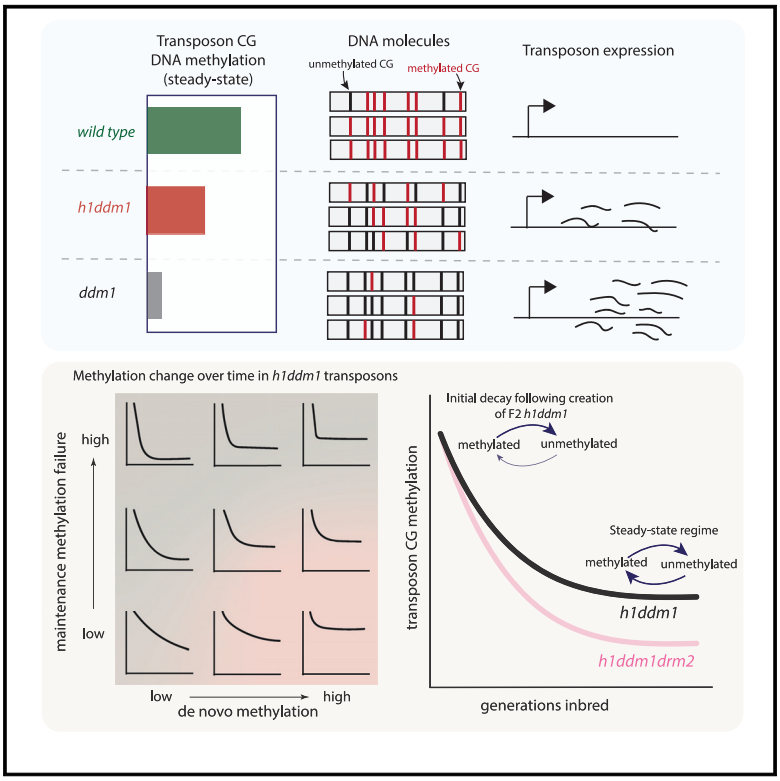
Cell Reports

Article

Extensive *de novo* activity stabilizes epigenetic inheritance of CG methylation in *Arabidopsis* transposons

Graphical abstract



Authors

David B. Lyons, Amy Briffa, Shengbo He, ..., Xiaoqi Feng, Martin Howard, Daniel Zilberman

Correspondence

martin.howard@jic.ac.uk (M.H.), daniel.zilberman@ist.ac.at (D.Z.)

In brief

By inbreeding DNA methylation-deficient *Arabidopsis thaliana* plants for many generations, Lyons et al. find that epigenetic inheritance of heterochromatic CG methylation relies on multiple methyltransferases and, using mathematical modeling, demonstrate that the inheritance process is much more dynamic and flexible than previously thought.

Highlights

- Stable epigenetic inheritance of intermediate CG methylation in heterochromatin
- Quantitative association of CG methylation with expression of transposable elements
- Mathematical modeling reveals extensive heterochromatic *de novo* CG methylation
- Heterochromatic *de novo* CG methylation mainly mediated by the RdDM pathway

Lyons et al., 2023, Cell Reports *42*, 112132
March 28, 2023
https://doi.org/10.1016/j.celrep.2023.112132

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Extensive *de novo* activity stabilizes epigenetic inheritance of CG methylation in *Arabidopsis* transposons

David B. Lyons,[1,4] Amy Briffa,[1,4] Shengbo He,[1] Jaemyung Choi,[1] Elizabeth Hollwey,[1,2] Jack Colicchio,[3] Ian Anderson,[3] Xiaoqi Feng,[1,2] Martin Howard,[1,*] and Daniel Zilberman[1,2,5,*]

[1]John Innes Centre, Norwich NR4 7UH, UK
[2]Institute of Science and Technology, 3400 Klosterneuburg, Austria
[3]Department of Plant & Microbial Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[4]These authors contributed equally
[5]Lead contact
*Correspondence: martin.howard@jic.ac.uk (M.H.), daniel.zilberman@ist.ac.at (D.Z.)
https://doi.org/10.1016/j.celrep.2023.112132

SUMMARY

Cytosine methylation within CG dinucleotides (mCG) can be epigenetically inherited over many generations. Such inheritance is thought to be mediated by a semiconservative mechanism that produces binary present/absent methylation patterns. However, we show here that, in *Arabidopsis thaliana h1ddm1* mutants, intermediate heterochromatic mCG is stably inherited across many generations and is quantitatively associated with transposon expression. We develop a mathematical model that estimates the rates of semiconservative maintenance failure and *de novo* methylation at each transposon, demonstrating that mCG can be stably inherited at any level via a dynamic balance of these activities. We find that DRM2—the core methyltransferase of the RNA-directed DNA methylation pathway—catalyzes most of the heterochromatic *de novo* mCG, with *de novo* rates orders of magnitude higher than previously thought, whereas chromomethylases make smaller contributions. Our results demonstrate that stable epigenetic inheritance of mCG in plant heterochromatin is enabled by extensive *de novo* methylation.

INTRODUCTION

Cytosine methylation provides a mechanism to heritably alter the genome without permanent modification of the DNA sequence.[1] DNA methylation represses transposable element (TE) transcription and transposition in plants and vertebrates.[2] Methylation also regulates endogenous genes: methylation close to the transcriptional start site generally causes gene silencing,[3] whereas methylation of other genic regions can promote or counteract expression.[4] Genetic defects in the methylation machinery lead to human disease such as cancer.[5] Disruption of methylation patterns during clonal propagation of plants causes developmental defects that hamper agriculture, and methylation patterns within genes account for a substantial fraction of phenotypic variation in natural *Arabidopsis thaliana* populations.[6–8] Faithful propagation of DNA methylation is clearly essential, yet our understanding of the underlying processes is far from complete.

Across eukaryotes, cytosine methylation is most prevalent within CG dinucleotides.[9] The core model for the epigenetic inheritance of CG methylation (mCG)—proposed over 40 years ago[10,11]—was inspired by the symmetry of the CG site, and posits that the methylation status of the old strand is used to reproduce the pattern on the strand synthesized during DNA replication. This elegant, semiconservative model has strong experimental support[12–14] but contains a potential flaw because it lacks a mechanism to recover DNA methylation following maintenance failure.[15]

Two classes of models have been proposed to address the above issue. The first, developed most explicitly with data from the fungus *Cryptococcus neoformans*, proposes that very high fidelity semiconservative maintenance (estimated failure rate of 9.3×10^{-5} per CG site per cell cycle) combined with rare, random, and potentially non-enzymatic *de novo* methylation and natural selection can produce stable epigenetic inheritance of mCG over million-year timescales.[14] The second, developed with mammalian data, de-emphasizes the semiconservative mechanism, and instead proposes a combination of inefficient maintenance (e.g., failure rate of 8.0% per site per cell cycle)[16] and high rates of untemplated *de novo* methylation (e.g., *de novo* rate of 4.7% per site per cell cycle)[16] that is specifically targeted to methylated regions by an unknown mechanism.[16–18] The maintenance methyltransferases in the two systems are different (Dnmt5 in *C. neoformans* and Dnmt1 in mammals),[19–21] which might account for the distinct mechanisms. The overall lack of long-term, transgenerational epigenetic inheritance of mammalian mCG[22–24] may also be compatible with higher rates of maintenance failure.

The existing plant data (mostly from *Arabidopsis*) appear to be more compatible with the former mechanism. Plants exhibit stable, transgenerational mCG inheritance,[25–27] and although plants use MET1 (a Dnmt1 family enzyme),[21] the global maintenance failure rate of 6.3×10^{-4} per generation[28] or about 1.9×10^{-5} per cell cycle (assuming 34 cell cycles per generation[29]) reported for *Arabidopsis* is lower than that reported for *C. neoformans* (9.3×10^{-5}).[14] Even the higher maintenance failure rate of 4.4×10^{-5} per cell cycle (1.5×10^{-3} per generation) reported for *Arabidopsis* genes[28] is two times lower than the *C. neoformans* rate. Epigenetic inheritance of plant mCG is thought to depend only on MET1[2,30,31] which, like *C. neoformans* Dnmt5, does not have known *de novo* activity.[2] The reported rates of *de novo* mCG are also low in *Arabidopsis* (global rate of 2.6×10^{-4} per generation or 7.6×10^{-6} per cell cycle),[28] and the source of the *de novo* activity is unknown.

Plant and animal genomes also contain cytosine methylation outside CG dinucleotides.[32] In plants, the CMT3 methyltransferase family catalyzes methylation of CNG trinucleotides, conventionally described as CHG (where H is any non-G base) to avoid overlapping CG sites.[33] The related CMT2 family can methylate cytosines outside CG and CNG contexts,[34,35] a pattern of specificity referred to as CHH. CMT2 and CMT3 rely on a positive feedback loop with dimethylation of lysine 9 of histone H3 (H3K9me2)[32,36,37]—a hallmark of heterochromatin[38]—and therefore preferentially methylate heterochromatic TEs.[34,35,39] Plants also possess an RNA-directed DNA methylation (RdDM) pathway, in which 24 nucleotide RNA molecules guide DRM methyltransferases (homologs of animal Dnmt3) to initiate DNA methylation in all sequence contexts and to maintain CHH methylation at relatively euchromatic TEs.[40]

Regardless of the mechanism, plants do not perfectly copy non-CG methylation each cell division, which leads to a probabilistic distribution of methylation states at CHG and CHH sites that contrasts with the more binary mCG patterns.[30,31] Although the RdDM pathway can establish DNA methylation in every context, including CG,[41–44] several studies have concluded that mCG maintenance is independent of the non-CG methylation pathways.[31,35,45] The *de novo* mCG rates reported for TEs (where the non-CG pathways operate) are also lower than those for genes,[28] ranging down to 6.8×10^{-8} per cell cycle (2.3×10^{-6} per generation).[46] Thus, inheritance of mCG within TEs is thought to be essentially semiconservative.

Previously, we described widespread intermediate mCG within heterochromatic TEs of *Arabidopsis h1dddm1* mutants lacking the nucleosome remodeler DDM1 and linker histone H1.[34,47] Lack of DDM1 is thought to inhibit MET1 access to heterochromatin, whereas lack of H1 is thought to increase MET1 access, especially to the linker DNA that separates nucleosomes.[34,47] Lack of H1 also causes increased RdDM activity within heterochromatic TEs of *h1ddm1* plants.[48] The reduced mCG maintenance efficiency in *h1ddm1* mutants combined with the reported low *de novo* rates in TEs[28,46] imply that the intermediate mCG patterns of *h1ddm1* plants should be unstable and eventually degrade to very low levels, but this has not been evaluated.

