNAs of RNA-directed DNA methylation. [Preprint] (2017). https://doi.org/10.1016/j.bbagrm.2016.08.004.

64. K. Osabe, Y. Harukawa, S. Miura, H. Saze, Epigenetic Regulation of Intronic Transgenes in Arabidopsis. Sci. Rep. 7, 1–13 (2017).

65. 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 (2024).

66. 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).

67. 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).

68. 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).

69. 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).

70. J. B. Hollick, Paramutation and related phenomena in diverse species. Nat. Rev. Genet. 18, 5–23 (2017).

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).

72. R. Martienssen, Epigenetic phenomena: paramutation and gene silencing in plants. Curr. Biol. 6, 810–813 (1996).

73. J. B. Hollick, J. E. Dorweiler, V. L. Chandler, Paramutation and related allelic interactions. Trends Genet. 13, 302–308 (1997).

74. 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).

75. 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).

76. 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).

77. 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).

78. M. W. Schmid, K. Kropivšek, S. E. Wuest, B. Schmid, U. Grossniklaus, Weak evidence for heritable changes in response to selection by aphids inArabidopsisaccessions, bioRxiv (2025). https://doi.org/10.1101/2025.01.21.634027.

79. 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).

80. 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).

81. 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).

82. 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 (2016).

83. F. Krueger, S. R. Andrews, Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27, 1571–1572 (2011).

84. 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).

85. 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).

86. Z. Sun, H. Li, L. Zhang, J. Wang, Estimation of recombination frequency in bi-parental genetic populations. Genet. Res. 94, 163–177 (2012).

87. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).

88. 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).

89. 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).

90. K. Katoh, K. Misawa, K.-I. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066 (2002).

91. 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).

92. J. Sukumaran, M. T. Holder, DendroPy: a Python library for phylogenetic computing. Bioinformatics 26, 1569–1571 (2010).

93. 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. Journal of Open Source Software 7, 3722 (2022).

94. 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).

95. M. Kolmogorov, J. Yuan, Y. Lin, P. A. Pevzner, Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. 37, 540–546 (2019).

96. 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).

97. 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).

98. 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).

99. A. Shumate, S. L. Salzberg, Liftoff: accurate mapping of gene annotations. Bioinformatics 37, 1639–1643 (2021).

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, 7 (2015).

101. Browning, S.R., Browning, B.L., 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).

102. C. L. Picard, M. Gehring, Proximal methylation features associated with nonrandom changes in gene body methylation. Genome Biol. 18, 1–18 (2017).

103. 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).

104. 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).

105. 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).

106. N. A. Ramawatar, B. Schwessinger, "DNA size selection (>3-4kb) and purification of DNA using an improved homemade SPRI beads solution. v1" (ZappyLab, Inc., 2018); https://doi.org/10.17504/protocols.io.n7hdhj6.

107. 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).

108. 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).

109. 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).

110. 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).

111. 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).

112. 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).

113. J. Denkena, F. Johannes, M. Colomé-Tatché, Region-level epimutation rates in Arabidopsis thaliana. Heredity 127, 190–202 (2021).

114. 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).

115. 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).

116. J. Liu, X. Zhong, Epiallelic variation of non-coding RNA genes and their phenotypic consequences. Nat. Commun. 15, 1375 (2024).

117. 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).

118. Z. Shahzad, J. D. Moore, J. Choi, D. Zilberman, Epigenetic inheritance mediates phenotypic diversity in natural populations, bioRxiv (2021)p. 2021.03.15.435374.

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 manuscript: PB, VC

DNA, RNA, and sRNA extractions of epiRILs as well as ONT-sequencing of 20 strains: EC

Analysis of WGBS-seq data, GWAS and GLM as well as de novo assembly of 20 genomes 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 and CX

Single-read DNA methylation analysis using ONT: MBa

Generation of *ddm1* and *ddm1* compound mutants and extraction of sRNAs: FKT 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.

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1 to S13

Tables S1 to S11

References (*83-118*)

Datasets S1 to S7