Transposable elements are prevalent vectors of transgenerational epigenetic inheritance in nature

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

[1] Institut de Biologie de l'École Normale Supérieure, Université Paris Sciences et Lettres (PSL), UMR 8197 CNRS, U1024 INSERM, 46 rue d'Ulm, 75005 Paris, France

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

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

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

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

[6] Present address: Institute of Plant Sciences of Paris-Saclay (IPS2), Université Paris-Saclay, Université Paris Cité, Université Evry Val d'Essonne, U1403 INRAE, UMR 9113 CNRS, 91190 Gif-sur-Yvette, France

[7] Present address: Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, 57 rue Cuvier, 75005 Paris, France

** correspondence to* pierre.baduel@ens.psl.eu; vincent.colot@ens.psl.eu

Abstract

Transgenerational epigenetic inheritance (TEI) mediated by transposable elements (TEs) is well established in plants. However, the molecular determinants and functional impact of this additional system of inheritance are poorly characterized. Through comprehensive analyses of severe DNA methylation loss over TEs in an experimental population and in >700 natural strains of *Arabidopsis thaliana*, we uncovered hundreds of TE epivariants in nature, most of which can be inherited independently of DNA sequence changes and thus are true epialleles. Furthermore, natural epivariants are enriched near genes and we uncovered a novel role for the Paf1 complex as a major facilitator of their prevalence. Conversely, related TE copies limit epiallelic variation, by producing the small RNAs that guide DNA methylation in *trans*. Crucially, natural epivariants mainly affect stress-responsive genes, making them unique targets for selection. Our findings therefore demonstrate that TEs are important mediators of transgenerational epigenetic variation in nature, with singular properties for plant adaptation.

Introduction

Transposable elements (TEs) are ubiquitous in eukaryotic genomes, contributing over 80 percent of nuclear DNA in some species[1]. Most TEs are ancestral copies that long lost their ability to transpose, hence their designation as "junk" DNA[2–4]. However, the rise of genomics has uncovered their key role as a source of biological innovations at the macroevolutionary scale[1,5–7]. In addition, the few mobile TEs contained within genomes can generate large-effect mutations (see[8–11] for examples) and within-species genomic studies have revealed that TE mobilization is widespread in natural settings[12,13]. Moreover, work in the model plant *Arabidopsis thaliana* has shown that in this species, mobile TEs that are the most active in nature insert preferentially near or within genes[14–16], often leading to complete knock outs in the latter case. Hence, novel TE insertions tend to be rapidly purged by natural selection in *A. thaliana*[14,15], which in turn suggests that the vast majority of ancestral TE insertions present across the genome have accumulated by genetic drift because they were neutral or only mildly deleterious at the outset.

In plants and mammals, DNA methylation plays important roles in limiting transposition, yet this chromatin modification targets ancestral TEs too[17–19]. Thus, despite being mere fossils, these sequences may therefore still exert an influence on the expression of the surrounding genes[20,21] and contribute to phenotypic diversity, simply as a result of variation in DNA methylation and independently of DNA sequence changes. Moreover, because this modification is not extensively reset across generations in plants[22–24], unlike in mammals[25,26], plants may be particularly prone to TE-mediated transgenerational epigenetic inheritance (TEI)[27,28]. Indeed, a number of "pure" TE-mediated epialleles[29,30] involving fossil TEs and with striking phenotypes have been reported in plants, including *Bad Karma* in oil palm, which blocks fruit development[31] and *fwa* in *A. thaliana,* which delays flowering[32–34]. By contrast, similar examples are few in mammals and all exhibit complex inheritance patterns, which questions their epiallelic nature[35].

While the existence of TE-mediated TEI in plants is undisputed[36], its modalities, extent, and phenotypic impact remain poorly understood[27,29]. Here we aimed to provide answers to these questions using *A. thaliana*, where mutants with severe loss of DNA methylation at TEs and other repeated sequences are viable and fertile[37–39], unlike in other plants with a much higher TE content[40,41].

66 In *A. thaliana*, cytosine DNA methylation is carried out by four main DNA methyltransferases
67 (MTases), MET1, CMT3, CMT2 and DRM2, which all target TEs and contribute to the stable
68 epigenetic inheritance of heavy DNA methylation over TEs[18,42–44]. MET1 is the ortholog of
69 mammalian CG-maintenance MTase Dnmt1, while CMT2 and CMT3 are plant-specific
70 chromodomain-containing DNA MTases that catalyze respectively CHH and CHG methylation
71 (where H=A, C or T), in feedforward loops with histone MTases that deposit the heterochromatic
72 mark H3K9me2. In contrast to these three DNA MTases, DRM2, the ortholog of the mammalian
73 *de novo* DNA MTase Dnmt3, is responsible for RNA-directed DNA methylation (RdDM) in all
74 cytosine contexts, a process guided by 24-nt small RNAs that mainly serves to perpetuate CHH
75 methylation at TEs, especially those not targeted by CMT2. Maintenance of high levels of DNA
76 methylation of TEs requires many other factors, among which the Snf2-like chromatin remodeler
77 DDM1 is pivotal[18,42], by allowing DNA MTases to access heterochromatin[45]. Indeed, *ddm1*
78 mutants exhibit severe hypomethylation almost exclusively over TE and other repeat
79 sequences[34,45,46]. Moreover, early work showed that *ddm1*-induced loss of DNA methylation can
80 be stably inherited even in the absence of the inducing mutation[37]. Capitalizing on this property,
81 which results from the disruption of the intricate interplay between DDM1 and histone variants[47],
82 a population of *ddm1*-derived epigenetic recombinant lines (so-called epiRILs) was generated
83 that led to the discovery of a large reservoir of potential epialleles affecting complex traits[48–50]
84 and of the key role of RdDM in limiting TEI at some DDM1 targets[51].

85 In the present study, we produced DNA methylation data at single cytosine resolution for 120
86 epiRILs to dissect out the molecular determinants of TE-mediated TEI. We then combined this
87 new information with genome sequence, DNA methylome, transcriptome, phenotypic and geo-
88 bioclimatic data available for hundreds of *A. thaliana* strains taken from across the world[52–55] to
89 characterize comprehensively this additional system of inheritance in nature. By uncovering the
90 genetic and genomic features that may limit or promote heritable DNA methylation variation at
91 TEs in this setting and assess its functional impact, we illuminate the molecular underpinnings of
92 TE-mediated TEI in plants as well as its biological and evolutionary significance.

93 # Results

94 ### Inheritance patterns of *ddm1*-induced hypomethylation differ markedly between TEs

95 To elucidate the molecular underpinnings of TE-mediated TEI in *A. thaliana*, we examined the
96 inheritance patterns of *ddm1*-induced severe loss of DNA methylation in the *ddm1*-derived
97 epiRIL population[48]. Briefly, this population was produced by crossing a *ddm1* parent with an
98 isogenic WT parent of the reference strain Col-0 and by propagating through single seed descent
99 over 500 backcross progeny without the *ddm1* mutant allele. Using WGBS-seq, we obtained
100 DNA methylome data at single nucleotide resolution for 120 epiRILs taken at generation F9 as
101 well as for the *ddm1* and WT parental lines. We identified in total 25,462 regions with marked
102 differences in DNA methylation (DMRs) between the two parental lines (Fig. S1-2, Dataset S1-
103 2; see Methods). Taking as genetic markers a subset of 1,092 regularly spaced DMRs that
104 segregate in a manner compatible with Mendelian inheritance (Dataset S3; see Methods), we
105 inferred for each epiRIL its epihaplotype along the five *A. thaliana* chromosomes (Fig. 1a, S2e,
106 S3; see Methods). We then selected a stringent set of 7,023 DMRs (Table S1) that are fully
107 contained within TE annotations (designated thereafter as TE-DMRs; see Methods) to avoid
108 ambiguities. None of the TE-DMRs belong to TE copies with transposition activity in the
109 epiRILs[16], thus limiting the risk of erroneous mapping of sequence reads.

110 For each TE-DMR, we measured its average CG methylation level in each of the parental lines
111 and epiRILs (Fig. S1c; see Methods) and used these values to define one of three possible DNA
112 methylation states in the epiRILs: WT-like, *ddm1*-like, or intermediate. TE-DMRs almost always
113 (97.6% of cases) exhibit WT-like methylation when harbored on WT-derived chromosome
114 intervals, with only few instances of intermediate or *ddm1*-like methylation instead (Fig. 1b).
115 Exceptions mainly involve TE-DMRs belonging to the two DNA transposon superfamilies
116 known to encode *trans* demethylation activities[56,57] (*VANDAL* or *MuDR*-like and *ATENSPM*,
117 Fig. S1d, S2e). Thus, at almost all TE-DMRs, the parental WT methylation state is stably
118 transmitted across generations in the epiRILs.

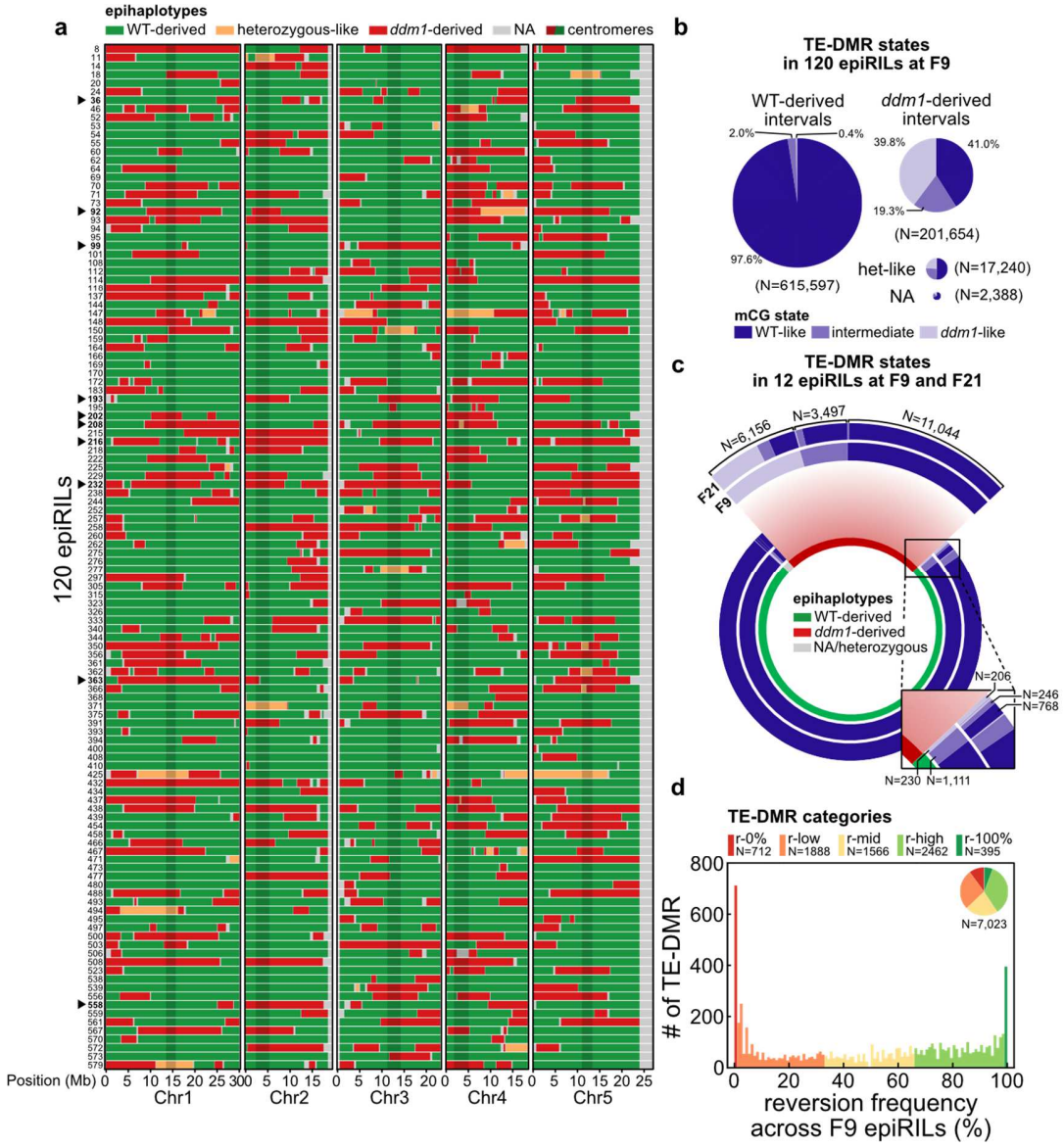


Figure 1. Inheritance patterns of *ddm1*-induced hypomethylation

(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.

In marked contrast, *ddm1*-induced hypomethylation is not systematically transmitted. Instead, the vast majority of *ddm1*-derived TE-DMRs (N=6,311 or 89.1%) exhibit at least one instance of

129 WT-like or intermediate DNA methylation, indicative respectively of full or partial reversion in
130 the epiRILs (Fig. 1b). To assess these inheritance patterns across a larger number of generations,
131 we further propagated a random subset of 12 epiRILs and obtained DNA methylome data at
132 generation F21. Once restored in F9, WT-like DNA methylation over *ddm1*-derived TE-DMRs is
133 almost always (99%) stably transmitted to F21 (Fig. 1c). In addition, intermediate DNA
134 methylation at F9 usually translates into WT-like levels by F21 (Fig. 1c), consistent with
135 previous work indicating that restoration of DNA methylation often occurs in a stepwise manner
136 across several generations[51]. By performing for one epiRIL long-read ONT sequencing and
137 single-molecule analysis of DNA methylation of the same DNA sample used for BS-Seq, which
138 was obtained from a pool of six seedlings, we also confirmed that intermediate DNA methylation
139 does not usually reflect the presence of a mixture of progeny with or without reversion, but
140 rather incomplete restoration of WT-like levels in individual progeny (Fig. S1f-g).

141 While over 90% of TE-DMRs show evidence of reversion in at least one epiRIL, only 395 have
142 reverted across all epiRILs, which indicates that most TE-DMRs can sustain TEI, albeit with
143 different degrees of penetrance. To investigate the significance of this observation, we
144 considered respectively cases of intermediate and WT-like methylation as half- and full-
145 reversion events and defined three broad categories of TE-DMRs with non-systematic reversion
146 (r-low, r-mid, and r-high) by generation F9 in addition to the two categories of TE-DMRs with
147 no (r-0%) or systematic reversion (r-100%; Fig. 1d, see Methods). Crucially, some r-0% TE-
148 DMRs had reverted by generation F21 (Fig. 1b-c, S1e), confirming that reversion dynamics
149 differ greatly between individual TE-DMRs and suggesting that at most TE-DMRs, TEI cannot
150 be sustained indefinitely in the absence of selection.

151 **RdDM-targeting in *trans* is the main determinant of reversion**

152 To elucidate precisely the molecular underpinnings of reversion, which requires RdDM[51], we
153 first measured in WT and *ddm1* the average CHH methylation (mCHH) level of TE-DMRs for
154 each of the five reversion categories. Although similar in WT across all categories, averages in
155 *ddm1* are much lower for the r-0%, r-low and r-mid categories and only mildly reduced for the r-
156 high and r-100% categories (Fig. 2a), consistent with the latter being enriched in RdDM targets
157 [58](Fig. S4a). Given the involvement of 23-24nt sRNAs in guiding RdDM, we sequenced sRNAs

extracted from the two parental lines, three *ddm1* compound mutant lines that affect to varying degrees the production of 23-24nt sRNAs and 10 epiRILs (see Methods). The abundance in the epiRILs of 23-24nt sRNAs that match *ddm1*-derived TE-DMRs correlates positively with reversion and in this case, levels tend to be even higher than in the WT parental line (Fig. 2b). Furthermore, reversion correlates most strongly with the abundance of 23-24nt sRNAs in *ddm1* (Fig. 2c), as a result of the drastic reduction of 23-24nt sRNAs matching TE-DMRs that belong to the r-0% or r-low categories in this mutant background (Fig. 2d, S4b-c). Correlations are also markedly higher for multiple- than single-mapping 23-24nt sRNAs (Fig. 2b,c) and related TE-DMRs often revert in concert (Fig. 2e). Moreover, a fraction of TE-DMRs (see Methods) that lose most or all of their single-mapping 23-24nt sRNAs in *ddm1* revert without regaining detectable levels of these (Fig. 2f, S4c-e; see example in Fig. 2g). Together, these findings reveal a predominant role for non-allelic interactions in driving reversion, consistent with results of numerous experiments with transgenes indicating that 23-24nt sRNAs can guide DNA methylation ectopically[59–63].

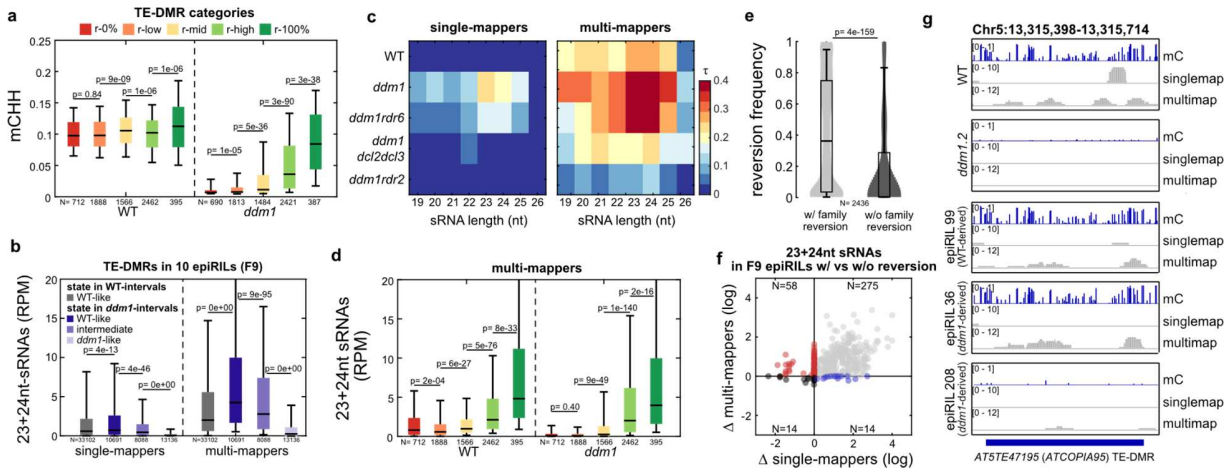


Figure 2. RdDM-targeting is the main determinant of reversion

(a) Levels of mCHH in WT and *ddm1* over the five TE-DMR reversion categories. (b) 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. (c) Correlation (Kendall τ) between single-mapping or multi-mapping sRNA levels (RPM) and reversion frequency in WT and different mutant backgrounds. (d) 23-24nt-long multi-mapping sRNA levels in WT and *ddm1* over the five TE-DMR reversion categories. (e) 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. (f) Changes in 23-24nt multi- vs single-mapping sRNAs associated with reversion, over TE-DMRs that lose most or all of their single-mapping 23-24nt sRNAs in

182 *ddm1*. Number of TE-DMRs in each quadrant is indicated. (g) DNA methylation levels and abundance of matching
183 23-24nt sRNAs over the *ATCOPIA95* TE-DMR (*AT5TE47195*) in WT, *ddm1*, epiRIL36 (where it was inherited
184 from *ddm1* and had reverted without regaining the single-mapping sRNAs lost in *ddm1),* epiRIL99 (where it was
185 inherited from WT), and epiRIL 208 (where it was inherited from *ddm1* and had not reverted).