Here, we analyze *h1ddm1* DNA methylation over ten generations of inbreeding. We find that, contrary to the above expectation, intermediate mCG patterns are stably inherited. Using a mathematical model, we find a predicted *de novo* component of mCG inheritance orders of magnitude stronger than implied by the previously reported transgenerational rates of mCG gain in TEs. This prediction is confirmed by analysis of *h1ddm1* compound mutants lacking either CMT2, CMT3, or the key RdDM methyltransferase DRM2. Our results contradict the established view that mCG epigenetic inheritance within TEs is essentially semiconservative. Instead, our data indicate that epigenetic inheritance of heterochromatic mCG is stabilized by extensive *de novo* methylation, with RdDM contributing most of the *de novo* activity and CMT2/3 making smaller contributions.

RESULTS

TE CG methylation decays to stable intermediate levels within ten *h1ddm1* generations

To understand DNA methylation inheritance in *h1ddm1* plants, we crossed a plant homozygous for mutations in both canonical *Arabidopsis* H1 genes to a heterozygous *ddm1* plant from a line in which the *ddm1* mutation had never been homozygous. The resulting *h1.1*/+;*h1.2*/+;*ddm1*/+ F1 was allowed to self-fertilize to generate an F2 founder in which all three mutant alleles were homozygous (and the *ddm1* allele was homozygous for the first time). As DDM1 predominantly functions in heterochromatin,[34] we focused our analysis on heterochromatic TEs (hTEs; as defined in Choi et al.[49]) that lose much of their DNA methylation in all sequence contexts when this remodeler is inactivated (Figures 1A–1C, S1A, and S1B; Table S1). Consistent with the positive association between DNA methylation loss and H3K9me2 in *ddm1* plants,[34] F2 *h1ddm1* mCG varies substantially between hTEs and shows a strong negative correlation with H3K9me2 in the parental *h1* genotype ($r = -0.60$, Figure 1D). Unlike *ddm1* where most mCG is lost by the F2 (Figure S1C), average and median *h1ddm1* mCG across hTEs decreases substantially from the F2 to the F3 generation (median change 11%), less so from the F3 to the F4 generation (~5%), and even less from the F4 to the F5 (Figures 1A–1C). Overall mCG remains stable at intermediate levels after the F5 generation, up to generation F11 (ten generations of inbreeding; Figures 1A–1C and S1D). Across all ten *h1ddm1* generations analyzed, intermediate mCG spans vast swaths of heterochromatin (>12 Mbp, ~10% of the genome), as can be appreciated from genome browser views (Figure 1B). CG methylation patterns are well correlated across generations (Figure 1E), indicating that the heterogeneous mCG landscape of *h1dddm1* hTEs is accurately reproduced.

CHG methylation (mCHG) behaves similarly to mCG over *h1ddm1* generations, dropping substantially initially, then leveling off in the subsequent generations (Figures S1A and S1E), whereas *h1ddm1* CHH methylation (mCHH) decreases overall, though not monotonically (Figures S1B and S1F). In both cases relative losses are much smaller than for mCG. Early-generation mCHG and mCHH patterns are poorly correlated with mCG, but the correlation improves over time (Figures 1E and S1G), suggesting that methylation in all contexts evolves in concert across *h1ddm1* generations. Although correlations are strongest between methylation patterns of the late generations, early-generation mCHG and mCHH patterns are

CellPress
OPEN ACCESS

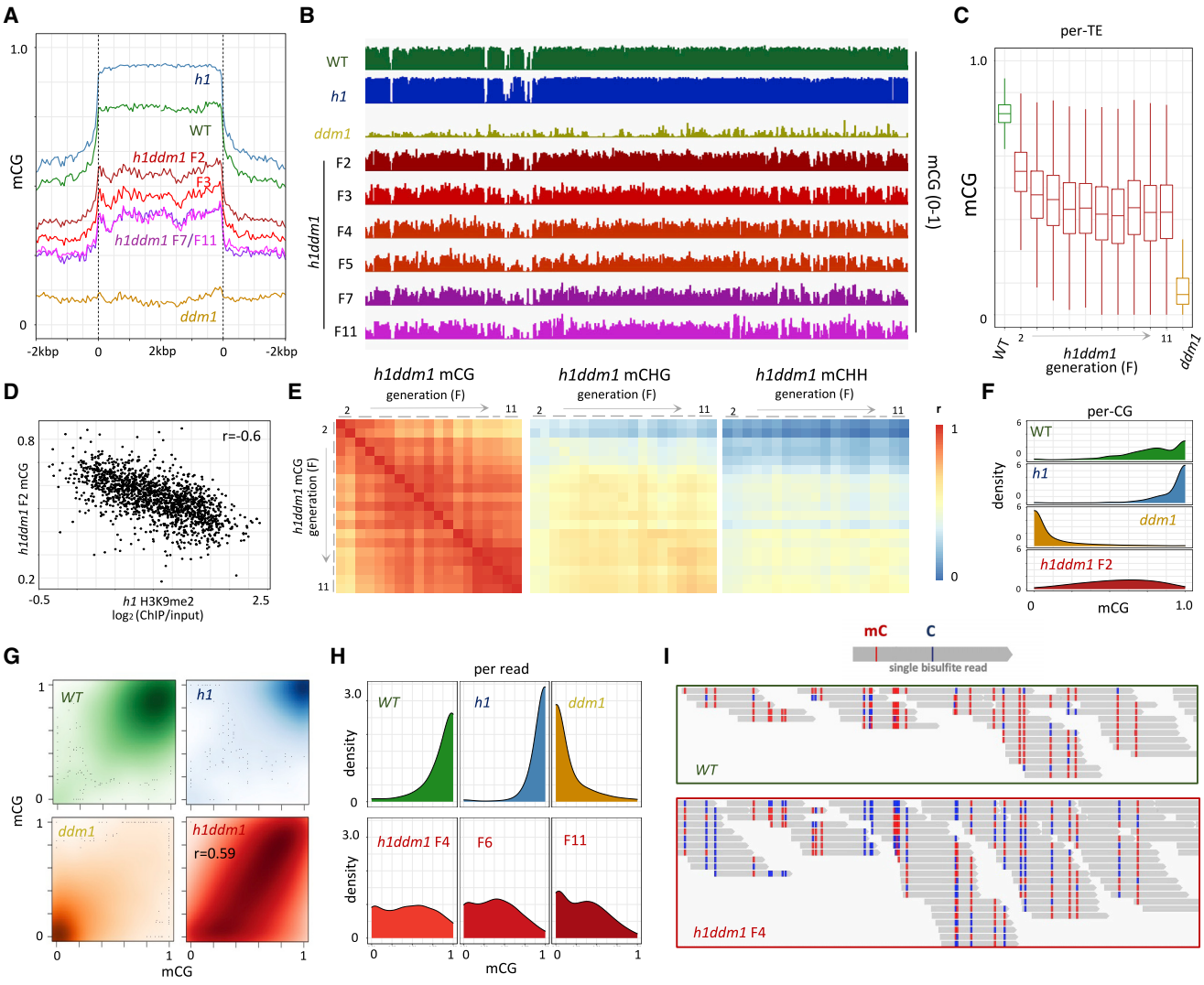


Figure 1. Heterochromatic mCG decays to an intermediate equilibrium in *h1ddm1*

(A) Average fractional CG methylation (mCG) is plotted for all heterochromatic TEs >250 bp (as in Choi et al.[49]) for the indicated genotypes and *h1ddm1* generations.

(B) Genome browser view of a 400-kb window of pericentromeric heterochromatin illustrating mCG across *h1ddm1* generations (Chr.5: 12,950,000–13,350,000).

(C) Boxplots of mCG for all *h1ddm1* heterochromatic TEs (red), beginning with *WT* (green, left) and ending with *ddm1* F5 (dark orange, right). Box depicts the interquartile range and whiskers extend to 1.5× interquartile range; horizontal line shows the median value.

(D) F2 *h1ddm1* per-TE mCG for all TEs >2 kb as a function of H3K9me2 (expressed as $\log_2$ chromatin immunoprecipitation [ChIP]/input) in the parental genotype, *h1*.

(E) Heatmaps of per-TE mCG Pearson's correlation (color bar) for mCG versus mCG, mCHG, and mCHH across *h1ddm1* generations (x axes, from left to right, respectively). Biological replicates are indicated with a gray bar over columns/rows (one replicate for the F10, two for all other generations). Note increased correlation between CG and non-CG methylation in later generations.

(F) Density plot of per-CG methylation for all CG sites with coverage >10 in heterochromatic TEs for the indicated genotypes.

(G) 2D density plots of per-CG methylation for all CG sites in heterochromatic TEs for F4 versus F5 generations (except for *h1ddm1*, where F2 versus F3 is shown to illustrate the correlation between intermediate values across generations at a stage when mCG is changing the most). Darker color indicates more data points.

(H) Density plots of mCG per-bisulfite-read with ≥3 CG sites per read from heterochromatic TEs.

(I) Genome browser view of bisulfite reads for *WT* (top) and *h1ddm1* (bottom); Chr.1: 14,918,800–14,918,900. Only (+) strand mapping reads are shown for clarity.

See also Figure S1 and Table S1.

better correlated with late-generation mCG than with early-generation mCG (Figure 1E). Therefore, to some extent, early-generation non-CG methylation can predict late-generation mCG. This observation suggests that the processes shaping mCG epigenetic inheritance in TEs are related to non-CG methylation.

Intermediate mCG is caused by methylation fluctuations at individual CG sites

A simple explanation for the observed intermediate mCG across *h1ddm1* generations would be methylation heterogeneity across CG sites, with fully methylated and fully unmethylated sites

producing an intermediate pattern when averaged in bins. In this scenario, the gradual decrease and stabilization of mCG in *h1ddm1* would be caused by an increasing number of CG sites permanently switching to an unmethylated state until only sites with stable mCG maintenance remain methylated around the F5 generation. For example, because DDM1 is more important for mCG maintenance in nucleosomes than in the connecting linker DNA,[47] the observed intermediate mCG patterns could be caused by stable mCG maintenance in linkers and unstable maintenance in nucleosomes. This should cause mCG in TEs with poorly positioned nucleosomes to stabilize at lower levels, because shifting nucleosome positions across cell cycles should prevent efficient maintenance at any CG site. However, this is not the case, as mCG behaves similarly across *h1ddm1* generations in TEs with various levels of nucleosome positioning (Figure S1H).

To test the broader possibility of bimodal CG site methylation heterogeneity, we plotted per-base methylation in wild-type (*WT*), *h1*, *ddm1*, and *h1ddm1* plants for all heterochromatic CG sites methylated above 5% in *WT*. The near-binary nature of mCG is evident in *WT* versus *ddm1* plants, and even more so in *h1* versus *ddm1* mutants (Figure 1F), because *h1* enhances heterochromatic mCG (Figures 1A and 1B).[34,47] In contrast, F2 *h1ddm1* hTE mCG is much more uniformly distributed, covering with substantial representation all methylation values from 0 to 1 (Figure 1F). This pattern largely persists over the generations (Figure S1I) and is correlated across generations (Figures 1G and S1J). Therefore, CG sites that are typically consistently methylated across a *WT* population of cells exist in a mixture of methylation states in *h1ddm1* cells.

Another possibility to generate intermediate mCG is methylation heterogeneity across entire loci (TEs, for example), with fully methylated and fully unmethylated loci producing an intermediate pattern when averaged. This scenario entails dynamic switching of locus methylation, with the fraction of loci that lose mCG increasing over the first few generations before stabilizing. To examine this possibility, we analyzed methylation of sequencing reads that correspond to individual DNA molecules, where locus-level variation should produce reads with either low or high methylation. When considering all heterochromatic reads with >3 CG dinucleotides in early, middle, and late generations, the observed distribution of per-read mCG is relatively uniform in *h1ddm1* plants (Figure 1H), similar to the per-CG *h1ddm1* mCG (Figure 1F), indicating that many individual DNA molecules have a mixture of methylated and unmethylated CG sites. This becomes apparent when mCG patterns of individual reads are visualized, with *h1ddm1* reads exhibiting a mixture of mCG levels and patterns (Figure 1I). These results demonstrate that the intermediate *h1ddm1* mCG patterns do not arise from averaging of fully methylated and unmethylated CG sites or from entire loci dynamically switching their methylation states during development. Instead, our results indicate that individual CG sites regularly switch their methylation states, implying a substantial flux of *de novo* mCG in *h1ddm1*.

Although *WT* CG sites tend to have higher methylation levels than in *h1ddm1*, individual CG sites nonetheless show methylation variability between DNA molecules in *WT* (Figure 1I). This is reflected in a broader distribution of per-read methylation levels in *WT* compared with *h1* (Figure 1H) and in a similarly broader distribution of per-CG methylation (Figures 1F and 1G). This suggests that CG sites regularly switch their methylation states, and therefore that mCG epigenetic inheritance in *WT* heterochromatin also involves substantial *de novo* methylation.

Intermediate CG methylation is heritable in *h1ddm1* plants

A possible explanation for the apparently stable inheritance of intermediate heterochromatic mCG in *h1ddm1* plants is that we are assaying the wrong tissue. We measure DNA methylation in leaves, which do not contribute to the subsequent generation. Cells that mediate inheritance, such as reproductive cells, show more robust maintenance of mCG than leaves and other somatic tissues.[50,51] Therefore, the *h1ddm1* genotype might primarily destabilize somatic but not reproductive methylation maintenance. In this scenario, the intermediate mCG patterns we observe in leaves may be produced in each generation through methylation decay during somatic development from an initially high level expected to be compatible with primarily semiconservative inheritance. To test this hypothesis, we quantified sperm DNA methylation in the F4 and F5 generations of *h1ddm1* plants and *WT*, *h1* (only the F5 generation), and *ddm1* siblings. *WT* plants have increased mCG in sperm compared with leaf (Figures 2A and 2B), in agreement with previous work,[50] whereas mCG is similarly low in *ddm1* sperm and leaf (Figures 2A and 2C). Sperm mCG is unchanged in *h1* compared with *WT* (Figures 2A–2C), consistent with the low H1 levels in male reproductive cells.[52] Additionally, we find a large reduction of mCHH (but not mCHG) in sperm compared with leaf in all genotypes (Figures S2A and S2B), indicating that mCHH reprogramming in sperm[53] does not require DDM1.

Importantly, *h1ddm1* plants possess nearly the same average mCG in sperm and leaf in the F4 and F5 generations (Figures 2A and 2B), a phenomenon readily apparent in the genome browser (Figure 2D). Furthermore, there is a high correlation of per-site mCG between sperm and leaf of the same generation (Figure 2E, top) as well as sperm across generations (Figure 2E, bottom). Similar to *h1ddm1* leaves (Figure 1F), *h1ddm1* sperm per-site mCG shows a continuous distribution of methylation frequencies (Figure 2F). Per-read sperm methylation patterns in *h1ddm1* (Figures 2G and 2H) are also similar to those of leaves (Figures 1H and 1I), indicating a mixture of mCG levels and patterns in *h1ddm1* sperm cells.

Because sperm cells directly initiate the next generation, our results demonstrate that patterns of intermediate methylation are indeed inherited. The extensive mCG heterogeneity between DNA molecules (Figures 2G and 2H), which represents heterogeneity between the haploid sperm, indicates that widespread *de novo* methylation of CG sites mediates stable inheritance of heterochromatic mCG in *h1ddm1* plants. Notably, per-read mCG heterogeneity is also apparent in *WT* sperm (Figures 2G and 2H), suggesting an important role for *de novo* methylation in maintaining *WT* heterochromatic mCG.

Transcriptional start site methylation is quantitatively associated with TE expression

Because loss of DNA methylation is accompanied by widespread TE activation in *ddm1* mutants,[54,55] we examined

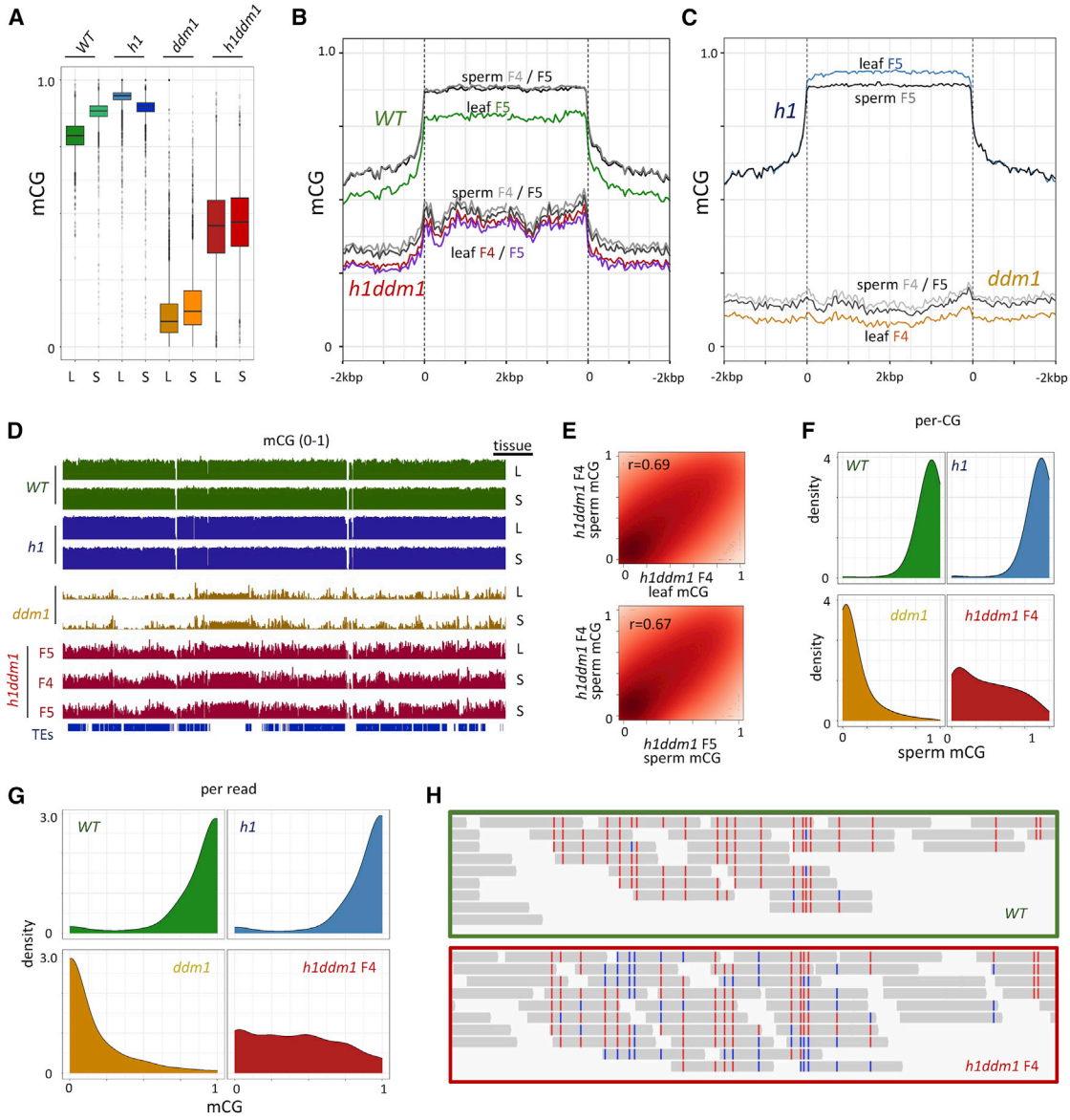


Figure 2. Intermediate *h1ddm1* mCG is heritable

(A) Boxplots of F5 mCG for all heterochromatic TEs with leaf values (L) displayed next to sperm (S) in the indicated genotypes.

(B and C) Average mCG in sperm and leaf for the indicated generations and genotypes plotted relative to the start and end sites of heterochromatic TEs, as in Figure 1A.

(D) Genome browser view of 500 kb of heterochromatin (Chr.5: 11,050,000–11,550,000) highlighting the similarities between sperm and leaf *h1ddm1* mCG. Source tissue is either leaf (L) or sperm (S). All non-*h1ddm1* genotypes shown are F5 generation.

(E) 2D density plots comparing per-CG methylation at sites with coverage >10 in F4 sperm versus F4 leaf (top) or F5 sperm (bottom) across heterochromatic loci. Pearson's r is shown. Darker color indicates more data points.

(F) Density plots of sperm per-CG methylation for all CG sites with coverage >5 in heterochromatic TEs.

(G) Density plots of sperm mCG per-bisulfite-read with ≥3 CG sites per read.