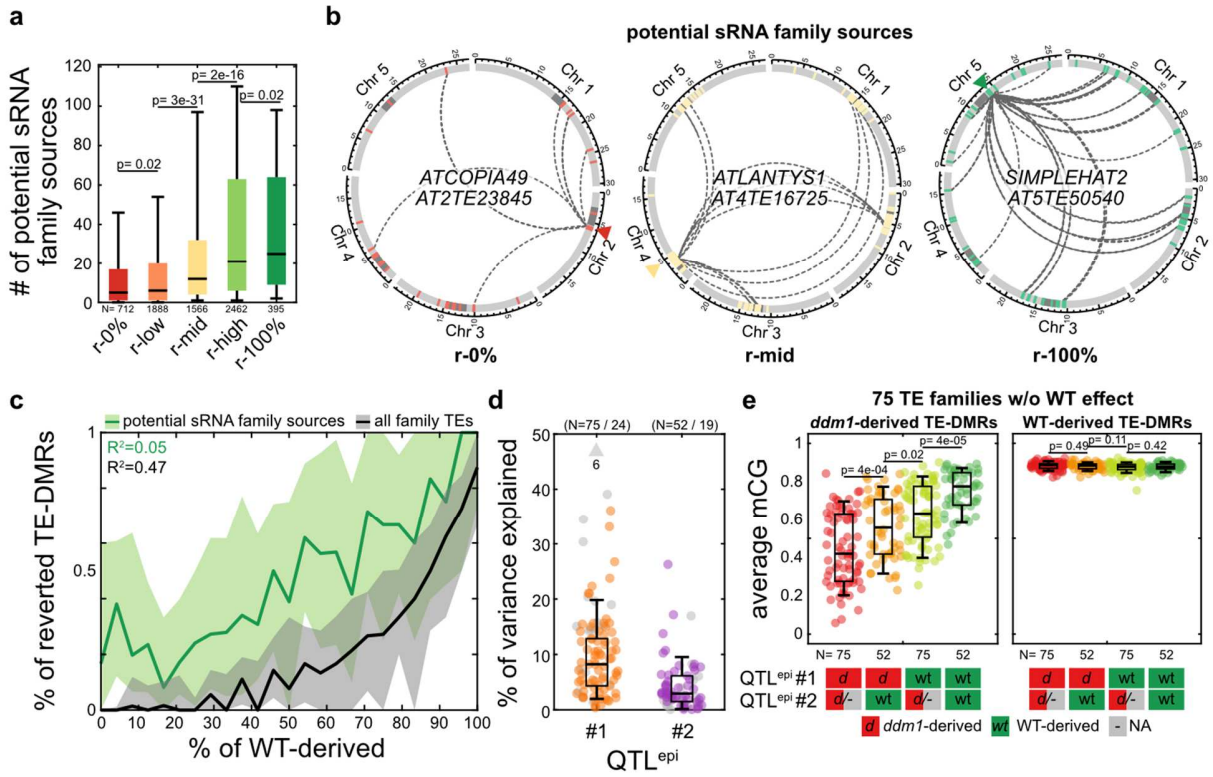
186 **WT-derived, methylated TEs are the main sources of *trans*-acting sRNAs driving reversion**

187 To identify the origin of the *trans*-acting 23-24nt sRNAs guiding reversion, we took into account
188 the fact that RdDM can be driven by sRNAs with imperfect homology[64] and considered all
189 multimapping 23-24nt sRNAs that can be aligned with up to one mismatch onto target TE-
190 DMRs. We then identified the potential sources of these sRNAs by looking for perfect matches
191 in the reference genome sequence (see Methods). As expected, most (92%) potential sources are
192 TEs that belong to the same families or superfamilies as the target TE-DMRs (Fig. S5a) and their
193 number, which increases with family size as well as the percentage of sequence identity with
194 target TE-DMRs, correlates positively with reversion (Fig. 3a-b, S5b-c). In addition, the
195 proportion of potential sources that are inherited from the WT parent also correlates positively
196 with reversion (Fig. 3c), consistent with the drastic loss of 23-24nt sRNAs in *ddm1* (Fig. 2b), and
197 this correlation is even stronger if all copies of a given TE family are considered (Fig. 3c), which
198 suggests the presence of additional sources missed by our homology search.

199 To determine if some WT-derived TEs are predominant sources of *trans*-acting 23-24nt sRNAs,
200 we took an epigenetic quantitative trait locus (QTL^{epi}) mapping approach. For each of the 105
201 TE families with at least 10 non-systematically reverting TE-DMRs, we used the average CG
202 methylation level of TE-DMRs inherited from the *ddm1* parent as a quantitative measure of
203 reversion within an epiRIL and asked whether variation between epiRILs in family-wide
204 reversion correlates with the parental origin of specific chromosomal segments (see Methods).
205 For 99 TE families, interval mapping revealed the presence of one or two major QTL^{epi}, each
206 explaining on average 13% of variation (Fig. 3d). These tend to be pericentromeric (Fig. S5d),
207 consistent with their TE-richness. In addition, they are significantly overrepresented in the
208 pericentromeres of chromosomes 1 and 2, which suggests the presence of more or more efficient
209 sources of 23-24nt sRNAs in these pericentromeres compared to the other three.

210 As expected, the WT-derived epigenotype at each peak QTL^{epi} marker is associated with higher
211 average CG methylation at TE-DMRs inherited from the *ddm1* parent and effects are additive

212 (Fig. 3e, S5e-f). Moreover, most (>80%) QTL^{epi} encompass one or several of the TEs we
213 identified as potential sources of *trans*-acting sRNAs (Fig. S5d), suggesting causality. As
214 expected also, the QTL^{epi} do not impact methylation of the TE-DMRs inherited from the WT
215 parental line (Fig. 3e), except of course in the case of TE families with *trans*-demethylation
216 activity. Indeed, 16 of the 24 TE families for which the *ddm1* epigenotype at the QTL^{epi} is
217 associated with lower methylation of WT-derived TE-DMRs (Fig. S5d-e, see Methods) belong to
218 the *VANDAL* or *ATENSPM* superfamilies, for which *trans*-demethylating activities have been
219 documented [56,57] (Fig. S5d). Hence, the QTL^{epi} identified for these superfamilies have a dual
220 effect, favoring either reversion when WT-derived or *trans*-demethylation when *ddm1*-derived.



221

222 **Figure 3. Reversion is driven mainly by *trans*-acting sRNAs**
223 (a) Number of potential 23-24nt sRNA sources belonging to the same TE family as the focal TE-DMRs, for each of
224 the five reversion categories. (b) Genomic localization of three representative r-0%, r-mid, and r-100% TE-DMRs
225 (indicated by triangles) and of related TE copies, including those that are potential sources of 23-24nt sRNAs
226 (indicated by arcs). (c) Percentage of reverted TE-DMRs per TE-family relative to the proportion of potential 23-
227 24nt sRNA sources belonging to the same TE family and that are WT-derived (in green) or of all WT-derived copies
228 of that TE family (in black). Coefficient of correlation (R^2) is indicated. (d) Percentage of variance explained
229 iteratively by the first and the second QTL^{epi} in the average mCG levels of *ddm1*-derived, non-systematically

230 reverting TE-DMRs within a given TE family. Values for the 24 TE families for which the QTL^{epi} affects the
231 methylation status of WT-derived TE-DMRs are shown in gray, with outliers and their number indicated with a
232 triangle. (e) Average mCG levels of *ddm1*- or WT-derived TE-DMRs per TE family in relation to the epihaplotype
233 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
234 not affect the methylation status of WT-derived TE-DMRs.

235 An important prediction from the above observations is that *ddm1*-induced hypomethylation at
236 non-systematically reverting TE-DMRs should be stably inherited in epiRILs where the main
237 producers of *trans*-acting 23-24nt sRNAs derive from the *ddm1* rather than the WT parental line.
238 To confirm this prediction, we grew >75 siblings of two randomly chosen F9 epiRILs (#92 and
239 #216) that harbor non-systematically reverting TE-DMRs in the *ddm1*-derived hypomethylated
240 state and measured DNA methylation of these in each sibling. We analyzed five TE-DMRs in
241 this manner, belonging to four distinct TE families and covering the range of population-wide
242 reversion frequencies (one r-low, two r-mid, and two r-high TE-DMRs). Moreover, three TE-
243 DMRs are members of TE families with at least one QTL^{epi} for family-wide reversion and as
244 expected, the top QTL^{epi} is inherited from *ddm1* in the epiRILs where they are hypomethylated
245 (Fig. S5g). Remarkably, even for the two r-high TE-DMRs, reversion was either extremely rare
246 or not detected at all (Fig. S5f). Thus, the epigenotype (WT- or *ddm1*-like) of trans-acting 23-
247 24nt sRNA sources is a key determinant of the transgenerational stability of *ddm1*-induced
248 hypomethylation of target TE-DMRs.

249 **Antagonist effects of RdDM and transcriptional activity modulate the inheritance of *ddm1*-**
250 **induced TE hypomethylation**

251 Although reversion is usually concerted within an epiRIL among TE-DMRs that belong to the
252 same TE family (Fig. 2d), we observed 369 TE-DMRs that remained hypomethylated even in
253 epiRILs where over >75% of related TE-DMRs have reverted (Fig. S6a, see Methods). Such
254 "stray" TE-DMRs tend to have higher levels of sequence divergence with their closest relative
255 and fewer close relatives altogether (Fig. 4a-b), in agreement with the known reduction of RdDM
256 efficiency when source and target exhibit sequence polymorphisms[65]. Furthermore, stray TE-
257 DMRs are enriched at or near RNA Pol II transcription units that are activated in *ddm1*
258 (EPigenetically Induced Consensus tAg clusTers or EPICATs[66]; Fig. S6b), in keeping with the
259 known hampering effect of Pol II transcription on RdDM[59,67,68]. The two closely related TE-

DMRs of the *ATGP2N* family located on chromosomes 1 and 4 provide a striking illustration of the negative impact of these EPICATs on reversion. Specifically, despite an abundance of multi-mapping 23-24nt sRNAs matching the two TE-DMRs in epiRIL #232, where both derive from *ddm1*, the copy present on chromosome 1 reverted whereas that on chromosome 4, which contains a strong and long EPICAT, did not (Fig. 4c). This pattern is in fact quite general, as reversion of TE-DMRs associated with EPICATs is less frequent than that of other TE-DMRs of the same family (Fig. S6c) and may require higher levels of matching multi-mapping 23-24nt sRNAs (Fig. 4d). Thus, inheritance of *ddm1*-induced hypomethylation is favored when there is a limited supply of *trans*-acting 23-24nt sRNAs and/or strong Pol II transcriptional activity at or near the TE-DMR.

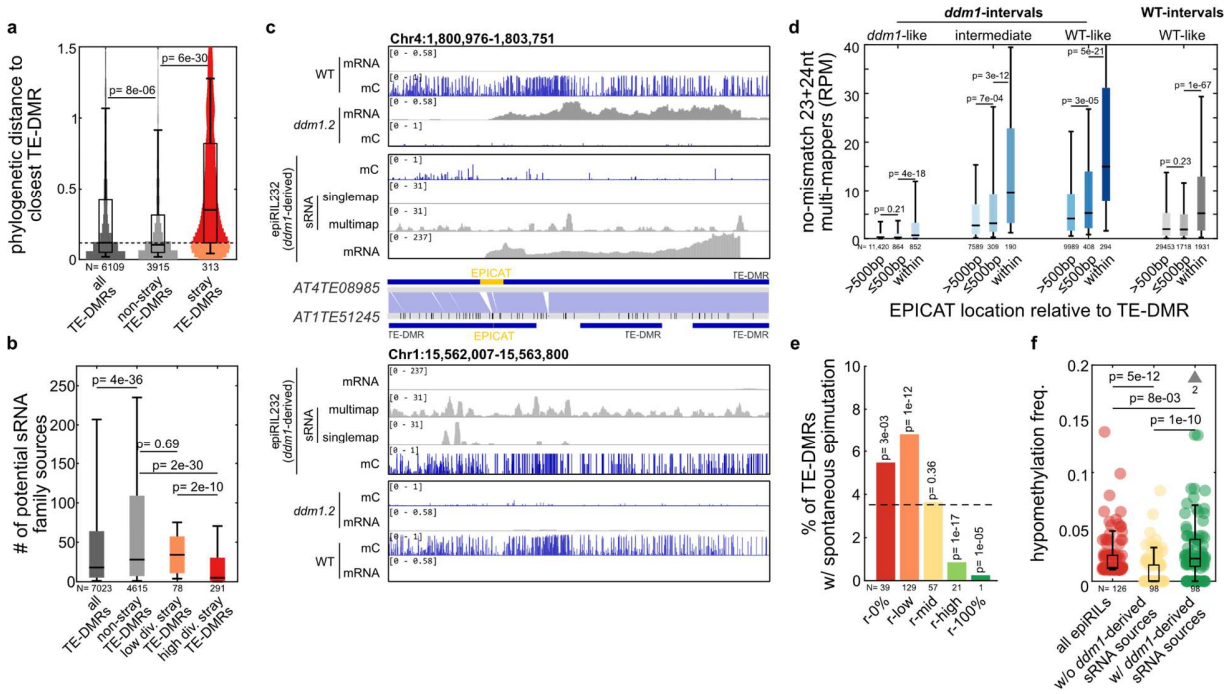


Figure 4. Inheritance and spontaneous occurrence of TE hypomethylation is modulated by two main antagonistic features

(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 [66], 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

278 high levels of multi-mapping sRNAs and the presence in the same epiRIL of a closely related TE-DMR
279 (*AT1TTE51245*) that has reverted. (d) Levels of multi-mapping 23-24nt sRNAs over TE-DMRs in 10 sequenced
280 epiRILs (F9), in relation to their methylation state when they are WT- or *ddm1*-derived and as a function of their
281 distance relative to the nearest EPICAT. (e) Percentage of TE-DMRs with spontaneous *ddm1*-like hypomethylation
282 in the 120 epiRILs across the five reversion categories. The p-values indicate the cumulative binomial probability of
283 observing similar or more extreme representation from random sampling (dashed line). (f) Frequency of
284 spontaneous *ddm1*-like hypomethylation among all epiRILs, only those where no related TE-DMR is *ddm1*-derived,
285 or only those where at least one related TE-DMR is *ddm1*-derived. The presence of outliers and their numbers are
286 indicated by a gray triangle.

287 **Weak RdDM targeting facilitates spontaneous TE epivariation**

288 In addition to the 1426 TE-DMRs that belong to TE families with known *trans*-demethylation
289 activity, another 126 TE-DMRs exhibit *ddm1*-like methylation in at least one of the F9 epiRILs
290 in which they are inherited from the WT (Fig. 1b, see Methods), suggestive of spontaneous
291 *ddm1*-like hypomethylation. To investigate this observation further, we exploited additional
292 DNA methylation and epihaplotypic data that were independently obtained at generation F9 for a
293 set of 169 epiRILs that minimally overlaps with ours[69]. After discarding potential *trans*-
294 demethylation events as well as possible epihaplotyping errors (see Methods, Fig. S6e), we
295 identified another 161 TE-DMRs with spontaneous *ddm1*-like epivariation within WT-derived
296 intervals. Remarkably, the 287 TE-DMRs overwhelmingly (68%) belong to the r-0% and r-low
297 categories (Fig. 4e) and they epivary spontaneously far less (3.8-fold less) when the potential
298 sources of *trans*-acting 23-24nt sRNAs are all WT-derived (Fig. 4f). In contrast, we did not
299 detect any preferential association of these TE-DMRs with EPICATs (Fig. S6d). Thus, we
300 conclude that whereas RdDM limits the spontaneous occurrence as well as the inheritance of
301 severe loss of DNA methylation at TE-DMRs, EPICATs only favor the latter.

302 **TE epivariants are common in nature and most are *bona fide* epialleles**

303 Having elucidated the molecular underpinnings of the inheritance patterns of *ddm1*-induced
304 epivariation in the epiRILs as well as of the spontaneous occurrence of similar epivariants in this
305 setting, we asked whether any of the 7,023 experimentally defined TE-DMRs exhibit *ddm1*-like
306 hypomethylation in nature. To this end, we took advantage of the BS-seq data available from leaf
307 tissues for 720 natural strains[53] and calculated for each TE-DMR its average CG methylation

308 level in each of the strains where the corresponding reference TE is present[15,70] (see Methods).
309 The number of such strains is typically very high as most TE-DMRs belong to TEs that are either
310 fixed or widely shared at the species level (Fig. S7a-b), consistent with most TEs in any given
311 genome being molecular fossils[14,15,71]. However, because we applied stringent filters to ensure
312 that average CG methylation is calculated over a sufficient number of informative CG sites
313 across the entire length of the TE-DMR (see Methods), we obtained reliable estimates of DNA
314 methylation for only around half of the TE-DMRs in any given strain (Fig. S7c, Dataset S5).

315 Most (85%) of TE-DMRs that passed our stringent filtering are always heavily methylated in
316 nature (Fig. 5a), consistent with dense methylation being the default state for TEs in *A.*
317 *thaliana*[53,72]. Nevertheless, we observed 1,068 TE-DMRs, hereafter designated as naturally
318 epivariable TE-DMRs, that exhibit *ddm1*-like CG hypomethylation in at least one strain (Table
319 S4). Crucially, hypomethylation systematically extends to CHG and CHH sites, as in *ddm1* (Fig.
320 S7d,e), although visual inspection indicates that the exact boundaries may differ between strains
321 (e.g. Fig. S7f), presumably in part because of DNA sequence polymorphisms within or near the
322 TE-DMRs.

323 Given that structural variants are difficult to detect using short read-sequencing and may
324 generate false calls of *ddm1*-like hypomethylation, we produced *de novo* genome assemblies
325 using long-read ONT sequencing for 20 genetically diverse strains (Table S5, see Methods). We
326 then determined the location and integrity of the TE-DMRs which we deemed present in the
327 genome of these strains based on short-read sequencing (see Methods). Using this approach, we
328 determined that on average, 18% of TE-DMRs in any strain are either truncated or missing from
329 the reference location. However, this percentage drops to 13% for the set of TE-DMRs with
330 natural epivariation (Dataset S6) and, whether the BS-seq data are mapped onto the *de novo*
331 assemblies or TAIR10 (R^2=0.86), methylation calls for the 82% of TE-DMRs with no structural
332 rearrangement are highly congruent (>95%, Fig. 5b). Thus, thanks to stringent filtering, our
333 detection of natural epivariation is particularly robust to the issue of mismapping caused by
334 related TEs[73].

335 Natural epivariation may nonetheless be influenced or dictated by other types of DNA sequence
336 variation[30]. To assess the possible contribution of single nucleotide polymorphisms (SNPs) in *cis*
337 (within 20kb), we carried out genome-wide association studies (GWAS) of natural epivariation

338 at each of the 490 TE-DMRs with sufficient epivariant counts (see Methods). Despite a few
339 significant associations (Fig. S9a-b), none were systematic (Fig. S9c), thus ruling out a strict
340 dependence of natural epivariation on the presence of sequence variants in *cis*.

341 We next leveraged the population of recombinant inbred lines (RILs) previously built between
342 Cvi-0 and Col-0 [74] to test directly the inheritance of natural epivariants and the potential
343 involvement of *cis* and *trans* sequence polymorphisms. Cvi-0 harbors the second largest number
344 of epivariants (113, Fig. S7c), all but three of which are also found in at least one other natural
345 strain (Table S6). These features therefore make the Cvi-0 x Col-0 RIL population particularly
346 informative. In a first step, we analyzed the segregation of 34 epivariable TE-DMRs distributed
347 across the genome and belonging to distinct TE superfamilies and reversion categories (Fig. 5c,
348 see Methods) in 36 RILs that together recapitulate most of the genetic differences between the
349 two parental lines[74]. Parental DNA methylation states systematically segregate together with
350 flanking sequence markers, except in one case (Fig. 5c). This exception is readily explained by
351 the absence of the TE-DMR in Cvi-0 and the presence of a closely related, unmethylated TE
352 copy elsewhere (Fig. S9d). Crucially, we also observed rare reversion for 19 TE-DMRs (Fig. 5c).
353 Together, these findings rule out any obligatory role of DNA sequence polymorphisms in *cis*
354 (including within the TE-DMRs) or *trans* in the inheritance of the 34 natural epivariants we
355 tested. Using whole-genome BS-seq data produced for a partially non-overlapping set of 10 Cvi-
356 0 x Col-0 RILs[75], we obtained for the vast majority (87%) of TE-DMRs similar evidence of
357 segregation of parental methylation differences independently of *cis* or *trans* sequence
358 polymorphisms (Fig. S9e). We therefore conclude that most natural epivariants at TE-DMRs are
359 *bona fide* epialleles.

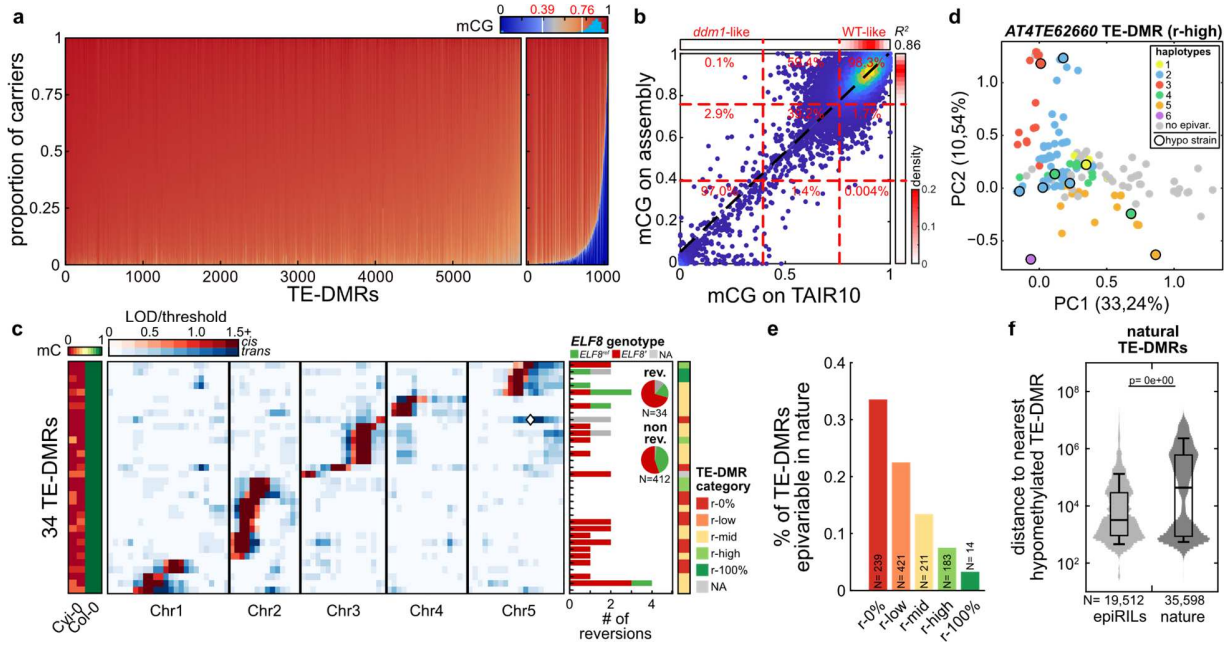


Figure 5. TE epivariants are recurrent in nature

(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) 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. (c) 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 non-reference TE insertion of *AT3TE90350 (ATCOPIA23)* present in Cvi-0 [15] 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. Results are summarized in pie charts. (d) PCA of SNPs within 1kb windows of a representative r-high TE-DMR. Strains are colored by haplotype if at least one carrier strain has an epivariant for the TE-DMR considered and are in gray otherwise. (e) Proportion of TE-DMRs with natural epivariation across the five reversion categories. (f) Distance to nearest epivariant in the epiRILs and in natural strains.

Natural epivariation results from recurrent loss of DNA methylation at TEs

Naturally epivariable TE-DMRs belong to TEs with similarly high population frequencies as the full set (Fig. S7b), yet natural epivariants tend to have low population frequencies (Fig. 5a; Fig. S8a). These observations indicate that most natural epivariants affect ancient TEs and reflect rare

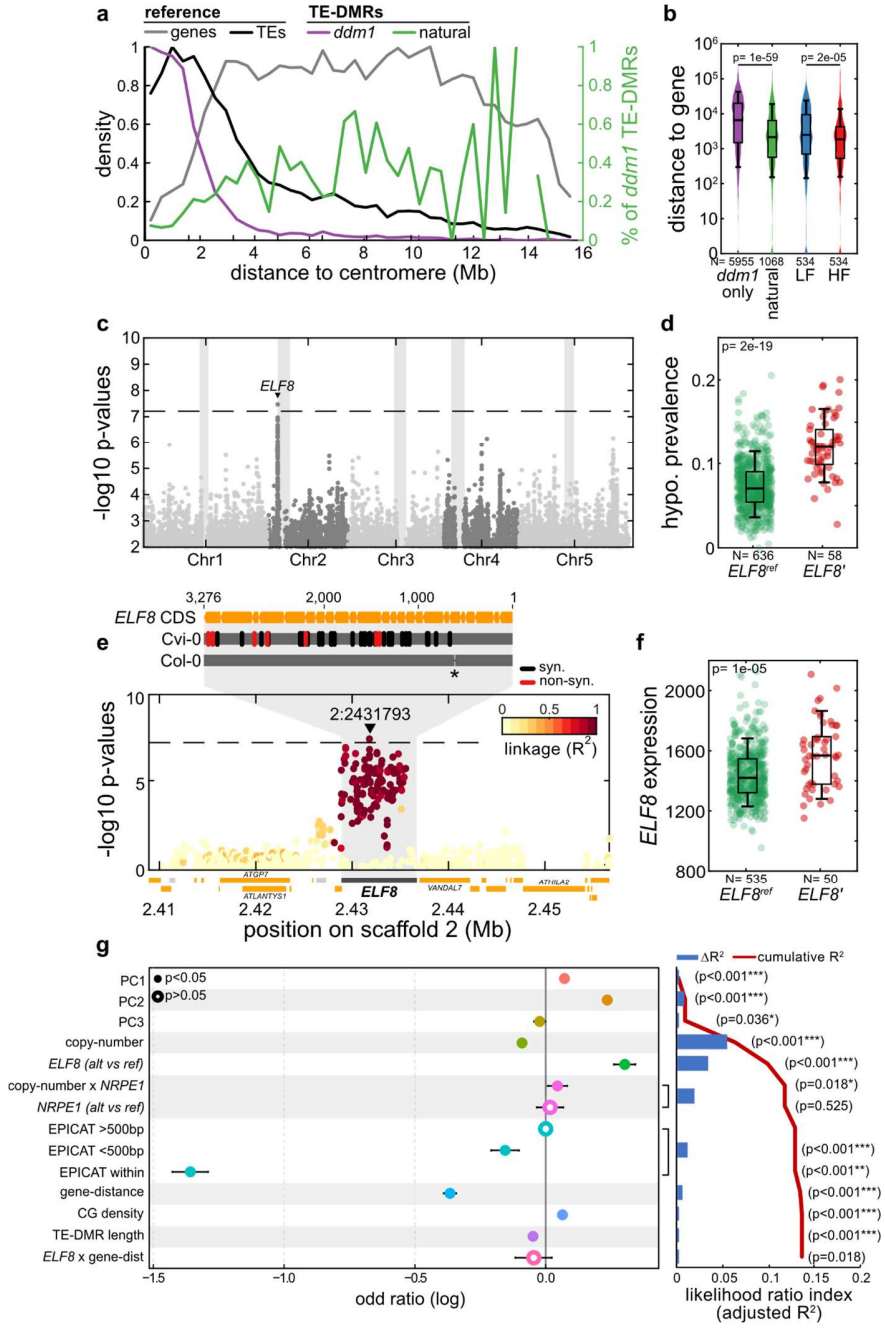
381 events of severe DNA methylation loss over these. Moreover, intermediate DNA methylation
382 levels are very rarely observed at TE-DMRs (Fig. 5a), suggesting that DNA methylation loss is
383 rapid rather than progressive across generations.

384 Natural epivariants found over the same TE-DMR in two or more strains could reflect identity by
385 descent or recurrent occurrence. To distinguish between these possibilities, we considered the
386 780 TE-DMRs with several instances of epivariation in nature and assessed the genetic
387 relatedness (kinship) of the corresponding strains (see Methods). While almost all (96%) TE-
388 DMRs show epivariation in at least two highly divergent strains (genome-wide kinship <0.25),
389 only half show it also in at least two closely related strains (genome-wide kinship >0.75; Fig.
390 S8b). Furthermore, local haplotyping of each of the corresponding TE-DMR-containing regions
391 (see Methods) indicates that the majority (66%) of epivariants are carried by at least two distinct
392 alleles (Fig. 5d, S8c-d). Thus, the patterns of epivariation observed in nature are shaped in large
393 part by recurrent occurrences.

394 Natural epivariation affects predominantly TE-DMRs that belong to the r-0% or r-low categories
395 (Fig. 5d), which tend to exhibit the highest population frequencies as well as (Fig. S8e). These
396 observations therefore indicate that the molecular determinants in the epiRILs (Fig. 1-4) of the
397 inheritance of *ddm1*-induced hypomethylation or of the spontaneous occurrence of similar
398 hypomethylation are also likely at play in nature. However, natural epivariants are typically few
399 and scattered in any given genome (49 on average, Fig. 5b, S7c) and tend to be uncorrelated,
400 except when close (<10kb) to each other (Fig. S8f-g),which is in marked contrast to their
401 distribution in the epiRILs. Thus, although natural epivariants are indistinguishable molecularly
402 from their *ddm1*-induced counterparts, they are unlikely to originate as a result of a *ddm1*-like
403 mutation.

404 **Paf-1 activity favors natural epivariation near genes**

405 Naturally epivariable TE-DMRs tend to be further away from the pericentromeres and closer to
406 genes than the whole set of *ddm1*-induced TE-DMRs (Fig. 6a) and their population frequencies
407 correlate positively with gene proximity (Fig. 6b). These observations suggest that the
408 occurrence or stable inheritance of natural epivariation at TE-DMRs is facilitated near genes or
409 else that it is selected for in this case.



410

411 **Figure 6. Natural epivariants are enriched near genes**

412 (a) Metaplot of the density of TEs, genes, *ddm1*-induced TE-DMRs in relation to their distance from centromeres (left
413 y-axis). The proportion of *ddm1*-induced TE-DMRs that are also epivariable in nature is represented in green (right
414 y-axis). (b) Distance to the nearest gene (+1bp for log representation) of all *ddm1*-induced TE-DMRs, naturally
415 epivariable TE-DMRs and those for which epivariation segregate at low or high frequency (LF, first two quartiles;
416 HF, last two quartiles). (c) Manhattan-plot of p-values of GWAS of prevalence of natural epivariation by genome. (d)
417 Prevalence of natural epivariation among strains carrying the reference *ELF8*^{ref} or non-reference *ELF8'* alleles. (e)

418 Close-up Manhattan-plot of GWAS p-values around *ELF8,* colored by linkage with the leading SNP (2:2431793).
419 Coding-sequence polymorphisms in the Cvi-0 allele of *ELF8.* Asterisk marks a single nucleotide deletion in the Col-
420 0 TAIR10 sequence that is not detected in re-sequenced Col-0 *ELF8* cDNA. (f) *ELF8* expression in strains carrying
421 the $ELF8^{ref}$ or *ELF8'* alleles. (g) Marginal effects and likelihood ratio index (McFadden adjusted R^2) of each
422 explanatory variable in GLM of natural epivariation prevalence ranked by R^2.

423 To look for genetic modifiers that may contribute to these distinctive properties of natural
424 epivariation, we performed a GWAS on their proportion relative to the total number of naturally
425 epivariable TE-DMRs present within a strain. This approach revealed one significant peak of
426 association (Fig. 6c-d), which encompasses *EARLY FLOWERING 8* (*ELF8*) and no other gene
427 (Fig. 6e). *ELF8* encodes a component of the highly conserved RNA polymerase-associated factor
428 1 (Paf-1) complex, which is involved in Pol II transcription[76], including in plants via
429 phosphorylation of Pol II[77] and deposition of the histone marks H3K4me2/me3[78]. The non-
430 reference *ELF8* allele (hereafter designated *ELF8'*) appears fully functional (Fig. 6e, S10a-f; see
431 Methods) but it is expressed at higher levels than *ELF8* (Fig. 6f), presumably causing its
432 association with a higher number of hypomethylated epialleles per strain (Fig. 6d) . Indeed, this
433 association is most pronounced for TE-DMRs located within 2kb of genes (Fig. S10g), a
434 configuration where RdDM targeting is strongly hampered by H3K4me2/me3-marking of
435 genes[79]. Together, these observations therefore suggest that ELF8 activity either promote or
436 stabilize epivariation at TE-DMRs near genes by antagonizing RdDM. However, examination of
437 the Cvi-0 x Col-0 RILs (see Methods), which segregate the *ELF8'* allele, effectively rules out a
438 major role of ELF8 activity on its own in promoting the occurrence of epivariants or in
439 stabilizing their inheritance (Fig. 5c, S10h-i). Thus, while allelic variation at *ELF8* is an
440 important determinant of the prevalence of epivariants between strains, additional factors, either
441 genetic or environmental, likely contribute alongside *ELF8* in nature.

442 **RdDM targeting is the main factor limiting epivariation in nature**

443 The epiRILs revealed that the number of methylated TE copies present in the genome is a major
444 factor limiting TEI as well as the occurrence of spontaneous epivariation (Figure 1-4). Although
445 most TEs are methylated in nature, their number can vary substantially between genomes[14],
446 which prompted us to ask if this variation may explain differences in the prevalence of
447 epivariants between strains. Specifically, we used a generalized linear model (GLM) and

448 performed a stepwise selection of ten parameters, including allelic variation at *ELF8* and
449 distance to genes, on the basis of their added explanatory power (see Methods). Results of the
450 GLM indicate that genetic or genomic factors explain almost 15% of variance in the proportion
451 of epivariants among the naturally epivariable TE-DMRs present in a strain (Fig. 6g), with allelic
452 variation at *ELF8* and TE copy-number and being the main parameters (respectively 3.5% and
453 5.5% of variance explained; Fig. 6g). Moreover, the negative correlation observed between the
454 latter parameter and epivariation prevalence is not as strong in strains that carry a weak allele of
455 the RdDM gene *NRPE1*[15,80], consistent with our finding in the epiRILs that RdDM strength,
456 which increases with TE copy-number[81,82], limits both the inheritance and the spontaneous
457 occurrence of hypomethylation at TE-DMRs (Fig. 3-4, S5-6). The stepwise GLM approach
458 revealed also that TE-DMRs containing or located near EPICATs (see above) are less frequently
459 present in the hypomethylated state (Fig. 6f), suggesting that the proportion of epivariants within
460 a strain is determined primarily by their rate of occurrence rather than their inheritance
461 properties. Finally, the GLM confirmed that natural epivariation is more frequent closer rather
462 than further away from genes, and that this effect tends to be stronger in strains carrying the
463 *ELF8'* rather than the *ELF8*^{ref} allele, in line with our GWAS results (Fig. S10g). Together, these
464 findings highlight the pivotal and antagonistic roles of RdDM and Paf1 activity in shaping the
465 prevalence of natural epivariants.