(H) Screenshot of per-read methylation from F4 sperm of *WT* and *h1ddm1* at Chr.3: 12,425,889–12,426,499. Only (+) strand is shown for clarity. Each gray rectangle is a bisulfite read. Blue lines therein represent unmethylated CG sites; red lines represent methylated CG sites.

See also Figure S2.

whether the increased levels of DNA methylation in *h1ddm1* (compared with *ddm1*) are associated with reduced TE expression. After assembling transcripts based on the Araport11 assembly,[56] which includes *ddm1*-specific TE transcripts,[57] we quantified expression in *WT*, *h1*, *ddm1*, and *h1ddm1* plants. Of the 1,375 transcripts upregulated in *ddm1*, 1,075 overlap heterochromatic TEs. Among these, 796 are bona fide expressed heterochromatic loci, based on their H3K9 methylation profiles and

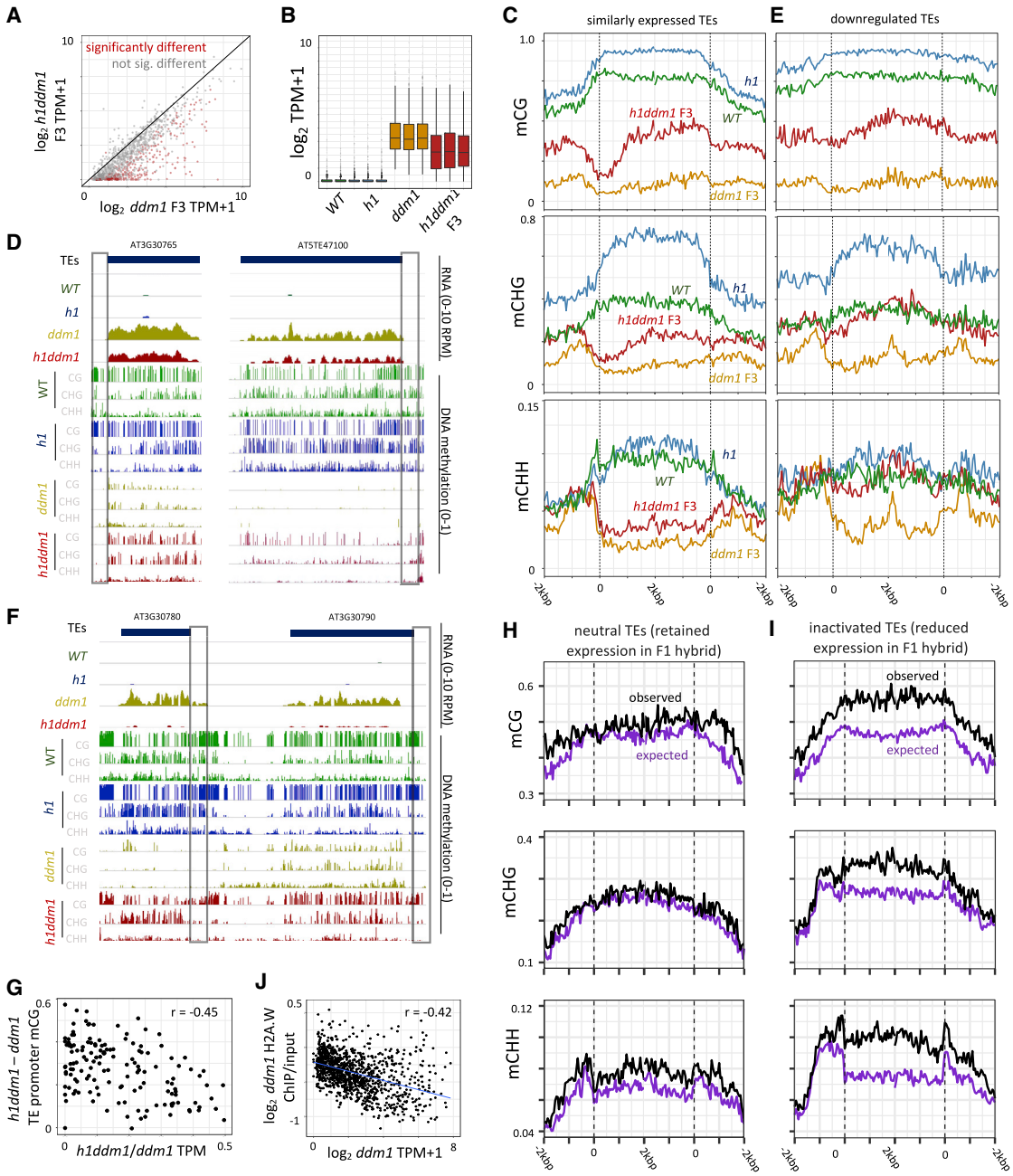


Figure 3. Promoter methylation is quantitatively associated with TE expression

(A) Scatterplot of *h1ddm1* versus *ddm1* average expression levels for heterochromatic transcripts. Color indicates whether transcript was found to be significantly changed (q value <0.05, likelihood ratio test as per Pimentel et al.[58]) in *h1ddm1*.

(B) Boxplots of expression data from (A) showing biological replicates.

(C) DNA methylation averages of indicated cytosine context (y axis) plotted from transcription start and end sites of TEs that are similarly expressed in *ddm1* and *h1ddm1* (q value >0.05 based on likelihood ratio test[58] and *ddm1* average expression level >2.5 TPM, n = 284).

(D) Genome browser views of loci showing depletion of DNA methylation at TE gene promoters and corresponding expression (promoter regions highlighted in boxes). Cytosine context noted in gray.

(E) Methylation averages as in (C) of TEs downregulated in *h1ddm1* (q value <0.05 and *ddm1* average expression level >2.5 TPM, n = 210).

(F) Genome browser view of neighboring loci showing promoter mCG at TE gene promoters and corresponding repression of expression in *h1ddm1* but not *ddm1* (promoter regions highlighted in boxes).

(G) Ratio of expression levels (*h1ddm1/ddm1*) per *ddm1* derepressed TE >2 kb (n = 146) compared with their respective change in promoter mCG, where the promoter region is ±500 bp of TSS.

(legend continued on next page)

expression level (H3K9me1 or -me2 >0.25 in *WT*, >2.5 transcripts per million [TPM] in *ddm1*; Figure S3A). Expression of these heterochromatic transcripts is generally reduced in *h1ddm1* compared with *ddm1* (Figures 3A and 3B), with 303 of these loci significantly downregulated (Figure S3B, q value <0.05). In contrast, expression of upregulated euchromatic *ddm1* transcripts (n = 131) is not, on average, substantially altered in *h1ddm1* (Figure S3C).

TEs that are similarly expressed between *ddm1* and *h1ddm1* possess a striking, localized depletion of DNA methylation in all contexts near the transcriptional start site (TSS) in *h1ddm1*, so that their methylation levels approach those of *ddm1* plants at the TSS (Figures 3C and 3D). In both CG and CHG contexts, average methylation downstream of the TSS increases to an intermediate level within 1 kb (Figures 3C and 3D), illustrating that, as in genes,[60,61] DNA methylation in the TE body is compatible with high levels of expression. In contrast, TEs with reduced expression in *h1ddm1* compared with *ddm1* show little or no methylation depletion near the TSS and overall non-CG methylation levels similar to *WT* (Figures 3E and 3F), indicating that the increased levels of CG and non-CG methylation in *h1ddm1* plants (compared with *ddm1*) attenuate TE expression. Consistently, the difference in mCG between *h1ddm1* and *ddm1* around the TSS of downregulated TEs (±500 bp; TEs longer than 2 kb) exhibits a negative correlation ($r = -0.45$) with the ratio of TE expression levels between *h1ddm1* and *ddm1* (Figure 3G). Taken together, our results support the hypothesis that mCG around the TSS quantitatively reduces TE expression.

Some TEs continue to be highly expressed following *ddm1* outcrossing to *WT* despite DDM1 restoration, whereas others are silenced.[44] A recent study reported that TE silencing in F1 progeny of *ddm1* and *WT* is not correlated with DNA methylation but is instead linked to deposition of the H2A.W histone variant.[59] However, we find a strong recovery of DNA methylation in all contexts around the TSS of TEs resilenced in the F1 hybrids, but not at expressed TEs (Figures 3H and 3I), as would be expected if TE expression is reduced due to increased DNA methylation, and in line with earlier results.[44]

Considering that loss of DDM1 disperses heterochromatic foci,[62] greatly reduces DNA methylation in all contexts,[34,63] and strongly activates TE expression,[54,55,64] whereas loss of H2A.W does not substantially disperse heterochromatin, alter DNA methylation, or activate TE expression,[65] reduced H2A.W occupancy in heterochromatin is not likely to be a primary cause of the *ddm1* phenotype. Instead, our data indicate that TE activation in *ddm1* mutants is caused by loss of DNA methylation, and reduced TE expression in *h1ddm1* and *ddm1* × *WT* plants is a consequence of increased methylation. As H2A.W occupancy within TE bodies is anticorrelated with TE expression (Figure 3J), transcription-coupled H2A.W eviction plausibly explains at least some of the observed H2A.W loss from *ddm1* heterochromatin.[59]

Morphological phenotypes are ameliorated in *h1ddm1* compared with *ddm1*

In addition to extensive disruption of heterochromatin, loss of DDM1 function causes CHG hypermethylation of genes and severe morphological phenotypes.[34,66–68] As heterochromatic methylation and TE silencing are partially restored in *h1ddm1* plants, we evaluated whether other *ddm1* defects are also ameliorated. We find that the widespread *ddm1* genic mCHG increase is largely abolished in *h1ddm1* (Figures 4A and 4B). This likely occurs because the CMT3 pathway that catalyzes mCHG is dosage sensitive,[45,69] and its elevated activity in *h1ddm1* versus *ddm1* heterochromatin (Figure S1A) reduces aberrant activity within genes. The *ddm1* morphological phenotypes are also markedly improved in *h1ddm1* plants, which have more rosette leaves with more total area, are not delayed in bolting, and grow taller and produce significantly more siliques (fruit) than *ddm1* (Figures 4C and 4D). We have propagated *h1ddm1* plants until generation F14, at which point they remain fertile and generally healthy (Figure S4A). Thus, in addition to stabilized intermediate DNA methylation of heterochromatin, *h1ddm1* plants have a stable, relatively normal morphology in comparison with *ddm1*.