466 **Natural epivariation impacts gene expression and is associated with specific habitats**

467 To assess the impact of natural epivariants on phenotypic diversity, we first used the RNA-seq
468 data obtained for 728 natural strains from leaf tissue of plants grown under standard conditions[53].
469 In line with the limited effects of *ddm1*-induced hypomethylation on gene expression overall[34,45],
470 most (>75%) natural epivariants are associated with minor (< 2-fold) expression changes of the
471 two nearest genes (Fig. S11a). However, 25% of epivariants are associated with major effects on
472 neighboring gene expression and typically with higher rather than lower expression, consistent
473 with the usually repressive impact of dense DNA methylation and suggesting causality. Indeed,
474 this bias is particularly strong for TE-DMRs located immediately upstream (≤500bp) or that
475 overlap with the coding region of genes (Fig. S11a-b) and causality was previously demonstrated
476 for the two genes *QQS* (Fig. S11c-d)[83] and *RPP7* (*AT1G58602;* Fig. S11e)[84].

477 To investigate further the impact of natural epivariants on gene expression, we produced RNA-
478 seq data for the same 10 F9 epiRILs that were used for the sRNA studies (Fig. 2e). For the 99
479 TE-DMRs that are differentially methylated among these epiRILs and for which meaningful
480 comparisons can be made (see Methods), we found that gene expression differences measured in
481 the epiRILs explain 40% of those observed in nature (Fig. 7a, Table S9). In addition, we
482 observed consistency in direction as well as magnitude in the majority of cases (59%) and almost
483 systematically when TE-DMRs are located upstream (≤500bp) or within genes (Fig. S11f).
484 Crucially, we also found that reversion to WT DNA methylation in the epiRILs is nearly always
485 (85%) accompanied by a partial or complete restoration of WT gene expression (Fig. S11g).
486 Thus, we conclude that most naturally epivariable TE-DMRs associated with gene expression
487 differences in cis are causal.

488 Causal TE-DMRs tend to be near or to overlap with genes of unknown function located in TE-
489 rich regions and that may be of recent origin, as was previously shown for three genes[83,85,86].
490 Therefore, whether epivariation at such TE-DMRs has any phenotypic impact is unclear, except
491 perhaps in the case of the *de novo* originated gene *QQS*[83], given its reported involvement in
492 starch metabolism[87]. Nonetheless, six genes have well-established functions (Table S9), among
493 which *AT2G05635 (RAD3*), which encodes a DEAD helicase, stands out, as it has been
494 implicated in natural variation for root length as well as tolerance to acid mine drainage[88].
495 Moreover, *RAD3* lies within the major QTL^{epi} interval identified for primary root length in the
496 epiRILs[50] and its expression is strongly upregulated when the three TEs positioned downstream
497 of it, which include two TE-DMRs, are all hypomethylated (Fig. S12a-c). Together, these
498 observations suggest that natural TE-mediated epiallic variation at *RAD3* impacts phenotypes in
499 an ecologically relevant manner.

500 To evaluate the functional impact of natural epivariation more directly, we searched for possible
501 correlations with a large set of environment variables. Specifically, we built for each of the 490
502 TE-DMRs with sufficient epivariant counts a GLM that accounts for population structure, the
503 potential *cis* modifiers we identified by GWAS (Fig. S9a-b) as well as the two main *trans*
504 modifiers of the prevalence of epivariants within strains, namely *ELF8* and TE copy-number. We
505 then tested whether any of the 196 AraClim variables compiled at the collection site of each
506 natural strain[55] bring additional explanatory power when included in the GLM (see Methods).

507 After correcting for multiple testing, we found 151 TE-DMRs where natural epivariation
508 correlates with at least one AraClim variable (Table S10). These "*araclim*" TE-DMRs comprise
509 the two located downstream of the *RAD3* gene, for which epivariation is enriched in habitats
510 with lower NO_2 emission levels in summer (Fig. 7b, S12d). As NO_2 emissions correlate
511 positively with levels of soil acidification[89], this last observation suggests that epivariation at
512 *RAD3* is maladaptive in native habitats characterized by acidic soils. More generally, the top
513 environmental variable explains on average 9% of variance at *araclim* TE-DMRs, a percentage
514 similar to that explained by *trans* genetic modifiers (13%; Fig. S12e). Thus, we conclude that
515 natural epivariation is likely of broad ecological relevance.

516 To reinforce this conclusion, we asked whether the environmental associations we uncovered are
517 correlated with natural differences in quantitative traits. Specifically, we considered the 417
518 quantitative phenotypic variables compiled in the AraPheno database
519 (https://arapheno.1001genomes.org/) and built for each of these a GLM with population structure
520 and up to 5 of the top significant SNPs identified by GWAS for each phenotype (AraGWAS)[54]
521 as the first six predictors. Given that none of these SNPs corresponds to a *cis*-modifier of natural
522 epivariation (Fig. S9a-c), we then tested as a seventh predictor each of the 490 naturally
523 epivariable TE-DMRs used for the AraClim analysis (see Methods). Using this approach, we
524 found that 89 (18%) TE-DMRs explain a substantial proportion of variance (8% on average,
525 Table S11) for at least one AraPheno trait, which is of the same order of magnitude as that
526 explained by top SNPs (~25% on average, Fig. 7c). Remarkably, more than half of these
527 "*arapheno*" TE-DMRs correspond also to *araclim* TE-DMRs (Fig. S12f), a significant
528 enrichment compared to the random expectation ($p<1e-6$) and a strong indication that natural
529 epivariation at these TE-DMRs contributes to local adaptation.

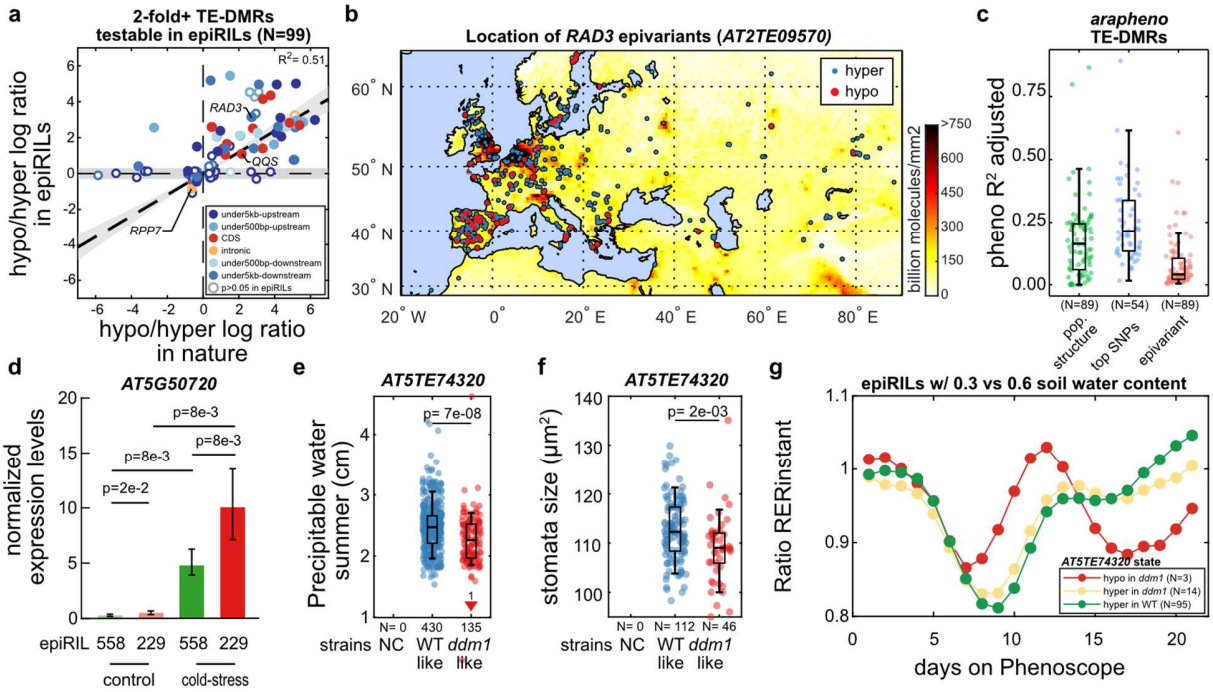


Figure 7. Natural epivariation has functional consequences

(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 neo.gsfc.nasa.gov) (c) Percentage of variance explained (MacFadden R^2 adjusted) by epivariation at 89 *arapheno* TE-DMRs in the GLM of their top AraPheno phenotype in comparison to that explained by population structure or by each of the top SNPs identified for each phenotype in AraGWAS. (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) Comparison of precipitable water in summer 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. (f) Comparison of stomata size of the same set of strains as in (i). (g) Comparison of the difference in daily growth rate (Delta RER) 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 can act as a specific enhancer of stress responses

Genes with naturally epivariable TE-DMRs are enriched in functions associated with responses to biotic or abiotic stress (Fig. S12g) and *araclim* as well as *arapheno* TE-DMRs tend to be the

549 closest to genes (Fig. S12h). However, based on the RNA-seq data of the epiRILs and natural
550 strains, which were all obtained from plants grown under standard conditions, only three of the
551 25 testable *arapheno/araclim* TE-DMRs have consistent and significant effects on neighboring
552 gene expression (Fig. S12f). This prompted us to test if some *arapheno/araclim* TE-DMRs could
553 exert their effect mainly or only in response to stress. To this end, we chose to investigate the
554 cold- and drought-stress responsive gene *HVA22E* (*AT5G50720*), which is located ~900bp
555 upstream of a r-mid category TE-DMR (Fig. S13a). We grew under standard conditions and cold
556 stress two of the four epiRILs that still harbor the TE-DMR in the *ddm1*-derived hypomethylated
557 state as well as two epiRILs that inherited it instead from the WT parental line (Fig. S13b) and
558 performed RT-qPCR on RNA extracted in each case (see Methods). While basal *HVA22E*
559 expression was similarly low across the four epiRILs tested, induction in response to cold was
560 much stronger for the two epiRILs with a *ddm1*-derived, hypomethylated TE-DMR (20-43-fold
561 vs 12-18-fold; Fig. 7d, S13c). Thus, epiallelic variation at *HVA22E* enables plants to
562 dramatically increase their response to cold stress without affecting basal expression.

563 While the TE-DMR downstream of *HVA22E* frequently reverts in the epiRILs, it is often
564 observed (24%) in the epivariant state in nature, which suggests the presence of stabilizing
565 genetic or genomic modifiers, or positive selection. Although discriminating between these
566 possibilities is beyond the scope of the present study, we exploited the fact that the TE-DMR is
567 in the epivariant state in Cvi-0 to assess its stability in the Cvi-0 x Col-0 RILs. Remarkably, we
568 observed reversion in 17% of the RILs that inherited it from Cvi-0 (Fig. S13d), which indicates a
569 strong destabilizing effect of the Col-0 genetic background on the epivariant form, consistent
570 with the frequent reversion observed in the epiRILs.

571 Strains containing the epivariant form of *HVA22E* are preferentially found in habitats
572 characterized by high summer frost and low summer precipitation (Fig. 7e, S13e). Furthermore,
573 analysis of the AraPheno data indicates that these strains tend to have smaller stomata (Fig. 7f).
574 On the basis of these observations, we compared the growth response of epiRILs depending on
575 the epiallelic state of the *HVA22E* TE-DMR using a high-throughput phenotyping platform that
576 is robust enough to establish physiologically-relevant mild drought stress conditions[90]. While all
577 epiRILs with a reverted *HVA22E* epiallele grew similarly to those where it is WT-derived under
578 both conditions, the epiRILs with an hypomethylated epiallele showed evidence of faster

579 recovery to mild drought (Fig. 7g, S13f). Thus, epiallelic variation at *HVA22E* likely acts as a
580 selectable engine of increased response to stress in nature.

581 Discussion

582 TE-mediated TEI is prevalent in nature

583 We have demonstrated that numerous TEs across the Arabidopsis genome are recurrently
584 subjected to heritable, severe loss of DNA methylation in nature. Furthermore, we have
585 demonstrated through comprehensive segregation analyses that most natural epivariants are *bona*
586 *fide* epialleles, as their inheritance is not strictly dictated by DNA sequence changes, neither in
587 *cis* nor in *trans.* Collectively, our findings establish TE-mediated TEI as an important additional
588 source of heritable variation in gene expression in nature, with unique properties for plant
589 adaptation.

590 Inheritance of TE epivariation is constrained by *trans*-acting RdDM

591 The in-depth characterization of TE-DMRs and their inheritance patterns in the epiRILs revealed
592 two key properties of TE-mediated TEI. First, we demonstrate that transgenerational inheritance
593 can vary markedly for the same TE-DMR between epiRILs (Fig. 1), because the 23-24nt sRNAs
594 required for RdDM-mediated reversion are mainly provided in *trans* by related TE copies
595 present elsewhere in the genome and which must be inherited in the methylated state (Fig. 2-3).
596 Incidentally, this observation confirms and extends to endogenous sequences the conclusion
597 reached using transgenes that efficient targeting of RdDM usually requires the production of
598 large amounts of 23-24nt sRNAs[59–63]. Second, we show that pericentromeric TEs play a
599 predominant role in limiting TEI, consistent with results obtained using a panel of F1
600 "epihybrids"[91]. In turn, this observation provides direct support to the proposition that the 23-24
601 nt sRNAs produced by pericentromeric TEs, which have little impact in *cis*, may mainly serve to
602 guide DNA methylation of related copies located in the chromosome arms[92–95].

603 More generally, the small number of TE-DMRs that systematically revert in the epiRILs (r-
604 100%) and the prevalence of non-allelic interactions in driving reversion in this setting provide
605 compelling proof that paramutation is an epigenetic process predominantly involving ectopic

606 rather than allelic interactions, as proposed over two decades ago[96–98]. In other words, classical
607 examples of paramutation between alleles or epialleles at the same locus[99] would be the
608 exception rather than the rule. In fact, numerous lines of evidence support the notion that a one-
609 to-one ratio of sources to targets of 23-24nt sRNAs is generally insufficient to trigger
610 paramutation[51], including the observation that RdDM-mediated paramutation at a transgenic
611 locus in *A. thaliana* can only be achieved in tetraploids with a three-to-one ratio of methylated to
612 unmethylated epialleles[68,100]. It is also consistent with the scarcity of paramutation events in the
613 F1 progeny of crosses between *A. thaliana* strains[101], which our analysis of the Cvi-0 x Col-0
614 RILs confirmed (Fig. 5, S9).

615 **Paf1 is a major facilitator of natural epivariation at TEs near genes**

616 Naturally epivariable TE-DMRs are enriched near genes (Fig. 6a-b) and we identified *ELF8*
617 (Fig. 6, S10), which encodes a component of the evolutionary conserved Paf1 complex, as the
618 second most important genetic or genomic modifier of the prevalence of epivariation in nature,
619 after the number of related TE copies (Fig. 6g). We also showed, in the epiRILs, that *ddm1*-
620 induced Pol II transcription at so-called EPICATs [66] located near or within a TE-DMR facilitates
621 the stable inheritance of hypomethylation, despite RdDM-targeting (Fig. 4). Therefore, gene
622 expression is likely a major *cis* determinant of the presence of epivariation at TEs, consistent
623 with the known antagonizing effect of Paf1 and active histone marks on small RNA-mediated
624 epigenetic silencing in *Schizosaccharomyces pombe*[102] and *A. thaliana*[79], respectively.
625 Remarkably, our findings mirror the discovery in *S. pombe* that Paf1 activity is required to
626 prevent the establishment and epigenetic inheritance of the force silencing of a euchromatic
627 locus by *trans*-acting sRNAs[103]. Thus, we hypothesize that in addition to its many known
628 functions in facilitating Pol-II transcription, the Paf1 complex also plays an important role in
629 ensuring that gene expression is not affected by the silencing of neighboring TEs.

630 **TE epivariation arises sporadically and recurrently in nature**

631 In any given strain, natural epivariants are usually few and widely dispersed across the genome
632 (Fig. 5a,f, S7c), which rules out an *en masse* origin as in *ddm1*. Moreover, results of our
633 haplotyping and kinship analyses exclude identity by descent as the main determinant of the

634 presence of epivariants at the same TE across multiple strains (Fig. 5d, S8). Instead, all evidence
635 points to a spontaneous, non-coordinated and recurrent origin of TE epivariants in nature.

636 Given the frequent associations we uncovered with climate variables at collection sites (Fig. 7b,
637 S12d-f), it is tempting to speculate that the environment acts not only as a selective force but also
638 as an inducer of TE epivariation in nature. However, severe and heritable loss of DNA
639 methylation at TEs in response to environmental cues has not been reported so far, despite clear
640 effects of abiotic stresses on epimutation rates genome-wide[104–107] as well as anecdotal evidence
641 of transient and moderate loss of DNA methylation at some TEs induced by biotic or abiotic
642 stresses (e.g.[108–110]). Furthermore, our comparison of RNA-seq data obtained for natural strains
643 and the epiRILs indicates that epivariation often causes changes in nearby gene expression rather
644 than the opposite (Fig. 7, S11-13). Thus, which environmental factors, if any, favor the
645 recurrence of heritable *ddm1*-like epivariants in nature remains to be determined.

646 **Natural TE epivariation exacerbates the induction of stress-response genes**

647 Natural epivariation at TEs is usually not associated with large changes (>2-fold) in the
648 expression of surrounding genes, at least when plants are grown under standard conditions
649 (S11a-b). This finding is of course consistent with the fact that genomes are populated by TEs
650 that mainly correspond to ancestral insertions, most of which must have accumulated because
651 they have no or mild deleterious effects[16]. Thus, the 151 TE-DMRs for which we detected an
652 association between epivariation and at least one biogeoclimatic variable (Fig. S12e-f) would
653 simply represent neutral "bio-markers", similar to the metastable DNA methylation variants
654 described at some TEs in the mouse[111]. However, the observation that naturally epivariable TE-
655 DMRs are particularly enriched near stress-responsive genes (Fig. S12g) suggests a more
656 nuanced picture. As we showed for the gene *HVA22E* (Fig. 7d, S13a-c), which is involved in
657 cold- and drought-stress responses, many natural epivariants may in fact affect gene expression,
658 but specifically in response to stress. Furthermore, our observation that epivariation at *HVA22E*
659 correlates with climate of native habitats (Fig. 7e-f, S13e) suggests that it is adaptive. Such
660 epivariants would therefore represent unique targets of natural selection because they could
661 revert much more readily than sequence variants and also because most of them should not cause
662 fitness trade-offs when plants are unchallenged by stress.

663 **TEI across the eukaryotic tree of life**

664 Natural epivariation at TEs as an additional source of heritable phenotypes is unlikely to be
665 limited to plants with small genomes and may be more significant as the genomic density of TEs
666 increases. Consistent with this notion, *ddm1* knock-out mutants are either non-viable or sterile in
667 maize and tomato[40,41], two crop species that have much larger, TE-laden genomes compared to
668 *A. thaliana*. Although this situation complicates considerably the characterization of TE-
669 mediated TEI at the genome- as well as population-wide scale, two studies in maize suggest that
670 TE-mediated epialleles are abundant and phenotypically consequential in this species as
671 well[112,113].

672 In mice, a number of TE-mediated, DNA methylation-based metastable epialleles have been
673 identified, but their inheritance properties as well as their usually minimal contribution to gene
674 expression differences suggest a minor role as a source of heritable phenotypes[35,111].
675 Nonetheless, experiments aimed at forcing DNA methylation or other chromatin changes at
676 endogenous transgenic loci or genome-wide in the mouse[114], Drosophila[115] and *C. elegans*[116–118]
677 indicated that such changes can be inherited across multiple generations in the absence of the
678 trigger in these species. However, similar heritable chromatin alterations have not been reported
679 in natural settings, leaving in doubt whether chromatin-based TEI plays any significant role in
680 these species. Evidence in *S. pombe*[119,120] that exposure to caffeine can promote spontaneous
681 chromatin changes that are adaptive but which cannot be sustained in the absence of the
682 antifungal agent suggests a similar conclusion in this species as well. Plants, perhaps because
683 they are sessile organisms, may therefore be relatively unique in exploiting chromatin-based TEI
684 as an additional system of inheritance with important ecological roles in nature.

685 **Limitations of the study**

686 Our study likely underestimates the extent of natural *ddm1*-like epivariation at TEs given the
687 stringent filters we used to define it. Moreover, we did not consider TEs not present in the
688 reference Col-0 genome, which typically correspond to recent, large-effect insertions[15] and
689 which can also epivary in nature[121]. Furthermore, *ddm1*-induced TE-DMRs do not include, by
690 design, any of the ~25% of TEs that are already hypo- or un-methylated in WT Col-0[122]. This
691 limitation effectively biased our GWAS and GLM approaches towards finding genetic modifiers

692 that increase rather than decrease the prevalence of epivariation. Also, while TEs are implicated
693 in most known examples of epiallelic variation with phenotypic consequences in plants, DNA
694 methylation variation affecting other types of sequences, such as non-coding RNA genes and
695 some protein-coding genes, may also contribute to phenotypic diversity in Arabidopsis[123–125].

696 Material and Methods

697 Genomic annotations

698 All genomic analyses were performed using the TAIR10 Col-0 reference genome sequence
699 assembly. The reference TE and gene annotations were used after removal from the latter of the
700 3,903 TE genes ('transposable_element_gene'; https://www.arabidopsis.org/download/).

701 Bisulfite-sequencing of 120 *ddm1*-derived epiRILs

702 Methylomes were produced for the *ddm1* and WT parental lines as well as for a set of 120 *ddm1*-
703 derived epiRILs[48] largely overlapping with those that were initially epihaplotyped using low-
704 resolution methods[49]. EpiRILs were taken one generation later (F9) than previously[49] and also at
705 generation F21 for 12 epiRILs. Whole-genome Bisulfite sequencing (WGBS-seq, Illumina
706 HiSeq X Ten) was performed on gDNA extracted from rosette leaves (pooled from six
707 individuals grown in long-days, 16h:8h light:dark, at 23°C and collected at bolting) of the 120 F9
708 epiRILs as well as 3 biological replicates of Col-0 WT and *ddm1* G4 parents. Numbers of
709 methylated and unmethylated reads over single cytosines were obtained using Bismark[126] and
710 pooled into single sites for Cs on opposite strands of CG and CWG sites (where W is A or T)
711 which we confirmed to be concordant (Fig. S2a) given the symmetry of both contexts (in
712 contrast to CCG where one of the strands is both a CHG and a CG site).

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

715 Differentially methylated sites (DMSs) between the three replicates of *ddm1* and WT parents
716 were identified using a generalized linear model (GLM, R package *glm*): out of 2.35 millions CG
717 and CWG symmetric sites with sufficient coverage (≥5 reads per site) in at least 20 epiRILs,

718 807,450 were found hypomethylated by at least 20% in *ddm1*. Consistent with previous
719 observations[45], most (72.8%) of these DMSs affect annotated TEs or their immediate flanks
720 (300bp on each side, Fig. S1a). Using the same GLM, we then determined the DNA methylation
721 status (i.e. WT-like, *ddm1*-like or intermediate) of these hypomethylated DMSs across the
722 informative epiRILs for each site by comparing the DNA methylation levels calculated at each
723 DMS in each epiRIL with the predictions from the GLM for each genotype (R function
724 predict.glm): all values below the upper 95% confidence interval (1.96 standard error) of *ddm1*
725 were considered *ddm1*-like, those above the lower 95% of WT were considered WT-like, and
726 those in between these two thresholds were considered intermediate. For each DMS, we then
727 compared to the expected ratio of Mendelian inheritance of 73-27% (F1 between *ddm1* and WT
728 is backcrossed to WT with 8% of selfing[48]) the ratio of WT-like epiRILs across all informative
729 epiRILs for the site. To take into account the number of informative epiRILs, we used a poisson-
730 binomial drawing from the selfing (50-50) or the backcrossing (75-25) distribution: for 20
731 informative epiRILs the 95% confidence interval of Mendelian segregation is 45%-95% while it
732 shrinks to 64%-82% when the full set of 120 epiRILs is informative (Fig. S1b). In total, 202,721
733 DMSs show WT-like methylation in a percentage of epiRILs consistent with Mendelian
734 inheritance, whereas for another 587,197 DMSs this percentage is instead indicative of either
735 partially or fully penetrant reversion, with some rare cases (2.2% of all DMSs) with either less
736 than 20 informative epiRILs or with a scarcity of epiRILs with WT-like methylation (Fig. S1b).

737 **Identification of differentially methylated regions (DMRs)**

738 As expected from previous analyses[49], parental DMSs close to each other show coherent
739 epiallelic inheritance across the epiRILs (Fig. S2b), a property that we used to iteratively merge
740 coherently-segregating DMSs into differentially methylated regions (DMRs) as long as they
741 were of the same class (Mendelian or reverting), separated by no more than one DMS of another
742 class, and within 207bp of each other. This threshold distance was chosen as the distance from
743 which the correlation between the methylation status of two adjacent DMSs across epiRILs starts
744 to significantly increase in variability (Fig. S2b). Out of these DMRs we removed those based on
745 less than 5 DMSs and thus obtained 9,422 Mendelian DMRs and 16,040 reverting (Dataset S1).

746 We then calculated the correlation between the methylation states of each pair of the 7,151
747 Mendelian DMRs that contain at least 5 DMSs covered in all 120 epiRILs (Dataset S2) to deduce
748 the rate of linkage decay across chromosome arms and pericentromeres (no decay observed in
749 pericentromeres given the scarcity of crossing-overs in these regions). Using a non-linear
750 regression model with variable variance to estimate the expectations of linkage decay as a
751 function of genomic distance - linear in pericentromeres where there is almost no decay (R
752 function nlme::gls) and exponential in chromosome arms (nlme::gnls) - we filtered out 1729
753 outlier DMRs for which methylation states are either too highly or too lowly correlated (outside
754 of 95% confidence intervals) with other DMRs across the epiRILs (Fig. S2c). High correlation at
755 a distance is observed notably for DMRs overlapping *VANDAL* and *ENSPM* TE annotations, in
756 line with the ability of TEs from these TE families to demethylate cognate TEs in *trans* [57] and we
757 therefore removed all the DMRs corresponding to these TE families. Redundant DMRs which
758 are closeby (<5kb) and highly correlated (R^2>0.6) were also trimmed down to one representative
759 DMR. For pericentromeric regions, given the high density of DMRs and the scarcity of crossing-
760 overs, we selected between 80 and 100 evenly distributed DMRs for each pericentromere. In
761 total, we obtained a subset of 1,092 evenly distributed and non-redundant Mendelian DMRs
762 (Dataset S3).

763 **Building the epihaplotypic map of 120 F9 epiRILs**

764 To infer the epihaplotypes from the 1,092 DMRs thus defined, we denoised the data using a
765 continuous-time Hidden Markov Model (HMM, R package MSM), with methylation status of
766 DMRs as observed states and the real parental haplotype as hidden states. The initial state
767 occupancy probabilities are the same as the Mendelian expectation (27% *ddm1* vs 73% WT).
768 Using recombination rates measured in female back-crosses[127] we calculated the expected
769 recombination rate in our F9 epiRIL population (assuming complete homozygosity by this
770 generation)[128]. From these recombination rates we calculated the expected transition matrix
771 (WT-derived to *ddm1*-derived and the opposite).

772 As pericentromeric regions rarely recombine, we used *ddm1*-like and WT-like pericentromeres
773 to calculate the distribution of methylation states of DMRs within *ddm1*-derived and WT-derived
774 intervals, respectively. We used these distributions as emission probability functions to optimize

775 the given transition matrix to adjust for the lack of recombination events in the pericentromeres.
776 Together with the initial state occupancy probabilities and the transition matrix, we used these
777 emission probability functions as parameters in our HMM to derive epihaplotypic intervals
778 delimited by DMRs. Short intervals (<1.5 Mb) were removed from further analysis as well as the
779 small fraction (2%) of larger intervals with DMRs mainly in the intermediate rather than in the
780 *ddm1*- or WT-like methylation state (see e.g. in Fig. S2e), which may correspond to regions of
781 residual heterozygosity. Finally, we noted five epiRILs (#62, #172, #323, #394 and #506) with a
782 high density of *ddm1*-like DMRs in the wild-type derived pericentromeric region of chromosome
783 4. Remarkably, *VANDAL* and *ENSPM* are also hypomethylated in these intervals, despite the
784 absence of a *trans*-demethylating copy present in a *ddm1*-derived interval. The most
785 parsimonious explanation is that these correspond to *ddm1*-derived intervals with high level or
786 reversion. These intervals were also removed from further analysis.

787 Overall, the new epihaplotypes (Dataset S3) are more resolutive and extend further than the
788 initial ones (Fig. 1a, S3). Indeed, despite excluding intervals that are poorly resolved or that
789 could reflect residual heterozygosity, we can nonetheless now infer unambiguously the WT or
790 *ddm1* origin ~93% of the genome sequence of each epiRIL on average, compared to 82% before
791 (Fig. S3).

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

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

800 To define lower and upper thresholds of WT-like and *ddm1*-like mCG levels, respectively, we
801 took the first percentile of average mCG levels over these TE-DMRs in the three WT replicates
802 and the last percentile in the three *ddm1* replicates (Fig. S1c). The methylation state of a TE-
803 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 (Dataset S4). 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. "Stray" TE-DMRs were identified, for non-*VANDAL* and non-*ENSPM* TE families with 4 or more 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.

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 (Table S2) 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).

831 **Single-molecule analysis of DNA methylation using ONT-sequencing.**

832 In order to enrich the ONT sequencing power over regions of interest, we used the adaptive
833 sampling option available on the GridION sequencer (Oxford Nanopore Technologies). High-
834 molecular-weight genomic DNA (>20 kb-long fragments) was extracted from grinded leaves of
835 the same pool of 6 epiRIL238 F9 plants used for BS-seq following a protocol adapted from
836 *Mayjonade et al, 2016* [129] for bigger extraction volumes. In order to reduce the library size for
837 adaptive sampling, DNA was first sheared at 10 ng.uL-1 by sonication for a single 30s interval at
838 the lowest sonication intensity on a standard Bioruptor (Diagenode) to obtain a range of DNA
839 fragments from 200 bp to >20 kb. Sonication was followed by size selection with Sera-Mag
840 SpeedBeads Carboxylate (GE Healthcare)[130], to eliminate fragments smaller than ~1500 bp.
841 Sequencing was performed on a LSK109 flow cell following manufacturer's procedures and
842 using for adaptive sampling a 1.07 Mb-long set of target regions (~1% of reference genome) that
843 overlap by at least 5 kb the 10 TE-DMRs localized in *ddm1*-derived intervals in epiRIL238 with
844 intermediate DNA methylation levels as well as 45 TE-DMRs that have fully reverted (Table
845 S3). mCG levels were calculated over TE-DMRs using DeepSignal-plant[131] either at the regional
846 level or at the single-read level over the TE-DMR interval for reads that cover >50% of the TE-
847 DMR.

848 **Small RNA mapping and analysis**

849 Total RNA was extracted with Trizol according to manufacturer's instructions (Invitrogen) from
850 pools of seedlings of *ddm1* and *ddm1rdr2*, *ddm1rdr6*, and *ddm1dcl2dcl3* compound mutants as
851 well as from inflorescences collected from pools of 5-10 siblings of 10 F9 epiRILs selected for
852 having contrasted epihaplotypic combinations (Fig. 1a, Dataset S3). sRNAs were sequenced by
853 Fasteris with Illumina for the mutants and by the BGI with DNBseq for the 10 F9 epiRILs.

854 Reads were mapped using bowtie2[132] using the --very-sensitive mode (-L 15 -N 1 -p 4) from
855 which unmapped reads and secondary alignments were removed using samtools view (-F
856 260)[133]. The total number of reads ranging in length from 15 to 30nt and mapping on nuclear
857 chromosomes was extracted using samtools view to obtain the sRNA library size of each sample.
858 Multi-mapping reads (attributed randomly to any of its mapping locations by bowtie2 in the
859 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. S4c). 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 mapping of mCG levels per TE family

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 (Dataset S3), 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, 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. S5d). 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

888 of LOD scores above the significance threshold minus 1 with the top marker was used as second
889 ^{epi}QTL.

890 The percentage of variance explained by the epiallelic segregation at each of the ^{epi}QTLs was
891 estimated by adding iteratively their epihaplotypes as parameters in a logistic GLM. We also
892 used the epihaplotype at the top ^{epi}QTL in a logistic GLM of average CG methylation levels over
893 the same TE-DMRs but within WT-derived intervals to identify the 24 TE families where it is
894 significantly associated with the epihaplotypic segregation at this locus (p-value<0.05) and thus
895 where effects within *ddm1*-derived intervals may be the result of *trans*-demethylating effects
896 (Fig. S5e).

897 **Phylogeny and divergence of TE sequences**

898 Divergence between related TE-DMRs was calculated by building a phylogeny of all the TE-
899 DMRs within each TE family using phyml (v3.3.3)[134] with parameters -s SPR -d nt -a e -q --
900 no_memory_check after trimming the MAFFT (v7.487)[135] alignment of fasta sequences using
901 trimal[136] in -phylip3.2. For each TE-DMR the distance to the closest TE-DMR was retrieved
902 from the phylogeny of the corresponding TE family using the DendroPy-4.4.0 python library[137].

903 **Identification of spontaneous hypomethylation events in the epiRILs**

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

911 Epihaplotyping of the 169 epiRILs that minimally overlaps with ours (37 epiRILs are in
912 common) for which single-cytosine resolution methylome data were obtained independently by
913 others[69] was performed following the method they described: i.e. based on the WT-like (>0.5) or
914 *ddm1*-like (<0.5) mCG levels of 140 of the 144 DMRs they identified as stably segregating (4
915 were removed as they deviate by more than 7% from the 75% of WT-like expected). Using this

916 epihaplotypic map, we identified 636 TE-DMRs with at least one event of *ddm1*-like mCG levels
917 within WT-derived intervals that do not belong to the *VANDAL* and *ENSPM* superfamilies.
918 However, we found events of hypomethylation within WT-derived intervals to occur almost
919 systematically on the same epiRILs (pearson correlation > 0.8) for two blocks of 266 and 196
920 TE-DMRs on Chr3 and Chr5, respectively (Fig. S6e), suggestive of an epihaplotyping error over
921 the corresponding intervals that are most likely *ddm1*-derived. 161 TE-DMRs remained after
922 filtering out these 475 TE-DMRs.

923 **Epivariation over TE-DMRs in natural strains**

924 Processed bisulfite sequencing data (BS-seq) and paired-end short reads whole-genome
925 sequencing (WGS) data from leaves of 720 genomes of *A.thaliana* were obtained from the 1,001
926 Genomes Project (1001genomes.org and signal.salk.edu/1001.php). For each strain, we
927 calculated the average CG methylation weighted by the sequencing depth at each site over TE-
928 DMRs within TEs that are present in the strain based on previous work[15]. In addition, WGS
929 coverage was calculated using the tool bam-readcount[138] in order to detect potential structural
930 variation unaccounted for by sole TE presence/absence polymorphisms. Specifically, for each
931 TE-DMR in each genome, we filtered out regions that have not enough WG-seq coverage (>5%
932 of TE-DMR with 0 coverage) or an excess of reads (median TE-DMR coverage > 2x genome
933 wide median coverage). In addition, we filtered out TE-DMRs with a number of CG sites overly
934 divergent from the reference genome (>50%) or with less than 5 BS-seq reads per CG on
935 average.

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

937 Long-read Oxford Nanopore Technology (ONT) sequencing data were produced for 20 strains
938 (chosen to represent each genetic group; Table S5) out of the 720 for which BS-seq data is
939 available. High-molecular-weight genomic DNA was extracted from grinded leaves[129] and
940 sequenced at ~30X (N50=37kb, Table S5) following recommended guidelines on ONT MinION
941 flow cell (R9.4.1).

942 Basecalling was performed using Guppy v6.4.6 with the dna_r9.4.1_450bps_sup_plant.cfg
943 model. Long-read fastq files were used for *de novo* assembly using two different assembly tools:

944 Flye[139] and NextDenovo[140]. The assembly with the best N50 was selected. To fix base errors in
945 the assemblies, we polished the genomes using medaka with long-read data
946 (https://github.com/nanoporetech/medaka), and two rounds of NextPolish[141] with short-read data.
947 Finally, the contigs of assemblies are scaffolded using Ragtag[142] based on alignment on the
948 reference genome. The location of reference genes and TEs were transposed on each genome
949 using Liftoff[143], then TE-DMRs were located within their TE by blast. TE-DMRs were
950 considered to be located at a syntenic region in the *de novo* assembly when surrounded by the
951 same reference genes in the genome assembly as in TAIR10. For this analysis we required that at
952 least one the three nearest genes (restricted to TAIR10 IDs ending in 0 to avoid mis-annotated
953 TE-genes that are themselves likely to be translocated between strains) on each side of the TE-
954 DMRs in the *de novo* assemblies were also found surrounding the TE-DMRs in the reference
955 genome. When located in synteny, TE-DMRs were considered full-length when their length in
956 the *de novo* assembly was greater than 80% of that in TAIR10. BSseq data was then remapped
957 on each *de novo* assembly using Bismark[126] to calculate mCG levels over the TE-DMRs.