Genic mCG declines across *ddm1* and *h1ddm1* generations

Although DDM1 is primarily important for TE methylation, some genes, especially those with low levels of expression, lose mCG in *ddm1* plants (Figure S4B).[47,68] Unlike *ddm1* hTEs, in which mCG declines precipitously in the F2 generation and shows little change thereafter (Figure S1C), these genes show a gradual transgenerational mCG decline (Figure 4E). As in TEs, mCG in *ddm1*-affected genes is substantially higher in *h1ddm1* than in *ddm1* (Figure 4E). Compared with TEs (Figure 1C), genic mCG declines more slowly and stabilizes more gradually across *h1ddm1* generations (Figure 4E). The slower decline of genic mCG in *ddm1* and *h1ddm1* compared with TEs suggests that mCG maintenance is more robust in genes in the absence of DDM1. This is consistent with most genes retaining roughly *WT* mCG in *ddm1*,[47] so that some loci (hTEs) have strongly compromised mCG maintenance, some (lowly expressed genes) have modestly reduced mCG maintenance, and others (most genes) have effectively normal mCG maintenance.

Quantification of *h1ddm1* mCG *de novo* and maintenance failure components

The pattern of transgenerational mCG loss and stabilization in *h1ddm1* resembles an exponential decay (Figure 1C) that should be sensitive to the rates of both maintenance failure and *de novo* methylation. To characterize *h1ddm1* mCG dynamics more precisely, we developed a mathematical model that describes the CG methylation reaction using three processes expressed as linear difference equations: *de novo* methylation per cell cycle

(H and I) DNA methylation averages of indicated cytosine context (y axis) plotted from transcription start and end sites of TE genes as characterized in Osakabe et al.[59] that have similar expression in *ddm1* and F1 *ddm1* × *WT* hybrid (H, "neutral") or reduced expression in the F1 hybrid (I, "inactivated").
(J) *ddm1* H2A.W ($\log_2$ ChIP/input) as a function of *ddm1* expression level per TE.
See also Figure S3.

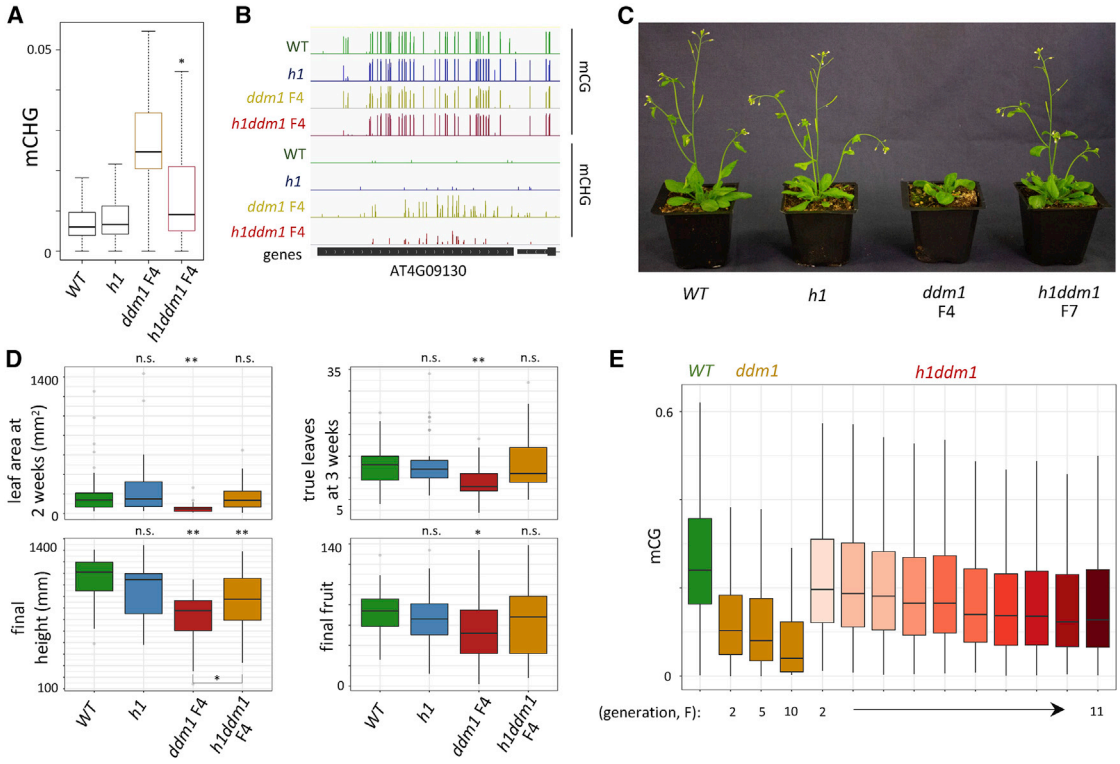


Figure 4. *h1ddm1* rescues diverse *ddm1* phenotypes

(A) Boxplot illustrating CHG hypermethylation in *ddm1* (F4) gene-body methylated genes is partially ameliorated in F4 *h1ddm1* (n = 561). Asterisk indicates significant decrease of mCHG in *h1ddm1* compared with *ddm1* (paired t test, $p < 2.2 \times 10^{-16}$).

(B) Genome browser view of *ddm1* CHG hypermethylation and its reduction in *h1ddm1* (F4) at AT4G09130.

(C) Plants at 4 weeks post germination; genotypes are indicated. All genotypes are F4 except *h1ddm1* is F7.

(D) Boxplots of plant measurements conducted on F4 plants of indicated genotypes. Sample numbers per genotype are as follows: *WT*, 43; *h1*, 38; *ddm1*, 35; *h1ddm1*, 49. Two-sample t tests were calculated comparing *WT* with each of the mutants; *$p < 0.05$; **$p < 1 \times 10^{-4}$; n.s., $p > 0.05$. For final height, an additional test is shown, indicating that *h1ddm1* is significantly improved compared with *ddm1*.

(E) Boxplots of mCG in genes hypomethylated in *ddm1* (n = 529, $p < 1.0 \times 10^{-13}$, Fisher's exact test). *ddm1* F10 obtained from Ito et al.[68]

See also Figure S4.

(delta, δ), maintenance failure per cell cycle (epsilon, ε), and DNA replication, assuming 34 cell cycles per generation based on published estimates for long-day conditions[29] (Figure 5A and STAR Methods). Over many cell cycles, any given CG site can be affected multiple times by these processes, dynamically changing back and forth between methylated and unmethylated. Overall, each TE has an initial *WT* mCG level, M_0, and a lower steady-state mCG level, termed M*, that is approached after multiple generations (n) of inbreeding, such that $M_n \rightarrow M^*$ for large n. The two unknowns δ and ε can be derived on a per-TE-basis by fitting the time-series methylation data to an exponential decay, with a timescale $[(\delta + \varepsilon)/2]^{-1}$, that tends to $M^* = \delta/(\delta + \varepsilon)$ (STAR Methods). Longer TEs (>2 kb) proved to be more amenable to modeling owing to lower levels of noise (n = 1,910). Excluding TEs with an average *WT* mCG $\leq$0.65 (n = 27, 1.4% of total), the mCG trajectory of the vast majority of TEs >2 kb (1,617/1,883, 86%) could be reliably fit to an exponential decay (STAR Methods). The fit to the generational averages is shown in Figure 5B. Although many short TEs were excluded, the remaining TEs comprise ~80% of total heterochromatic sequence length in the *Arabidopsis* genome.

By individually modeling the above TEs, we find that *h1ddm1 de novo* and maintenance failure rates are within comparable ranges and show an intriguing linear correlation of 0.51 (δ_{median} = 0.022, δ_{mean} = 0.021, SD = 0.010; ε_{median} = 0.033, ε_{mean} = 0.035, SD = 0.015) (Figure 5C). We were unable to calculate δ and ε for *ddm1* owing to the very rapid loss of mCG in this genotype (Figure S1C). However, we find that TEs that most readily regain mCG in *ddm1* × *WT* F1 plants[59] have the highest δ values in *h1ddm1* (Figure S5A, r = 0.48), suggesting that the *h1ddm1* δ values are relevant for other genotypes.

By altering the values of δ and ε, essentially any steady-state methylation level can be achieved, as shown in the breadth of observed M* at individual TEs, ranging from near 0 to greater than 60% (Figure 5D). Furthermore, different δ and ε combinations can achieve the same M*. To help visualize this, we overlaid lines on a plot of ε versus δ that correspond to given steady-state mCG states (M*) (green lines in Figure 5C). These lines emphasize that a low ε (efficient maintenance) is usually required for the *WT* mCG level, and that even a small increase of the absolute maintenance failure rate can have a substantial impact on the steady-state methylation level. Thus, although the average

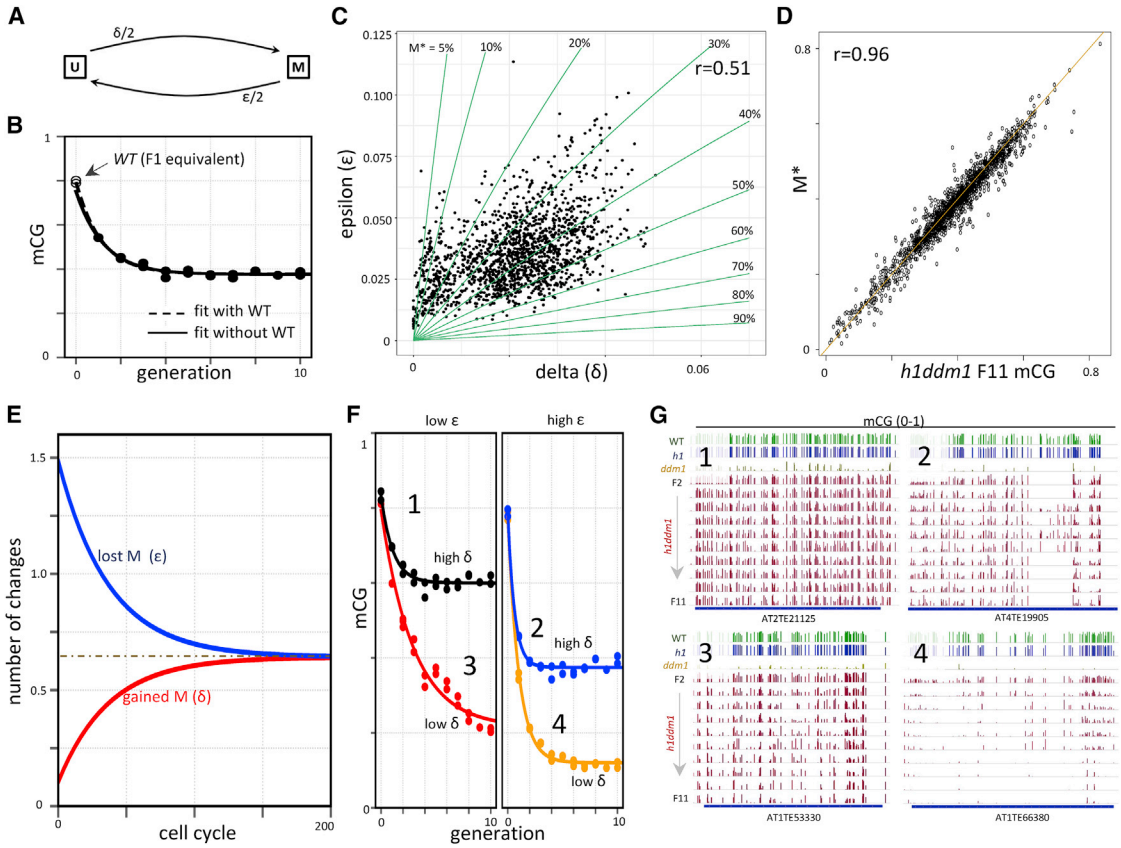


Figure 5. Estimation of per-TE *de novo* and maintenance mCG rates in *h1ddm1*

(A) Gain and loss rates per cell cycle as a function of *de novo* methylation rate (δ) and maintenance methylation failure rate (ε). U, unmethylated; M, methylated.

(B) The average mCG of all TEs >2 kb with *WT* mCG >0.65 across *h1ddm1* generations (generation 1 = *h1ddm1* F2) fits an exponential decay. Solid line indicates the fit without including *WT* methylation values as the initial data point (generation 0); dashed line represents the *WT* + *h1ddm1* fit.

(C) Scatterplot of δ versus ε at those TEs fit by an exponential (n = 1,617). The overall Pearson's correlation coefficient (r = 0.51) between the δ and ε values is shown. Green lines indicate the δ and ε required to achieve the indicated steady-state methylation level, M*.

(D) Scatterplot of modeled steady-state methylation (M*) versus *h1ddm1* mCG at F11.

(E) Plot of changes in methylation per cell cycle for a hypothetical TE with 100 CG sites beginning with 90% methylation and with average *h1ddm1* δ and ε values (0.021 and 0.033, respectively). Dotted line illustrates the number of changes at steady state (0.64).

(F) Different δ and ε regimes produce a variety of mCG dynamics in *h1ddm1*. Quintiles of δ and ε were used to subset the modeled TEs and fit their respective average mCG. Generation 0 = WT, generation 1 = *h1ddm1* F2, and so forth.

(G) Genome browser views of representative TEs from each of the four groups in (F). Numbers listed next to curves in (F) correspond to browser views shown. See also Figure S5.

h1ddm1 ε is only about 3%, this compounds over many cell cycles to produce a substantially lower steady-state mCG compared with *WT* (Figures 1A–1C).

To appreciate the implications of this model, consider a TE composed of 100 CG sites, beginning with 90% methylation, and where δ and ε are equal to their *h1ddm1* mean values. *De novo* activity affects unmethylated CG sites (uCG), and therefore plays only a small role in the initial cell cycles following the creation of the mutant state (generation 0, only 10 uCG, Figure 5E). However, as time progresses, unmethylated sites increase in prevalence and the absolute contribution of *de novo* methylation increases, whereas maintenance failure becomes relatively less important as the number of mCG sites decreases. A steady state is approached around cell cycle 150 (corresponding to just over four generations), when the numbers of gains and losses are balanced, both converging to a constant equal to (on average) 0.64 CG sites gaining/losing per cell cycle (dashed line in Figure 5E).

Different regimes of δ and ε manifest as different steady-state levels and/or mCG trajectories (with faster or slower decays) (Figures 5F, 5G, and S5B). Efficient maintenance and low *de novo* each cause a slower decay to the mCG steady state, so that TEs with excellent maintenance and low *de novo* fail to reach steady state even after ten generations (Figures 5F and 5G, #3, n = 146 TEs), whereas those with highest *de novo* but poor maintenance converge to steady state rapidly (Figures 5F and 5G, #2, n = 149 TEs). At the most mCG-depleted TEs *de novo* rates are low, with larger maintenance failure, and average mCG stabilizes at ~10% (Figures 5F and 5G, #4, n = 17 TEs). At the other end of the spectrum, TEs with high δ and low ε lose ~20% of their *WT* mCG (Figures 5F and 5G, #1, n = 13 TEs). However, because

δ and ε are positively correlated (Figure 5C), TEs with poor maintenance tend to have better *de novo* rates, and vice versa, so that the above extremes (Figures 5F and 5G, #1 and #4) are unusual. Although we could not fit genic methylation dynamics well enough to confidently quantify δ and ε, the more gradual decline of mCG in genes (Figure 4E) than in TEs (Figure 1C) is consistent with lower δ and ε (less *de novo* activity and better maintenance) in genes.

Contributions of DRM and CMT methyltransferases to CG methylation in *h1ddm1*

Our model indicates that the initial mCG decline is dominated by maintenance failure, whereas steady-state mCG has a strong dependence on *de novo* methylation (Figures 5E and 5F). The correlation of late-generation (steady-state) mCG with mCHG and mCHH (Figures 1E and S1G) therefore suggests that the RdDM and/or CMT pathways that mediate mCHG/CHH contribute (directly or indirectly) to *de novo* mCG activity. Consistent with this, δ shows strong linear correlations with mCHG and mCHH (r_{mCHG} = 0.66, r_{mCHH} = 0.59) (Figures S5C and S5D). To elucidate the enzymatic origins of *de novo* mCG, we generated three *h1ddm1* mutant lines, each lacking one of the three principal *Arabidopsis* non-CG methyltransferases: DRM2, CMT2, and CMT3. We inbred each line through the F7 generation and analyzed leaf methylomes as we did for *h1ddm1*. The *h1ddm1drm2* and *h1ddm1cmt2* lines were created in the same way as the initial *h1ddm1* line: a plant homozygous for all mutations except *ddm1* was allowed to self-fertilize to create the founder F2 in which the *ddm1* mutation was homozygous for the first time. We were unable to generate *h1ddm1cmt3* plants this way, and instead created an F2 *h1ddm1*(*cmt3*/+) plant (first generation of *ddm1* homozygosity), which was allowed to self-fertilize to create F3-equivalent *h1ddm1cmt3* plants (first generation of *cmt3* homozygosity).

Average hTE mCG is greatly reduced in *h1ddm1drm2* compared with *h1ddm1*, which becomes obvious after the F2 generation (Figures 6A and 6B). Consistently, the overall morphology of *h1ddm1drm2* plants resembles *ddm1* more than *h1ddm1* (Figures S6A and S4A). In contrast, neither *h1ddm1cmt2* nor *h1ddm1cmt3* lines exhibit an overall mCG decrease compared with *h1ddm1* (Figures 6A and 6B), despite corresponding strong losses of mCHH and mCHG, respectively (Figures S6B and S6C), and these lines morphologically resemble *h1ddm1* (Figure S6A).

Although the loss of CMT3 does not substantially reduce average heterochromatic mCG, we identified 86 TEs with mCG in *h1ddm1cmt3* mutants comparable with *h1ddm1drm2* (>50% mCG decrease; $p < 0.01$, Fisher's exact test; Figures 6C–6E and S6D). Because non-CG methylation promotes RdDM,[48,70,71] we suspected the loss of CMT3 might reduce RdDM activity at these loci. Indeed, mCHH (the hallmark of RdDM) is nearly eliminated at these TEs in F7 *h1ddm1cmt3* plants, as it is in F7 *h1ddm1drm2* (Figures 6F and 6G, mCHG shown in Figure S6E), indicating that RdDM activity depends on CMT3 at a subset of heterochromatin. These CMT3-dependent TEs also have reduced mCHH in F7 *h1ddm1cmt2* plants (Figures 6F and 6G) as well as reduced mCG (Figures 6D and 6E), so that steady-state mCG is affected by all three of the tested methyltransferases. Notably, non-CG methylation at these TEs does not collapse immediately in the compound methyltransferase mutants, but can be similar to *h1ddm1* in early generations and tends to decrease in tandem with mCG (Figures 6E, 6F, and S6F–S6H). This is consistent with our observation that CG and non-CG methylation changes are correlated across *h1ddm1* generations (Figures 1E and S1G), and with the known dependence of non-CG methylation on mCG at some TEs.[39,48,49] The CG and non-CG methylation decreases may therefore be mutually reinforcing. Together, these data demonstrate that DRM2 is a key contributor to intermediate mCG in *h1ddm1* heterochromatin but also that epigenetic inheritance of mCG involves integration of the RdDM, CMT, and MET1 pathways, at least at some loci.

DRM2 mediates most of the *de novo* mCG in heterochromatin

Application of our discrete methylation model to the methyltransferase mutant data allows us to calculate the relative contributions of each methyltransferase to *de novo* and maintenance mCG (Figures 7A–7C). Surprisingly, we find that ε decreases for all three compound mutants compared with *h1ddm1* (Figure 7D), indicating that maintenance becomes more efficient when these methyltransferases are removed. This is likely related to the observation that F2 *h1ddm1* mCG shows a strong negative correlation with H3K9me2 (Figure 1D). As the initial mCG decline is dependent in large part on ε (Figures 5E and 5F), ε should also correlate with H3K9me2, which is indeed the case (Figure 7E). Because non-CG methylation and H3K9me2 vary in tandem,[35] our data suggest that the *drm2*, *cmt2*, and *cmt3* mutations improve mCG maintenance by rendering TEs less heterochromatic (less non-CG methylation and H3K9me2) and therefore less dependent on DDM1.