958 **Kinship, local haplotyping, and GWAS**

959 Genome wide IBS and aBN kinship matrix were calculated using the emmax-kin-intel64
960 function of EMMAX (https://genome.sph.umich.edu/wiki/EMMAX) with parameters -v -s -d 10
961 and -v -d 10, respectively, on the PLINK-transposed (v1.90p)[144] vcf of the short variants
962 identified in the 1001 Genomes (1001genomes_snp-short-indel_with_tair10_only_ACGTN.vcf
963 from https://1001genomes.org/data/GMI-MPI/releases/v3.1/) out of which we only kept bi-allelic
964 SNPs with a minimum MAF of 0.05 and a minimum missing genotyping rate of 0.1 and found
965 within the 720 strains with BS-seq data (SNP720.MAF005.g01 vcf).
966
967 Local haplotyping of each TE-DMRs was calculated using BEAGLE[145] on the
968 SNP720.MAF005.g01 vcf in windows of 1 kb (500 bp on each side of the most central SNP
969 within each TE-DMR). Local PCAs (Fig. S7l-m) were performed on the SNPs from
970 SNP720.MAF005.g01 located within 1 kb around the TE-DMR.
971
972 GWASs were performed using emmax-intel64 with parameters -v -d 10 using the aBN kinship
973 matrix and the SNP720.MAF005.g01 vcf. For epivariation prevalence at each TE-DMR, we only

974 analyzed the 490 naturally epivariable TE-DMRs with at least 5 *ddm1*-like epivariants and with
975 low genomic inflation (λ_{GC}<1.1). For heatmap representation of the results (Fig. S9a) we
976 subsampled to the SNPs with the highest p-values within 20 kb windows across the genome. For
977 genome-wide epivariation prevalence, we used as phenotype the proportion of natural *ddm1*-like
978 epivariant per strain among all naturally epivariable TE-DMRs present in the strain.

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

980 We selected for an extended segregation analysis 40 of the 115 naturally hypomethylated TE-
981 DMRs identified in Cvi-0 to represent a variety of superfamilies and reversion categories (Fig.
982 5c). Genomic DNA was extracted from frozen leaf tissue using Qiagen DNeasy Plant Mini Kit,
983 and 30 to 100ng was used for McrBC (NEB) overnight digestion (15hr). Levels of unmethylated
984 DNA over a target region were calculated as the fraction of DNA concentration measured by
985 qPCR remaining following McrBC digestion compared to the same sample where McrBC is
986 replaced by water (undigested). Efficiency of digestion was confirmed using the *ddm1*-
987 independent DNA methylation at the 3' end of *AT5G13440* as a positive control, and the
988 unmethylated *AT5G13440* as a negative control (primers listed in Table S7).

989 Using primers designed with primer3 (https://github.com/primer3-org/primer3) to specifically
990 amplify each region (Table S7), we first confirmed for 34 out of the 38 TE-DMRs that they are
991 differentially methylated between Col-0 and Cvi-0 in freshly harvested tissue of two sibling
992 plants. The remaining 4 TE-DMRs either displayed WT-like methylation in Cvi-0 (1 TE-DMR),
993 hypomethylation in Col-0 (1 TE-DMR), or did not amplify in Cvi-0 (2 TE-DMRs), and we
994 therefore excluded them from further analysis. Using McrBC-qPCR, we then measured the
995 methylation levels of the 34 differentially methylated TE-DMRs across a diverse set of 36 F8
996 Cvi x Col RILs (Table S8), that include the 20 of the minimal set identified by Simon et al. [74] as
997 well as 16 additional with maximal genetic differences to the initial set of 20. TE-DMRs were
998 considered hypomethylated when McrBC-qPCR methylation levels were below 40%, WT-like
999 methylated when above 60%, and intermediate otherwise. TE-DMRs and ELF8 alleles were
1000 considered as Cvi-derived when both bordering markers are of the Cvi-genotype (B), Col-
1001 derived when both are A, and NA otherwise (C or D genotypes or non-agreeing bordering

1002 markers). Reversion events were identified as Cvi-derived TE-DMRs with WT-like methylation
1003 levels.

1004 Genetic associations between TE-DMR methylation levels and the genetic maps of these RILs[74]
1005 were computed using the mqmscan function of the qtl R package with a window.size of 10 and a
1006 step.size of 2. The significance threshold at alpha = 0.05 was calculated using mqmpermutation
1007 with the same parameters as well as n.perm = 10 and batchsize = 25. LOD scores were
1008 normalized by the significance threshold for heatmap visualization (Fig. 5c). For one TE-DMR,
1009 carried by an *ATCOPIA23* on chromosome 3 in Col-0, associations are strongest with markers in
1010 *trans* on chromosome 5 which correspond to a TE insertion polymorphism of *ATCOPIA23*
1011 identified at this location on chromosome 5 in Cvi-0[15]. By mapping BS-seq reads on the Cvi-0
1012 ONT assembly *ATCOPIA23* was confirmed to be absent in Cvi-0 from its original location on
1013 Chr3 and present on Chr5 in an hypomethylated state (Fig. S9d), thus explaining the *trans*
1014 association of hypomethylation with markers bordering that locus.

1015 For the 10 RILs for which BS-seq data was publicly available[75], mCG levels were calculated
1016 over all naturally epivariable TE-DMRs with *ddm1*-like methylation in Cvi-0 (N=104, Fig. S9e,
1017 S10i) or with WT-like methylation in Cvi-0 (N=465, Fig. S10h). Average mCG levels were
1018 calculated over both replicates of each RIL and compared to each replicates of the Col-0 and
1019 Cvi-0 parental lineages sequenced as part of the same study (Fig. S9e, S10h-i). Hypomethylation
1020 and WT-like methylation were classified using the thresholds defined in the epiRILs (<0.3944
1021 for *ddm1*-like and >0.7586 for WT-like). Genotypes were again determined using concordant
1022 bordering markers, with reversion events identified as Cvi-derived TE-DMRs with WT-like
1023 methylation levels, while spontaneous hypomethylation events as Col-derived TE-DMRs with
1024 *ddm1*-like methylation levels. Correlations between methylation levels and genotypes were
1025 calculated as a Pearson correlation coefficient.

1026 **Generalized linear modeling of genetic and environmental determinants**

1027 Association between the presence of *ddm1*-like epivariants over the 1068 naturally epivariable
1028 TE-DMRs and a set of selected genetic parameters was calculated using a binomial Generalized
1029 linear model (logit link) using R glm function. In order to limit the risk of overfitting, we
1030 restricted our analysis to the following set of the genetic parameters which we deemed most

1031 biologically relevant: copy number of each TE family in each genome, genotype at *ELF8* (using
1032 the top SNP at 2:2431793 identified in our GWAS), genotype at *NRPE1* (using the top SNP at
1033 2:16719082 identified in previous studies[15,80]), presence of an EPICAT as defined previously[66]
1034 within, under 500bp, or >500bp from the TE-DMR, distance to the nearest gene, density of CG
1035 sites, and length of the TE-DMR. TE family copy-numbers were estimated using average read-
1036 depth after remapping short-read sequencing to a library of consensus TEs as described
1037 previously[14]. For all 10 parameters, the variables were tested as fixed effects in the glm, and for
1038 *NRPE1* and *ELF8* genotypes we also included interaction effects with copy-number and gene
1039 distance respectively. Discrete parameters such as genotypes or associations with EPICATs were
1040 considered as categorical variables, while quantitative parameters were Z-scored (R scale
1041 function) to avoid range effects. For each fixed or interaction effect, an individual GLM was
1042 built to assess their significance in explaining the prevalence of *ddm1*-like epivariation over the
1043 1068 naturally epivariable TE-DMRs after including the three first components of the PCA of
1044 the aBN kinship matrix calculated from the `SNP720.MAF005.g01` vcf to take into account
1045 population structure (together the three PCs explain ~55% of the variance). If significant
1046 ($p<0.05$), the variable with the largest R^2 (McFadden adjusted) was included in the GLM, to then
1047 iteratively test the effect of the remaining variables. For significant interaction effects, the
1048 variable was included along with its fixed effect.
1049
1050 196 quantitative AraClim biovariables were downloaded from
1051 https://gramene.org/CLIMtools/arabidopsis_v2.0/AraCLIM-V2/. After Z-scoring (R scale
1052 function), they were included individually as an explanatory variable in individual binomial
1053 GLMs (logit link) built for each of the 490 TE-DMRs with sufficient epivariant counts (see
1054 GWAS methods) which included the first three PCs of the kinship matrix, the potential *cis*
1055 modifiers identified by GWAS (Fig. S9a-b), and the two main *trans* modifiers of the number of
1056 *ddm1*-like epivariation per strain, namely TE copy-number and *ELF8* (Fig. 6f). P values were
1057 then corrected using the Benjamini and Hochberg correction for false discovery rate using the R
1058 p.adjust function. When several AraClim variables were found to be significant (adjusted
1059 $p<0.05$), the one explaining the largest fraction of variance (McFadden adjusted R^2) was
1060 selected.

1061 **RNA-seq analyses in natural strains and in the epiRILs**

1062 For analysis of gene expression in natural strains, RNA-seq data of the 728 strains analyzed
1063 previously[53] was downloaded from GEO (Accession GSE80744). For analysis of gene
1064 expression in the epiRILs, total RNA was extracted with Trizol according to manufacturer's
1065 instructions (Invitrogen) from rosette leaves collected from pools of 5-10 siblings of the same set
1066 of 10 F9 epiRILs selected for sRNA analysis and sequenced by the BGI. Reads were mapped
1067 using STAR[146] with options --outFilterMultimapNmax 1 --alignIntronMax 10000. Coverage over
1068 exons was calculated using bedtools coverage and read counts were then normalized over entire
1069 transcripts using R package DEseq2. A pseudo count of 0.001 was added to all normalized
1070 expression levels to compute log ratios.

1071 For each naturally epivariable TE-DMRs, expression levels of the two closest genes were
1072 compared between strains carrying the TE-DMR with *ddm1*-like or with WT-like methylation by
1073 computing the log10 ratio between the median expression in the first set over that in the second
1074 set of strains (hypo/hyper log ratios) as well as the p-value of a Wilcoxon rank-sum test between
1075 the two sets. To obtain random expectations of hypo/hyper log ratios (gray lines in Fig. 7a and
1076 S11a) we shuffled randomly the strain labels among all strains with RNA-seq data and repeated
1077 the reshuffling for a total of 10 times.

1078 For TE-DMRs where epivariation in nature is associated with ≥2-fold expression changes at one
1079 of the two nearest genes, we searched in the set of 10 epiRILs sequenced for those that were
1080 stably inherited hypomethylated from the *ddm1* parent in at least one of the epiRILs. To limit the
1081 possibility of indirect epihaplotypic effects that could be due to the co-occurrence of multiple
1082 hypomethylated TE-DMRs in the epiRILs, we restricted our analysis to naturally epivariable TE-
1083 DMRs that are within 5kb of the nearest gene and which are not separated from it by another
1084 naturally epivariable TE-DMR. For the 99 TE-DMRs matching these criteria we then calculated
1085 hypo/hyper log10 ratios of gene expression averages between the epiRILs where they are *ddm1*-
1086 derived and hypomethylated and the epiRILs where they are WT-derived and WT-like
1087 methylated. For the TE-DMRs that have reverted to WT-like methylation in at least one but not
1088 all of the epiRILs where it is *ddm1*-derived, we also calculated the log10 ratio of gene expression
1089 averages between the epiRILs where they are *ddm1*-derived and hypomethylated and the epiRILs
1090 where they are *ddm1*-derived and WT-like methylated (rev/hypo log ratios, Fig. S11f).

1091 **Functional analysis of *ELF8* alleles**

1092 Although the coding-sequence of *ELF8'* differs from that of the reference allele by 40 SNPs
1093 (Fig. 6f), none of these are missense. To confirm that the frame-shifting single nucleotide
1094 deletion present in the TAIR10 sequence of *ELF8* (indicated with a * in Fig. 6f) is a sequencing
1095 error, we sequenced the cDNA of the Col-0 *ELF8* allele both upstream and downstream. RNA
1096 extractions were performed on flash-frozen ground tissue (shoots) from Col-0 WT using
1097 Macherey-Nagel Nucleospin RNA plant kit following manufacturer's instructions. Reverse-
1098 transcriptase was performed using SuperScript IV (ThermoFisher Scientific) with oligo $d(T)_{20}$
1099 primers and cDNA was amplified by PCR using primers listed in Table S12.

1100 Comparison of the predicted folding of the protein sequence of both alleles was performed using
1101 AlphaFold2[147,148] finding no appreciable alteration (Fig. S10a). *ELF8* expression levels in strains
1102 carrying either *ELF8* alleles were compared using matching published RNAseq data[53] and
1103 compared to the expression levels in the same sets of strains of *FLOWERING LOCUS C* and
1104 *GENERAL REGULATORY FACTOR11* (*GRF11*), for which a direct role for ELF8 in the
1105 establishment and maintenance of H3K4 methylation over promoter regions has been
1106 obtained[149–151] (Fig. S10b-c). *FLC* expression levels in the two sets of strains were compared to
1107 their flowering time (Fig. S10d) using published FT16 phenotypic data of an overlapping set of
1108 strains[52]. TEs that lose CHH methylation in the *jmj14 ldl1 ldl2* triple mutant were obtained from
1109 a previous study of this mutant line[79] to compare mCHH and mCG methylation levels in strains
1110 carrying either the *ELF8'* or the $ELF8^{ref}$ allele (Fig. S10e-f).

1111 **Phenotypic associations with AraPheno**

1112 417 quantitative phenotypic variables were downloaded from the AraPheno database
1113 (https://arapheno.1001genomes.org/) and for each of them we retrieved from the
1114 SNP720.MAF005.g01 vcf the genotypes at up to the 5 top significant SNPs identified in
1115 araGWAS (https://aragwas.1001genomes.org/#/). For each Z-scored arapheno variable a
1116 gaussian GLM (identity link) was built with R glm function using as explanatory variables the
1117 first three principal components of the kinship matrix as well as the genotypes at the top
1118 significant SNPs if significant in the model. After confirming that none of these SNPs
1119 correspond to a *cis*-modifier of epivariation identified by GWAS (Fig. S9a-c), each of the 490

1120 naturally epivariable TE-DMRs used for the AraClim analysis were then tested individually as a
1121 seventh categorical predictor. P values were then corrected using the Benjamini and Hochberg
1122 correction for false discovery rate using the R p.adjust function. When an epivariation was found
1123 to be significantly associated (adjusted $p < 0.05$) with multiple phenotypic variables, the one
1124 where the epivariation explains the largest fraction of phenotypic variance (McFadden adjusted
1125 R^2) was selected.

1126 **_HVA22E_ expression analysis in epiRILs**

1127 F10 offsprings of epiRILs #150 and #229 (*ddm1*-like at *AT5TE74320*) as well as of epiRILs
1128 #558 and #193 (WT-like at *AT5TE74320*) were obtained from F9 siblings propagated in bulk in
1129 standard growth conditions. F10 progeny of each epiRIL was grown on MS-media in Petri dishes
1130 in standard growth conditions (23°C) for 2 weeks. Cold-stress (CS) was performed by
1131 transferring 2-weeks old F10 seedlings at 4°C for 24 hours. Shoots of seedlings exposed (CS) or
1132 not (CTL) to the cold-stress were simultaneously collected at the end of the 24h cold-stress and
1133 flash-frozen either individually for methylation analysis or in pools of 8 seedlings for *HVA22E*
1134 expression. Methylation levels of *AT5TE74320* were measured in individual F10 seeldings using
1135 McrBC-qPCR on flash-frozen shoot tissue as described above (primers listed in Table S7)
1136 confirming that the TE-DMR is stably inherited in all F10 offsprings either in the *ddm1*-like
1137 (epiRILs 150 and 229) or WT-like (epiRILs 558 and 193) methylation states independently of
1138 the cold-stress (Fig. S13b).

1139 RNA extractions were performed on flash-frozen ground tissue (shoots) from five pools of 8 F10
1140 seedlings using Macherey-Nagel Nucleospin RNA plant kit following manufacturer's
1141 instructions. Reverse-transcriptase was performed using SuperScript IV (ThermoFisher
1142 Scientific) with oligo $d(T)_{20}$ primers and cDNA was cleaned up for qPCR using Macherey-Nagel
1143 Nucleospin PCR Cleanup. Expression levels of *HVA22E* were normalized by the average of
1144 three reference genes (*AT3G02065, AT2G41020, AT1G13320*). Primers used for qPCR are listed
1145 in Table S12.

1146 **Phenotyping of epiRILs growth rates in mild-drought or well-watered conditions**

1147 Growth and phenotyping of epiRILs on the Phenoscope (https://phenoscope.versailles.inrae.fr/)

1148 were performed as described previously[90]. Briefly, using seeds obtained from the Versailles
1149 Arabidopsis Stock Center (https://publiclines.versailles.inrae.fr/), six plants per epiRIL were
1150 grown in individual pots, three of which were maintained at 60% (control) of the maximum soil
1151 water content (SWC) and the other three at 30% (mild-drought) of this maximum. First, seeds
1152 were germinated at soil saturation (= 100% of the maximum SWC) and seedlings were set up on
1153 the robot at 8 days after sowing (DAS). Control and mild-drought SWC were stably reached and
1154 maintained from 12 and 16 DAS, respectively. Instant Relative Expansion Rates (RER instant)
1155 were calculated as the relative growth rate of the projected rosette area, integrated over 3 days
1156 windows. The projected rosette area was extracted by segmentation from daily zenithal images.

1157 **Statistical analyses**

1158 Unless specified otherwise, p-values are obtained by two-sided Wilcoxon rank sum test and
1159 statistical analyses and graphics were performed and obtained in MATLAB.

1160 Data availability

1161 Raw BS-seq, RNA-seq, sRNA-seq, and ONT-seq read-data (fastq and fasta of 20 *de novo*
1162 genome assemblies, fast5 of epiRIL238) generated and analyzed for this study are available in
1163 the European Nucleotide Archive (ENA) under project PRJEB77668.