The *cmt* compound mutants exhibit modest decreases in δ that are comparable with the decrease in ε. In *h1ddm1cmt2*, δ decreases by 33% (from 0.021 to 0.014) and ε decreases by 26% (from 0.035 to 0.026), and in *h1ddm1cmt3*, δ decreases by 43% (to 0.012) and ε decreases by 34% (to 0.023; Figures 7D, 7F, and 7G). The changes between δ and ε are correlated (Figures S7A–S7C), resulting in a form of dynamic compensation that, in the case of *cmt* mutants, keeps steady-state mCG (which depends on the ratio of δ and ε) nearly the same as in *h1ddm1* (Figures 7H–7J). However, in *h1ddm1drm2*, δ falls 75% (to 0.0053; Figures 7F and 7G) but maintenance does not improve at a similar level (ε decreases by 23% to 0.027; Figures 7D and 7G). The much greater relative decrease in δ causes greatly reduced steady-state mCG compared with *h1ddm1* (Figures 7H and 7K) and, accordingly, TEs with both high δ and low ε are very uncommon in *h1ddm1drm2* as compared with low δ and high ε (Figures 7L and 7M, black (n = 19) and orange (n = 42) points, respectively). Our results indicate that although CMT2, CMT3, and DRM2 all participate in the epigenetic dynamics of heterochromatic mCG, DRM2 contributes the bulk of the *de novo* mCG activity.

De novo rate estimates are consistent between *h1ddm1* and *met1*

Because we calculate TE *de novo* mCG rates in the *h1ddm1* background, it is important to ascertain how relevant these rates are for *WT*, especially because loss of histone H1 substantially

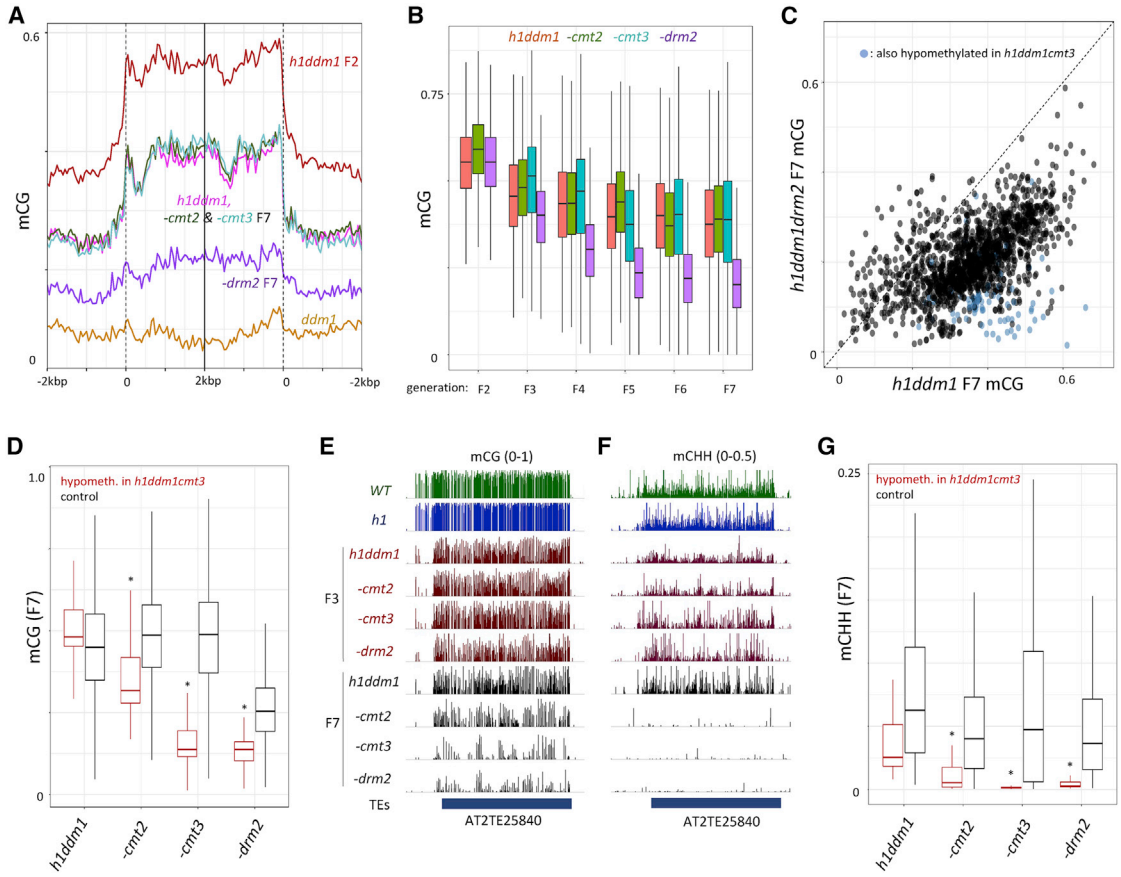


Figure 6. Steady-state mCG in *h1ddm1* depends on DRM2

(A) Average mCG is plotted for all heterochromatic TEs for indicated genotypes around TE start and stop sites (TEs >250 bp). *h1ddm1*, *h1ddm1cmt2*, and *h1ddm1cmt3* resemble one another at F7, whereas mCG is much reduced in *h1ddm1drm2*.

(B) Boxplots of per-TE mCG over seven generations for *h1ddm1* and compound *h1ddm1* methyltransferase mutants (*h1ddm1cmt3* F2 is absent; see text).

(C) Scatterplot comparing *h1ddm1drm2* F7 with *h1ddm1* F7 average mCG of previously modeled TEs (n = 1,619), with blue points highlighting the TEs significantly hypomethylated in *h1ddm1cmt3* relative to *h1ddm1* (n = 86, $p < 0.01$, Fisher's exact test).

(D) Boxplots comparing mCG in modeled TEs that are either hypomethylated in *h1ddm1cmt3* (red) or control TEs that are not significantly hypomethylated (black). Average mCG in these TEs is also significantly lower in *h1ddm1cmt2* and *h1ddm1drm2* (*$p < 2.2 \times 10^{-16}$, paired t test).

(E and F) Genome browser view illustrating mCG (E) and mCHH (F) at TE with CMT3-dependent mCG.

(G) Boxplots comparing mCHH levels as in (D). Note the near complete depletion of mCHH in *h1ddm1cmt3* but not *-cmt2* at CMT3-dependnent TEs, indicating that CMT3 is required for DRM2 activity at these loci. Average mCHH in these TEs is also significantly lower in *h1ddm1cmt2* and *h1ddm1drm2* (*$p < 2.2 \times 10^{-16}$, paired t test).

See also Figure S6.

perturbs RdDM activity.[48,72] To address this issue, we calculated *de novo* rates in *met1* and *h1met1* mutants using published data.[49] These mutants are assumed to lack maintenance activity ($\varepsilon = 1$), and δ values can therefore be straightforwardly inferred based on measured steady-state mCG (δ = mCG, STAR Methods). The *met1* and *h1met1* δ values are correlated with *h1ddm1* δ values (Figures S7B and S7C, r = 0.28 for *met1* and r = 0.34 for *h1met1*), supporting the conclusion that *h1ddm1* δ values are relevant for other genotypes. Some TEs retain RdDM and CMT activity in *met1* and *h1met1* mutants (MET1-independent TEs), whereas others lose methylation in all sequence contexts (MET1-dependent TEs).[39,48,49] As expected, *h1ddm1* δ values show similar distributions at these TE categories (Figure 7N). However, the *met1* and *h1met1* δ values are much higher in MET1-independent TEs, where they are much more comparable with those for *h1ddm1* (Figure 7N), as would be expected if *de novo* activity is mediated by the non-CG methylation pathways.

Bisulfite sequencing has an intrinsic error rate caused by incomplete chemical conversion of unmethylated cytosine, as well as PCR and sequencing errors.[73] This rate can be estimated on the basis of the measured methylation frequency in the unmethylated chloroplast genome, and is 0.17% for both *met1* and *h1met1* (dashed lines in Figure 7N). Because per-TE error rates will be distributed around 0.17%, *met1* and *h1met1* δ estimates that are close to this value should be treated as upper bounds, with real δ values potentially much smaller. Therefore, many MET1-dependent TEs may experience very low *de novo* rates

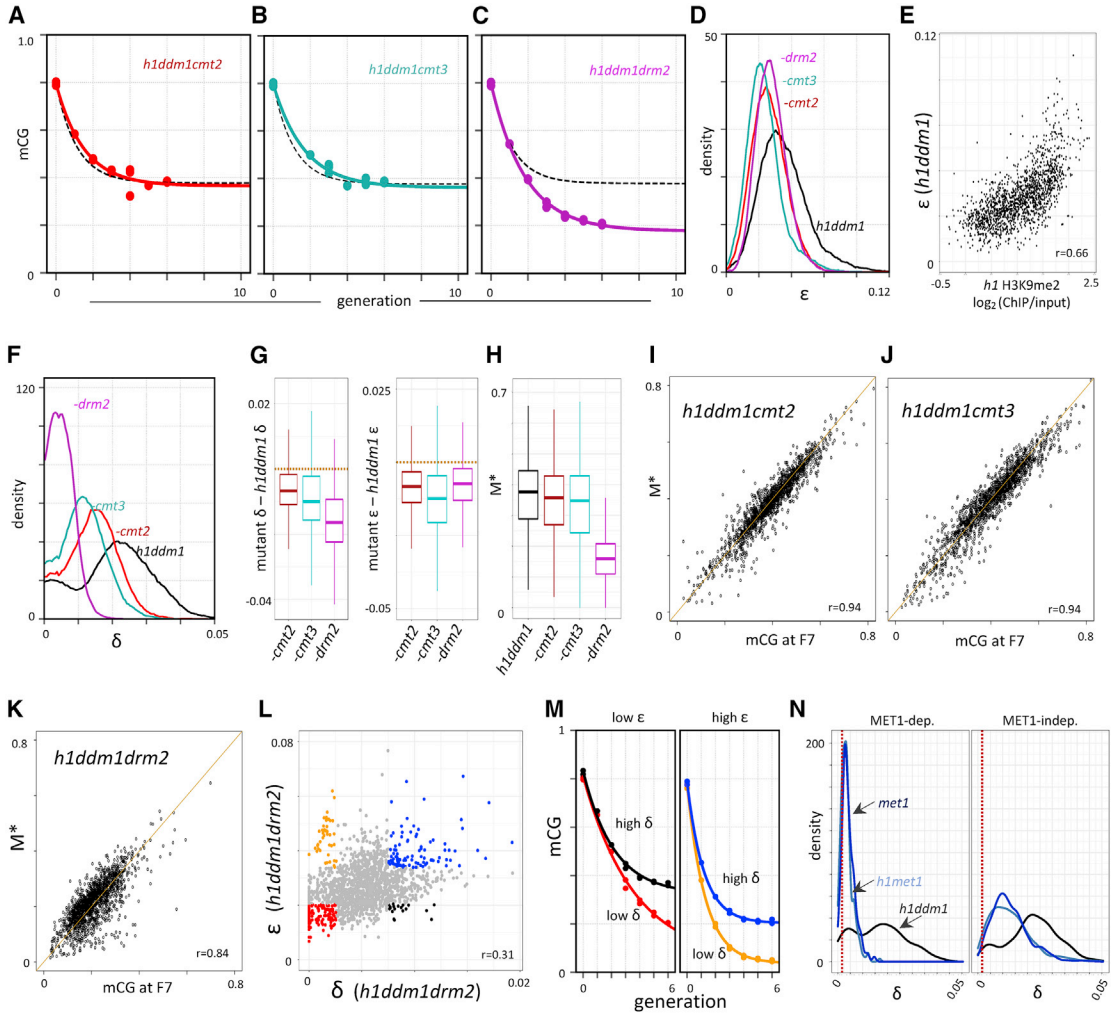


Figure 7. Estimation of mCG rates in compound *h1ddm1* methyltransferase and *met1* mutants

(A–C) Comparison of average modeled TE mCG decay fits (colored by genotype) with *h1ddm1* (black dashed lines).