1164 WGS data was obtained in FASTQ format for the 1047 strains of the 1001genomes.org project
1165 from NCBI SRA Project PRJNA273563. Raw RNA-seq data was obtained from NCBI GEO
1166 Accession GSE80744. Processed BS-seq data was obtained from NCBI GEO Accession
1167 GSE43857. Raw DAP-seq data was obtained from NCBI GEO Accession GSE60143.

1168 Author contributions

1169 PB and VC designed the study, supervised the work and wrote the manuscript. EC performed
1170 DNA, RNA, and sRNA extractions of epiRILs as well as ONT-sequencing of 20 strains. LDO
1171 performed analysis of WGBS-seq data, GWAS and GLM as well as de novo assembly of 20
1172 genomes using ONT-sequencing data. VS contributed to the initial analysis of the WGBS-seq
1173 data of the epiRILs. GBV performed epihaplotyping of F9 epiRILs. PB analyzed epivariation

segregation over TE-DMRs in the epiRILs and RILs. CX performed McrBC analysis of RILs. AS and PB analyzed sRNA seq data. MEM and CX performed expression and DNA methylation analysis at *HVA22E*. MBa performed single-read DNA methylation analysis using ONT. MBo generated and extracted sRNAs of *ddm1* compound mutants. ADF and LQ estimated TE family copy-numbers in the genomes of the 1001 Genomes Project.vEG and OL performed phenotyping of epiRILs using the Phenoscope. All authors read and approved the manuscript.

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. High-throughput phenotyping benefited from the support of the IJPBPlant Observatory facilities PO-PHENO and PO-VASC.

Funding

This work was supported in part by the European Union (EpiGeneSys FP7 Network of Excellence #257082, to VC), the 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). PB was funded in part by a postdoctoral fellowship from the Fondation pour la Recherche Médicale (FRM SPF20170938626). The IJPB benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007, to OL).

Declaration of interests

The authors declare no competing interests.

1196

1197

1198 **References**

1199 1. Wells, J.N., and Feschotte, C. (2020). A Field Guide to Eukaryotic Transposable Elements. Annu.
1200 Rev. Genet. *54*, 539–561.

1201 2. Ohno, S. (1972). So much "junk" DNA in our genome. Brookhaven Symp. Biol. *23*, 366–370.

1202 3. Orgel, L.E., and Crick, F.H.C. (1980). Selfish DNA: the ultimate parasite. Nature *284*, 604–607.

1203 4. Doolittle, W.F., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome
1204 evolution. Nature *284*, 601–603.

1205 5. Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements:
1206 from conflicts to benefits. Nat. Rev. Genet. *18*, 71–86.

1207 6. Rosspopoff, O., and Trono, D. (2023). Take a walk on the KRAB side: (Trends in Genetics, 39:11
1208 p:844-857, 2023). Trends Genet. https://doi.org/10.1016/j.tig.2023.12.007.

1209 7. Qiu, Y., and Köhler, C. (2020). Mobility connects: transposable elements wire new transcriptional
1210 networks by transferring transcription factor binding motifs. Biochem. Soc. Trans. *48*, 1005–1017.

1211 8. Kazazian, H.H., Jr, Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G., and Antonarakis, S.E.
1212 (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel
1213 mechanism for mutation in man. Nature *332*, 164–166.

1214 9. Bhattacharyya, M.K., Smith, A.M., Ellis, T.H., Hedley, C., and Martin, C. (1990). The wrinkled-
1215 seed character of pea described by Mendel is caused by a transposon-like insertion in a gene
1216 encoding starch-branching enzyme. Cell *60*, 115–122.

1217 10. Van't Hof, A.E., Campagne, P., Rigden, D.J., Yung, C.J., Lingley, J., Quail, M.A., Hall, N., Darby,
1218 A.C., and Saccheri, I.J. (2016). The industrial melanism mutation in British peppered moths is a
1219 transposable element. Nature *534*, 102–105.

1220 11. Kobayashi, S., Goto-Yamamoto, N., and Hirochika, H. (2004). Retrotransposon-induced mutations
1221 in grape skin color. Science *304*, 982.

1222 12. Bourgeois, Y., and Boissinot, S. (2019). On the Population Dynamics of Junk: A Review on the
1223 Population Genomics of Transposable Elements. Genes *10*, 419.

1224 13. Pulido, M., and Casacuberta, J.M. (2023). Transposable element evolution in plant genome
1225 ecosystems. Curr. Opin. Plant Biol. *75*, 102418.

1226 14. Quadrana, L., Bortolini Silveira, A., Mayhew, G.F., LeBlanc, C., Martienssen, R.A., Jeddeloh, J.A.,
1227 and Colot, V. (2016). The Arabidopsis thaliana mobilome and its impact at the species level. Elife *5*.
1228 https://doi.org/10.7554/eLife.15716.

1229 15. Baduel, P., Leduque, B., Ignace, A., Gy, I., Gil, J., Jr, Loudet, O., Colot, V., and Quadrana, L.
1230 (2021). Genetic and environmental modulation of transposition shapes the evolutionary potential of
1231 Arabidopsis thaliana. Genome Biol. *22*, 138.

1232 16. Quadrana, L., Etcheverry, M., Gilly, A., Caillieux, E., Madoui, M.-A., Guy, J., Bortolini Silveira, A.,
1233 Engelen, S., Baillet, V., Wincker, P., et al. (2019). Transposition favors the generation of large effect

1234 mutations that may facilitate rapid adaption. Nat. Commun. *10*, 3421.

1235 17. Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of
1236 the genome. Nat. Rev. Genet. *8*, 272–285.

1237 18. Zhang, H., Lang, Z., and Zhu, J.-K. (2018). Dynamics and function of DNA methylation in plants.
1238 Nat. Rev. Mol. Cell Biol. *19*, 489–506.

1239 19. Ernst, C., Odom, D.T., and Kutter, C. (2017). The emergence of piRNAs against transposon invasion
1240 to preserve mammalian genome integrity. Nat. Commun. *8*, 1411.

1241 20. Hirsch, C.D., and Springer, N.M. (2017). Transposable element influences on gene expression in
1242 plants. Biochim. Biophys. Acta Gene Regul. Mech. *1860*, 157–165.

1243 21. Choi, J.Y., and Lee, Y.C.G. (2020). Double-edged sword: The evolutionary consequences of the
1244 epigenetic silencing of transposable elements. PLoS Genet. *16*, e1008872.

1245 22. Gehring, M. (2019). Epigenetic dynamics during flowering plant reproduction: evidence for
1246 reprogramming? New Phytol. *224*, 91–96.

1247 23. He, S., and Feng, X. (2022). DNA methylation dynamics during germline development. J. Integr.
1248 Plant Biol. *64*, 2240–2251.

1249 24. Chow, H.T., and Mosher, R.A. (2023). Small RNA-mediated DNA methylation during plant
1250 reproduction. Plant Cell *35*, 1787–1800.

1251 25. Seisenberger, S., Peat, J.R., Hore, T.A., Santos, F., Dean, W., and Reik, W. (2013). Reprogramming
1252 DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. Philos.
1253 Trans. R. Soc. Lond. B Biol. Sci. *368*, 20110330.

1254 26. Greenberg, M.V.C., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian
1255 development and disease. Nat. Rev. Mol. Cell Biol. *20*, 590–607.

1256 27. Quadrana, L., and Colot, V. (2016). Plant Transgenerational Epigenetics. Annu. Rev. Genet. *50*,
1257 467–491.

1258 28. Bošković, A., and Rando, O.J. (2018). Transgenerational Epigenetic Inheritance. Annu. Rev. Genet.
1259 *52*, 21–41.

1260 29. Baduel, P., and Colot, V. (2021). The epiallelic potential of transposable elements and its
1261 evolutionary significance in plants. Philos. Trans. R. Soc. Lond. B Biol. Sci. *376*, 20200123.

1262 30. Richards, E.J. (2006). Inherited epigenetic variation — revisiting soft inheritance. Nat. Rev. Genet.
1263 *7*, 395–401.

1264 31. Ong-Abdullah, M., Ordway, J.M., Jiang, N., Ooi, S.-E., Kok, S.-Y., Sarpan, N., Azimi, N., Hashim,
1265 A.T., Ishak, Z., Rosli, S.K., et al. (2015). Loss of Karma transposon methylation underlies the
1266 mantled somaclonal variant of oil palm. Nature *525*, 533–537.

1267 32. Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and
1268 Peeters, A.J. (2000). The late flowering phenotype of fwa mutants is caused by gain-of-function
1269 epigenetic alleles of a homeodomain gene. Mol. Cell *6*, 791–802.

1270 33. Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W.J.J., Koornneef, M., and Kakutani, T.
1271 (2007). Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. Plant
1272 J. *49*, 38–45.

1273 34. Lippman, Z., Gendrel, A.-V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K.,
1274 Mittal, V., May, B., Kasschau, K.D., et al. (2004). Role of transposable elements in heterochromatin
1275 and epigenetic control. Nature *430*, 471–476.

1276 35. Bertozzi, T.M., and Ferguson-Smith, A.C. (2020). Metastable epialleles and their contribution to
1277 epigenetic inheritance in mammals. Semin. Cell Dev. Biol. *97*, 93–105.

1278 36. Heard, E., and Martienssen, R.A. (2014). Transgenerational epigenetic inheritance: myths and
1279 mechanisms. Cell *157*, 95–109.

1280 37. Vongs, A., Kakutani, T., Martienssen, R.A., and Richards, E.J. (1993). Arabidopsis thaliana DNA
1281 methylation mutants. Science *260*, 1926–1928.

1282 38. Saze, H., Mittelsten Scheid, O., and Paszkowski, J. (2003). Maintenance of CpG methylation is
1283 essential for epigenetic inheritance during plant gametogenesis. Nat. Genet. *34*, 65–69.

1284 39. Kankel, M.W., Ramsey, D.E., Stokes, T.L., Flowers, S.K., Haag, J.R., Jeddeloh, J.A., Riddle, N.C.,
1285 Verbsky, M.L., and Richards, E.J. (2003). Arabidopsis MET1 cytosine methyltransferase mutants.
1286 Genetics *163*, 1109–1122.

1287 40. Li, Q., Eichten, S.R., Hermanson, P.J., Zaunbrecher, V.M., Song, J., Wendt, J., Rosenbaum, H.,
1288 Madzima, T.F., Sloan, A.E., Huang, J., et al. (2014). Genetic perturbation of the maize methylome.
1289 Plant Cell *26*, 4602–4616.

1290 41. Corem, S., Doron-Faigenboim, A., Jouffroy, O., Maumus, F., Arazi, T., and Bouché, N. (2018).
1291 Redistribution of CHH Methylation and Small Interfering RNAs across the Genome of Tomato
1292 ddm1 Mutants. Plant Cell *30*, 1628–1644.

1293 42. Xu, G., and Law, J.A. (2024). Loops, crosstalk, and compartmentalization: it takes many layers to
1294 regulate DNA methylation. Curr. Opin. Genet. Dev. *84*, 102147.

1295 43. He, L., Huang, H., Bradai, M., Zhao, C., You, Y., Ma, J., Zhao, L., Lozano-Durán, R., and Zhu, J.-K.
1296 (2022). DNA methylation-free Arabidopsis reveals crucial roles of DNA methylation in regulating
1297 gene expression and development. Nat. Commun. *13*, 1335.

1298 44. Lyons, D.B., Briffa, A., He, S., Choi, J., Hollwey, E., Colicchio, J., Anderson, I., Feng, X., Howard,
1299 M., and Zilberman, D. (2023). Extensive de novo activity stabilizes epigenetic inheritance of CG
1300 methylation in Arabidopsis transposons. Cell Rep. *42*, 112132.

1301 45. Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer,
1302 S.L., and Zilberman, D. (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA
1303 methyltransferases to access H1-containing heterochromatin. Cell *153*, 193–205.

1304 46. Jeddeloh, J.A., Stokes, T.L., and Richards, E.J. (1999). Maintenance of genomic methylation
1305 requires a SWI2/SNF2-like protein. Nat. Genet. *22*, 94–97.

1306 47. Lee, S.C., Adams, D.W., Ipsaro, J.J., Cahn, J., Lynn, J., Kim, H.-S., Berube, B., Major, V., Calarco,
1307 J.P., LeBlanc, C., et al. (2023). Chromatin remodeling of histone H3 variants by DDM1 underlies
1308 epigenetic inheritance of DNA methylation. Cell *186*, 4100–4116.e15.

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

49. Colomé-Tatché, M., Cortijo, S., Wardenaar, R., Morgado, L., Lahouze, B., Sarazin, A., Etcheverry, M., Martin, A., Feng, S., Duvernois-Berthet, E., et al. (2012). 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.

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

51. Teixeira, F.K., Heredia, F., Sarazin, A., Roudier, F., Boccara, M., Ciaudo, C., Cruaud, C., Poulain, J., Berdasco, M., Fraga, M.F., et al. (2009). A Role for RNAi in the Selective Correction of DNA Methylation Defects. Preprint, https://doi.org/10.1126/science.1165313 https://doi.org/10.1126/science.1165313.

52. 1001 Genomes Consortium. Electronic address: magnus.nordborg@gmi.oeaw.ac.at, and 1001 Genomes Consortium (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell *166*, 481–491.

53. Kawakatsu, T., Huang, S.-S.C., Jupe, F., Sasaki, E., Schmitz, R.J., Urich, M.A., Castanon, R., Nery, J.R., Barragan, C., He, Y., et al. (2016). Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. Cell *166*, 492–505.

54. Togninalli, M., Seren, Ü., Meng, D., Fitz, J., Nordborg, M., Weigel, D., Borgwardt, K., Korte, A., and Grimm, D.G. (2018). The AraGWAS Catalog: a curated and standardized Arabidopsis thaliana GWAS catalog. Nucleic Acids Res. *46*, D1150–D1156.

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

56. Hosaka, A., Saito, R., Takashima, K., Sasaki, T., Fu, Y., Kawabe, A., Ito, T., Toyoda, A., Fujiyama, A., Tarutani, Y., et al. (2017). Evolution of sequence-specific anti-silencing systems in Arabidopsis. Nat. Commun. *8*, 2161.

57. Sasaki, T., Ro, K., Caillieux, E., Manabe, R., Bohl-Viallefond, G., Baduel, P., Colot, V., Kakutani, T., and Quadrana, L. (2022). Fast co-evolution of anti-silencing systems shapes the invasiveness of Mu-like DNA transposons in eudicots. EMBO J. *41*, e110070.

58. Zhang, Y., Jake Harris, C., Liu, Q., Liu, W., Ausin, I., Long, Y., Xiao, L., Feng, L., Chen, X., Xie, Y., et al. (2018). Large-scale comparative epigenomics reveals hierarchical regulation of non-CG methylation in Arabidopsis. Preprint, https://doi.org/10.1073/pnas.1716300115 https://doi.org/10.1073/pnas.1716300115.

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

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

1348 61. Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of
1349 increasing complexity. Preprint, https://doi.org/10.1038/nrg3683 https://doi.org/10.1038/nrg3683.

1350 62. Wendte, J.M., and Pikaard, C.S. (2017). The RNAs of RNA-directed DNA methylation. Preprint,
1351 https://doi.org/10.1016/j.bbagrm.2016.08.004 https://doi.org/10.1016/j.bbagrm.2016.08.004.

1352 63. Osabe, K., Harukawa, Y., Miura, S., and Saze, H. (2017). Epigenetic Regulation of Intronic
1353 Transgenes in Arabidopsis. Sci. Rep. *7*, 1–13.

1354 64. Martins, L.M., and Law, J.A. (2023). Moving targets: Mechanisms regulating siRNA production and
1355 DNA methylation during plant development. Curr. Opin. Plant Biol. *75*, 102435.

1356 65. Dalakouras, A., Dadami, E., Wassenegger, M., Krczal, G., and Wassenegger, M. (2016). RNA-
1357 directed DNA methylation efficiency depends on trigger and target sequence identity. Plant J. *87*,
1358 202–214.

1359 66. Le, N.T., Harukawa, Y., Miura, S., Boer, D., Kawabe, A., and Saze, H. (2020). Epigenetic regulation
1360 of spurious transcription initiation in Arabidopsis. Nat. Commun. *11*, 3224.

1361 67. Li, J., Yang, D.-L., Huang, H., Zhang, G., He, L., Pang, J., Lozano-Durán, R., Lang, Z., and Zhu, J.-
1362 K. (2020). Epigenetic memory marks determine epiallele stability at loci targeted by de novo DNA
1363 methylation. Nat Plants *6*, 661–674.

1364 68. Bente, H., Foerster, A.M., Lettner, N., and Mittelsten Scheid, O. (2021). Polyploidy-associated
1365 paramutation in Arabidopsis is determined by small RNAs, temperature, and allele structure. PLoS
1366 Genet. *17*, e1009444.

1367 69. Zhang, Y., Jang, H., Xiao, R., Kakoulidou, I., Piecyk, R.S., Johannes, F., and Schmitz, R.J. (2021).
1368 Heterochromatin is a quantitative trait associated with spontaneous epiallele formation. Nat.
1369 Commun. *12*, 6958.

1370 70. Stuart, T., Eichten, S.R., Cahn, J., Karpievitch, Y.V., Borevitz, J.O., and Lister, R. (2016).
1371 Population scale mapping of transposable element diversity reveals links to gene regulation and
1372 epigenomic variation. Elife *5*. https://doi.org/10.7554/eLife.20777.

1373 71. Maumus, F., and Quesneville, H. (2014). Ancestral repeats have shaped epigenome and genome
1374 composition for millions of years in Arabidopsis thaliana. Nat. Commun. *5*, 1–9.

1375 72. Vaughn, M.W., Tanurdzić, M., Lippman, Z., Jiang, H., Carrasquillo, R., Rabinowicz, P.D., Dedhia,
1376 N., McCombie, W.R., Agier, N., Bulski, A., et al. (2007). Epigenetic natural variation in Arabidopsis
1377 thaliana. PLoS Biol. *5*, e174.

1378 73. Jaegle, B., Pisupati, R., Soto-Jiménez, L.M., Burns, R., Rabanal, F.A., and Nordborg, M. (2023).
1379 Extensive sequence duplication in Arabidopsis revealed by pseudo-heterozygosity. Genome Biol. *24*,
1380 1–19.

1381 74. Simon, M., Loudet, O., Durand, S., Bérard, A., Brunel, D., Sennesal, F.-X., Durand-Tardif, M.,
1382 Pelletier, G., and Camilleri, C. (2008). Quantitative trait loci mapping in five new large recombinant
1383 inbred line populations of Arabidopsis thaliana genotyped with consensus single-nucleotide
1384 polymorphism markers. Genetics *178*, 2253–2264.

1385 75. Picard, C.L., and Gehring, M. (2017). Proximal methylation features associated with nonrandom
1386 changes in gene body methylation. Genome Biol. *18*, 1–18.

1387 76. Van Oss, S.B., Cucinotta, C.E., and Arndt, K.M. (2017). Emerging Insights into the Roles of the
1388 Paf1 Complex in Gene Regulation. Trends Biochem. Sci. *42*, 788–798.

1389 77. Mori, S., Oya, S., Takahashi, M., Takashima, K., Inagaki, S., and Kakutani, T. (2023).
1390 Cotranscriptional demethylation induces global loss of H3K4me2 from active genes in Arabidopsis.
1391 EMBO J. *42*, e113798.

1392 78. Oh, S., Park, S., and van Nocker, S. (2008). Genic and global functions for Paf1C in chromatin
1393 modification and gene expression in Arabidopsis. PLoS Genet. *4*, e1000077.

1394 79. Greenberg, M.V.C., Deleris, A., Hale, C.J., Liu, A., Feng, S., and Jacobsen, S.E. (2013). Interplay
1395 between active chromatin marks and RNA-directed DNA methylation in Arabidopsis thaliana. PLoS
1396 Genet. *9*, e1003946.

1397 80. Sasaki, E., Kawakatsu, T., Ecker, J.R., and Nordborg, M. (2019). Common alleles of CMT2 and
1398 NRPE1 are major determinants of CHH methylation variation in Arabidopsis thaliana. PLoS Genet.
1399 *15*, e1008492.

1400 81. Deans, N.C., Talbot, J.-E.R.B., Li, M., Sáez-González, C., Hövel, I., Heavens, D., Stam, M., and
1401 Hollick, J.B. (2024). Paramutation at the maize pl1 locus is associated with RdDM activity at distal
1402 tandem repeats. PLoS Genet. *20*, e1011296.

1403 82. Chandler, V.L., Eggleston, W.B., and Dorweiler, J.E. (2000). Paramutation in maize. Plant Mol.
1404 Biol. *43*, 121–145.

1405 83. Silveira, A.B., Trontin, C., Cortijo, S., Barau, J., Del Bem, L.E.V., Loudet, O., Colot, V., and
1406 Vincentz, M. (2013). Extensive natural epigenetic variation at a de novo originated gene. PLoS
1407 Genet. *9*, e1003437.

1408 84. Berthelier, J., Furci, L., Asai, S., Sadykova, M., Shimazaki, T., Shirasu, K., and Saze, H. (2023).
1409 Long-read direct RNA sequencing reveals epigenetic regulation of chimeric gene-transposon
1410 transcripts in Arabidopsis thaliana. Nat. Commun. *14*, 3248.

1411 85. Li, Z.-W., Chen, X., Wu, Q., Hagmann, J., Han, T.-S., Zou, Y.-P., Ge, S., and Guo, Y.-L. (2016). On
1412 the Origin of De Novo Genes in Arabidopsis thaliana Populations. Genome Biol. Evol. *8*, 2190–
1413 2202.

1414 86. Wang, J., Marowsky, N.C., and Fan, C. (2013). Divergent evolutionary and expression patterns
1415 between lineage specific new duplicate genes and their parental paralogs in Arabidopsis thaliana.
1416 PLoS One *8*, e72362.

1417 87. Donoghue, M.T., Keshavaiah, C., Swamidatta, S.H., and Spillane, C. (2011). Evolutionary origins of
1418 Brassicaceae specific genes in Arabidopsis thaliana. BMC Evol. Biol. *11*, 47.

1419 88. Ghimire, B., Saminathan, T., Bodunrin, A., Abburi, V.L., Kshetry, A.O., Shinde, S., Nimmakayala,
1420 P., and Reddy, U.K. (2021). Genome-Wide Association Study of Natural Variation in Arabidopsis
1421 Exposed to Acid Mine Drainage Toxicity and Validation of Associated Genes with Reverse
1422 Genetics. Plants *10*, 191.

1423 89. Prakash, J., Agrawal, S.B., and Agrawal, M. (2023). Global Trends of Acidity in Rainfall and Its
1424 Impact on Plants and Soil. J Soil Sci Plant Nutr *23*, 398–419.

1425 90. Tisné, S., Serrand, Y., Bach, L., Gilbault, E., Ben Ameur, R., Balasse, H., Voisin, R., Bouchez, D.,

1426 Durand-Tardif, M., Guerche, P., et al. (2013). Phenoscope: an automated large-scale phenotyping
1427 platform offering high spatial homogeneity. Plant J. *74*, 534–544.

1428 91. Kakoulidou, I., Piecyk, R.S., Meyer, R.C., Kuhlmann, M., Gutjahr, C., Altmann, T., and Johannes,
1429 F. (2024). Mapping parental DMRs predictive of local and distal methylome remodeling in
1430 epigenetic F1 hybrids. Life Sci Alliance *7*. https://doi.org/10.26508/lsa.202402599.

1431 92. Zhong, X., Hale, C.J., Law, J.A., Johnson, L.M., Feng, S., Tu, A., and Jacobsen, S.E. (2012). DDR
1432 complex facilitates global association of RNA polymerase V to promoters and evolutionarily young
1433 transposons. Nat. Struct. Mol. Biol. *19*, 870–875.

1434 93. Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014).
1435 Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat. Struct. Mol. Biol.
1436 *21*, 64–72.

1437 94. McCue, A.D., Nuthikattu, S., and Slotkin, R.K. (2013). Genome-wide identification of genes
1438 regulated in trans by transposable element small interfering RNAs. RNA Biol. *10*, 1379–1395.

1439 95. Li, S., Vandivier, L.E., Tu, B., Gao, L., Won, S.Y., Li, S., Zheng, B., Gregory, B.D., and Chen, X.
1440 (2015). Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in Arabidopsis reveals
1441 features and regulation of siRNA biogenesis. Genome Res. *25*, 235–245.

1442 96. Martienssen, R. (1996). Epigenetic phenomena: paramutation and gene silencing in plants. Curr.
1443 Biol. *6*, 810–813.

1444 97. Hollick, J.B., Dorweiler, J.E., and Chandler, V.L. (1997). Paramutation and related allelic
1445 interactions. Trends Genet. *13*, 302–308.

1446 98. Settles, A.M., Baron, A., Barkan, A., and Martienssen, R.A. (2001). Duplication and suppression of
1447 chloroplast protein translocation genes in maize. Genetics *157*, 349–360.

1448 99. Hollick, J.B. (2017). Paramutation and related phenomena in diverse species. Nat. Rev. Genet. *18*,
1449 5–23.

1450 100. Mittelsten Scheid, O., Afsar, K., and Paszkowski, J. (2003). Formation of stable epialleles and their
1451 paramutation-like interaction in tetraploid Arabidopsis thaliana. Nat. Genet. *34*, 450–454.

1452 101. Greaves, I.K., Groszmann, M., Ying, H., Taylor, J.M., Peacock, W.J., and Dennis, E.S. (2012).
1453 Trans chromosomal methylation in Arabidopsis hybrids. Proc. Natl. Acad. Sci. U. S. A. *109*, 3570–
1454 3575.

1455 102. Kowalik, K.M., Shimada, Y., Flury, V., Stadler, M.B., Batki, J., and Bühler, M. (2015). The Paf1
1456 complex represses small-RNA-mediated epigenetic gene silencing. Nature *520*, 248–252.

1457 103. Duempelmann, L., Mohn, F., Shimada, Y., Oberti, D., Andriollo, A., Lochs, S., and Bühler, M.
1458 (2019). Inheritance of a Phenotypically Neutral Epimutation Evokes Gene Silencing in Later
1459 Generations. Mol. Cell *74*, 534–541.e4.

1460 104. Jiang, C., Mithani, A., Belfield, E.J., Mott, R., Hurst, L.D., and Harberd, N.P. (2014).
1461 Environmentally responsive genome-wide accumulation of de novo Arabidopsis thaliana mutations
1462 and epimutations. Genome Res. *24*, 1821–1829.

1463 105. Yao, N., Zhang, Z., Yu, L., Hazarika, R., Yu, C., Jang, H., Smith, L.M., Ton, J., Liu, L., Stachowicz,

1464 J.J., et al. (2023). An evolutionary epigenetic clock in plants. Science *381*, 1440–1445.

1465 106. Ganguly, D.R., Crisp, P.A., Eichten, S.R., and Pogson, B.J. (2017). The Arabidopsis DNA
1466 Methylome Is Stable under Transgenerational Drought Stress. Plant Physiol. *175*, 1893–1912.

1467 107. Wibowo, A., Becker, C., Marconi, G., Durr, J., Price, J., Hagmann, J., Papareddy, R., Putra, H.,
1468 Kageyama, J., Becker, J., et al. (2016). Hyperosmotic stress memory in Arabidopsis is mediated by
1469 distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA
1470 glycosylase activity. Elife *5*. https://doi.org/10.7554/eLife.13546.

1471 108. Van Dooren, T.J.M., Silveira, A.B., Gilbault, E., Jiménez-Gómez, J.M., Martin, A., Bach, L., Tisné,
1472 S., Quadrana, L., Loudet, O., and Colot, V. (2020). Mild drought in the vegetative stage induces
1473 phenotypic, gene expression, and DNA methylation plasticity in Arabidopsis but no
1474 transgenerational effects. J. Exp. Bot. *71*, 3588–3602.

1475 109. Raingeval, M., Leduque, B., Baduel, P., Edera, A., Roux, F., Colot, V., and Quadrana, L. (2023).
1476 Retrotransposon-driven environmental regulation of*FLC*leads to adaptive response to herbicide.
1477 bioRxiv. https://doi.org/10.1101/2023.09.06.556499.

1478 110. Yu, A., Lepère, G., Jay, F., Wang, J., Bapaume, L., Wang, Y., Abraham, A.-L., Penterman, J.,
1479 Fischer, R.L., Voinnet, O., et al. (2013). Dynamics and biological relevance of DNA demethylation
1480 in Arabidopsis antibacterial defense. Proc. Natl. Acad. Sci. U. S. A. *110*, 2389–2394.

1481 111. Kazachenka, A., Bertozzi, T.M., Sjoberg-Herrera, M.K., Walker, N., Gardner, J., Gunning, R.,
1482 Pahita, E., Adams, S., Adams, D., and Ferguson-Smith, A.C. (2018). Identification, Characterization,
1483 and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. Cell *175*,
1484 1259–1271.e13.

1485 112. Xu, J., Chen, G., Hermanson, P.J., Xu, Q., Sun, C., Chen, W., Kan, Q., Li, M., Crisp, P.A., Yan, J.,
1486 et al. (2019). Population-level analysis reveals the widespread occurrence and phenotypic
1487 consequence of DNA methylation variation not tagged by genetic variation in maize. Genome Biol.
1488 *20*, 243.

1489 113. Eichten, S.R., Briskine, R., Song, J., Li, Q., Swanson-Wagner, R., Hermanson, P.J., Waters, A.J.,
1490 Starr, E., West, P.T., Tiffin, P., et al. (2013). Epigenetic and genetic influences on DNA methylation
1491 variation in maize populations. Plant Cell *25*, 2783–2797.

1492 114. Takahashi, Y., Morales Valencia, M., Yu, Y., Ouchi, Y., Takahashi, K., Shokhirev, M.N., Lande, K.,
1493 Williams, A.E., Fresia, C., Kurita, M., et al. (2023). Transgenerational inheritance of acquired
1494 epigenetic signatures at CpG islands in mice. Cell *186*, 715–731.e19.

1495 115. Ciabrelli, F., Comoglio, F., Fellous, S., Bonev, B., Ninova, M., Szabo, Q., Xuéreb, A., Klopp, C.,
1496 Aravin, A., Paro, R., et al. (2017). Stable Polycomb-dependent transgenerational inheritance of
1497 chromatin states in Drosophila. Nat. Genet. *49*, 876–886.

1498 116. Greer, E.L., Maures, T.J., Ucar, D., Hauswirth, A.G., Mancini, E., Lim, J.P., Benayoun, B.A., Shi,
1499 Y., and Brunet, A. (2011). Transgenerational epigenetic inheritance of longevity in Caenorhabditis
1500 elegans. Nature *479*, 365–371.

1501 117. Schwartz-Orbach, L., Zhang, C., Sidoli, S., Amin, R., Kaur, D., Zhebrun, A., Ni, J., and Gu, S.G.
1502 (2020). Caenorhabditis elegans nuclear RNAi factor SET-32 deposits the transgenerational histone
1503 modification, H3K23me3. Elife *9*. https://doi.org/10.7554/eLife.54309.

1504 118. Klosin, A., Casas, E., Hidalgo-Carcedo, C., Vavouri, T., and Lehner, B. (2017). Transgenerational
1505 transmission of environmental information in C. elegans. Science *356*, 320–323.

1506 119. Torres-Garcia, S., Yaseen, I., Shukla, M., Audergon, P.N.C.B., White, S.A., Pidoux, A.L., and
1507 Allshire, R.C. (2020). Epigenetic gene silencing by heterochromatin primes fungal resistance. Nature
1508 *585*, 453–458.

1509 120. Yaseen, I., White, S.A., Torres-Garcia, S., Spanos, C., Lafos, M., Gaberdiel, E., Yeboah, R., El
1510 Karoui, M., Rappsilber, J., Pidoux, A.L., et al. (2022). Proteasome-dependent truncation of the
1511 negative heterochromatin regulator Epe1 mediates antifungal resistance. Nat. Struct. Mol. Biol. *29*,
1512 745–758.

1513 121. He, L., Wu, W., Zinta, G., Yang, L., Wang, D., Liu, R., Zhang, H., Zheng, Z., Huang, H., Zhang, Q.,
1514 et al. (2018). A naturally occurring epiallele associates with leaf senescence and local climate
1515 adaptation in Arabidopsis accessions. Nat. Commun. *9*, 460.

1516 122. Ahmed, I., Sarazin, A., Bowler, C., Colot, V., and Quesneville, H. (2011). Genome-wide evidence
1517 for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. Nucleic
1518 Acids Res. *39*, 6919–6931.

1519 123. Liu, J., and Zhong, X. (2024). Epiallelic variation of non-coding RNA genes and their phenotypic
1520 consequences. Nat. Commun. *15*, 1375.

1521 124. Kornienko, A.E., Nizhynska, V., Molla Morales, A., Pisupati, R., and Nordborg, M. (2023).
1522 Population-level annotation of lncRNAs in Arabidopsis reveals extensive expression variation
1523 associated with transposable element-like silencing. Plant Cell *36*, 85–111.

1524 125. Shahzad, Z., Moore, J.D., Choi, J., and Zilberman, D. (2021). Epigenetic inheritance mediates
1525 phenotypic diversity in natural populations. bioRxiv, 2021.03.15.435374.
1526 https://doi.org/10.1101/2021.03.15.435374.

1527 126. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for
1528 Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

1529 127. Giraut, L., Falque, M., Drouaud, J., Pereira, L., Martin, O.C., and Mézard, C. (2011). Genome-wide
1530 crossover distribution in Arabidopsis thaliana meiosis reveals sex-specific patterns along
1531 chromosomes. PLoS Genet. *7*, e1002354.

1532 128. Sun, Z., Li, H., Zhang, L., and Wang, J. (2012). Estimation of recombination frequency in bi-
1533 parental genetic populations. Genet. Res. *94*, 163–177.

1534 129. Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C., Langlade, N., and
1535 Muños, S. (2016). Extraction of high-molecular-weight genomic DNA for long-read sequencing of
1536 single molecules. Biotechniques *61*, 203–205.

1537 130. Ramawatar, N.A., and Schwessinger, B. (2018). DNA size selection (>3-4kb) and purification of
1538 DNA using an improved homemade SPRI beads solution. v1 (ZappyLab, Inc.)
1539 https://doi.org/10.17504/protocols.io.n7hdhj6.

1540 131. Ni, P., Huang, N., Nie, F., Zhang, J., Zhang, Z., Wu, B., Bai, L., Liu, W., Xiao, C.-L., Luo, F., et al.
1541 (2021). Genome-wide detection of cytosine methylations in plant from Nanopore data using deep
1542 learning. Nat. Commun. *12*, 5976.

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

133. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

134. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

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

136. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

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

138. Khanna, A., Larson, D.E., Srivatsan, S.N., Mosior, M., Abbott, T.E., Kiwala, S., Ley, T.J., Duncavage, E.J., Walter, M.J., Walker, J.R., et al. (2022). Bam-readcount - rapid generation of basepair-resolution sequence metrics. Journal of Open Source Software *7*, 3722.

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

140. Hu, J., Wang, Z., Sun, Z., Hu, B., Ayoola, A.O., Liang, F., Li, J., Sandoval, J.R., Cooper, D.N., Ye, K., et al. (2024). NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. Genome Biol. *25*, 107.

141. Hu, J., Fan, J., Sun, Z., and Liu, S. (2020). NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics *36*, 2253–2255.

142. Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z.B., Schatz, M.C., and Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. *23*, 258.

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

144. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

145. Browning, S.R., Browning, B.L. (2007). 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.

146. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

147. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with

1581 AlphaFold. Nature *596*, 583–589.

1582 148. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022).
1583 ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682.

1584 149. He, Y., Doyle, M.R., and Amasino, R.M. (2004). PAF1-complex-mediated histone methylation of
1585 FLOWERING LOCUS C chromatin is required for the vernalization-responsive, winter-annual habit
1586 in Arabidopsis. Genes Dev. *18*, 2774–2784.

1587 150. Oh, S., Zhang, H., Ludwig, P., and van Nocker, S. (2004). A mechanism related to the yeast
1588 transcriptional regulator Paf1c is required for expression of the Arabidopsis FLC/MAF MADS box
1589 gene family. Plant Cell *16*, 2940–2953.

1590 151. Singh, S., Kailasam, S., Lo, J.-C., and Yeh, K.-C. (2021). Histone H3 lysine4 trimethylation-
1591 regulated GRF11 expression is essential for the iron-deficiency response in Arabidopsis thaliana.
1592 New Phytol. *230*, 244–258.

1593