(D) Density plots of modeled ε for individual TEs.

(E) Scatterplot showing *h1ddm1* ε as a function of parental *h1* H3K9me2.

(F) Density plots of modeled δ (see STAR Methods regarding the low-δ peaks).

(G) Boxplots of the change in δ (left) or ε (right) for the indicated genotypes. Orange horizontal lines are at y = 0. All methyltransferase mutants exhibit significantly lower δ and ε means ($p < 2.2 \times 10^{-16}$ for all comparisons except for *h1ddm1cmt3* ε, for which $p = 6.1 \times 10^{-10}$, two-tailed t test).

(H) Boxplots of modeled steady-state methylation (M*) for the indicated mutants.

(I–K) Scatterplots of modeled steady-state methylation (M*) versus mCG at F7 for the indicated genotypes, with Pearson's r shown.

(L) Scatterplot of δ versus ε in *h1ddm1drm2* TEs, with differently colored points indicating the different extremes possible in this mutant, grouped per quintile combination.

(M) Color-matched exponential decay fits to average mCG of the quintile combinations shown in (L). Generation 0 = *WT*, generation 1 = *h1ddm1drm2* F2, and so forth.

(N) Density plots depicting the distributions of δ for the indicated genotypes in either MET1-dependent (left, n = 271) or MET1-independent (right, n = 1,258) modeled TEs (see STAR Methods). The 5.5% of modeled TEs not in these categories were excluded from this analysis. Red dashed lines indicate the estimated false-positive methylation rate for the *met1* and *h1met1* data (see text).

See also Figure S7.

in *met1* and *h1met1* plants (Figure 7N, left panel). In contrast, MET1-independent TEs generally have δ values well above the error rate (Figure 7N, right panel), with the average *met1* and *h1met1* δ values (0.013 for both) in MET1-independent TEs only 1.7-fold lower than that for *h1ddm1* in MET1-independent TEs ($\delta_{mean} = 0.022$). Because RdDM activity at MET1-independent TEs is similar between *met1* and *WT*, and—unlike *h1ddm1*—*met1* does not cause RdDM to broadly enter heterochromatin,[48] *met1* δ values cannot be explained by greatly enhanced RdDM activity, as they might be in *h1ddm1*. *WT* and *met1* heterochromatic *de novo* rates should therefore be in a comparable range.

DISCUSSION

Despite the similarities between the reported rates of mCG change in *C. neoformans*[14] and *Arabidopsis*,[26–28,46] our data indicate that epigenetic inheritance of mCG in *Arabidopsis* TEs involves a strong *de novo* component comparable with mammalian systems.[16–18] The *de novo* activity is primarily mediated by RdDM, with smaller contributions from CMT2/3. Both of these pathways are self-reinforcing, and are attracted—directly or indirectly—by mCG,[74] thereby solving the problem of targeting *de novo* mCG to regions with existing methylation.

We derive our *de novo* rates primarily from *h1ddm1* data, which may overestimate the *de novo* activity contributed by RdDM in *WT* because heterochromatic RdDM is enhanced in *h1ddm1* plants.[48] However, the *h1ddm1* rates are in good agreement with the rates in *met1* mutants (Figure 7N), which do not have enhanced heterochromatic RdDM,[48] and therefore the *de novo* rates reported here should be a fair estimate of *WT* rates. The *de novo* rate concordance between *h1ddm1* and *met1* may seem at odds with the importance of RdDM for *de novo* mCG. However, the activities of the RdDM and CMT pathways fluctuate across cell types and tissues,[75] including during *Arabidopsis* reproductive and embryonic development.[53,76–80] Embryonic development especially is characterized by a gradual rise of mCHH in TEs that peaks at the onset of germination and is mediated by RdDM and CMT2.[77,78] The functional significance of this process has been unclear, and one of its consequences may be to reinforce mCG inheritance by restoring CG sites within TEs to a methylated state. CG methylation in heterochromatic TEs may thus be stabilized by germline bursts of RdDM activity that resemble the constitutive behavior of RdDM in *h1ddm1* mutants.

The *de novo* mCG rates we find in heterochromatin are about 1% per site per cell cycle ($\delta/2$, Figure 5 and STAR Methods), which translates to around 30% per generation (1–0.99^{34}). These rates are at least three orders of magnitude higher than the overall *de novo* rate of 3.2×10^{-4} per generation reported for TEs[28] and the range of 2.3×10^{-6} to 1.7×10^{-4} recently reported within different TE-associated chromatin states (CSs).[46] This large difference is likely accounted for by two related factors. First, CS rates were calculated using the AlphaBeta model[81] that produces a good fit for CSs associated with genes, explaining up to 88.6% of the data variance in CS5.[46] However, the model performs less well with the TE-associated states CS30–CS36. AlphaBeta provides reasonable fits for CS30 (67.6% of variance explained) and CS31 (57% of variance explained), and these states have relatively high reported *de novo* rates (3.5×10^{-5} for CS30 and 1.7×10^{-4} for CS31). AlphaBeta does not handle other TE CSs well, ranging from 28.3% of variance explained for CS32 to effectively 0% of variance explained for CS36 (variance explained values are from Table S2 in Hazarika et al.[46]). This indicates that the AlphaBeta model is not reliable for these regions of the genome.

More generally, AlphaBeta and other published rate estimates rely on the assumption that the rates are low enough that all or nearly all changes are captured by measurements that are separated by one or several generations. This should be a good assumption for genes, but it is not for TEs. Per-generation *de novo* rates of 30% mean that the half-life of an unmethylated site is short (30% rate implies a half-life of about two generations), leading to underestimates of methylation failure rates. This could partially explain why the reported per-generation maintenance failure rate for genes (1.5×10^{-3}) is 100-fold higher than the rate for TEs (1.2×10^{-5}).[28] Calculation of *de novo* rates may be confounded by an even more serious issue. Because *de novo* rate calculations rely on ancestrally unmethylated CG sites and the half-life of unmethylated sites is inversely related to the *de novo* rate, CG sites with unusually low *de novo* rates will dominate the calculation, leading to potentially very large rate underestimates. This might explain the 1,000-fold difference between the reported per-generation *de novo* rate in TEs (3.2×10^{-4})[28] and our estimate of around 3×10^{-1}.

Our results indicate that the low levels of mCG divergence observed at plant TEs[28,46] are not caused by especially low rates of methylation loss and gain. Instead, this phenomenon is caused by a very high *de novo* rate that is much greater than the maintenance failure rate, so that maintenance errors are rapidly corrected—an idea consistent with the earlier proposal that TE mCG is stabilized by an epigenetic process that favors methylation gain over methylation loss.[28] Our data indicate that stable epigenetic inheritance of TE mCG is ensured by a combination of DDM1-supported MET1 maintenance activity with the strong *de novo* activity of RdDM (with smaller contributions from the CMT pathway). Using the reported per-generation maintenance failure rate for genes (1.5×10^{-3}) as a baseline, the *de novo* activity at TEs is strong enough to stabilize mCG in *h1ddm1* despite a >100-fold increase in the maintenance failure rate, to $\sim 4 \times 10^{-1}$ per generation (equivalent to 1–0.985^{34}, Figure 5; per-cell-cycle loss rate = $\varepsilon/2$, STAR Methods). Importantly, there is no compelling reason to assume that maintenance efficiency is the same or similar between genes and TEs in *WT* plants. Maintenance efficiency at TEs could be higher than at genes, but—even with the support of DDM1—may also be considerably lower, especially at heterochromatic TEs. One of the main conclusions of our study is that the high *de novo* rates in TEs enable stable epigenetic inheritance of mCG within a wide range of maintenance rates.

Our results also bear on the functionality of DDM1 and of plant heterochromatin more broadly. The remarkable alleviation of the *ddm1* phenotype (e.g., TE activation, decreased fecundity, increased gene-body mCHG; Figures 3A–3G and 4A–4D) that emerges in *h1ddm1* concomitantly with the restoration of heterochromatic DNA methylation illustrates the central importance of chromatin homeostasis in the regulation of organismal viability, and the crucial role of DNA methylation in heterochromatin homeostasis and function. This is consistent with the reports that the mere presence of sufficiently large segments of demethylated heterochromatin in genetically *WT* plants is sufficient for phenotypic disruption.[82,83]

Our data do not support the proposal that the *ddm1* phenotype, including TE activation, is primarily caused by loss of the H2A.W histone variant from heterochromatin[59]—a conclusion that is also inconsistent with the published phenotypes of *ddm1* and *h2a.w* loss-of-function mutants.[34,54,65–67,84] Unlike *ddm1*, loss of H2A.W alone does not substantially alter DNA methylation or activate TEs[65] but does promote TE activation

when combined with mutations in other chromatin silencing factors, including H1.[85] That is, loss of H1 has opposite effects on TE activation when combined with loss of H2A.W or DDM1. If the primary direct function of DDM1 were to deposit H2A.W into heterochromatin, the *ddm1* mutation should not cause extensive loss of DNA methylation or TE activation, and the additional inactivation of H1 should promote, rather than dampen, TE expression. The available data are thus inconsistent with the proposal that TE activation in *ddm1* mutants is caused primarily or directly by H2A.W loss, although H2A.W depletion from heterochromatin may contribute to the *ddm1* phenotype. Instead, our results indicate that *ddm1* phenotypes are primarily caused by loss of heterochromatic DNA methylation, and the ameliorative effects of H1 removal in the *ddm1* background stem directly from increased heterochromatic DNA methylation, especially steady-state mCG.

Limitations of the study

Although we present evidence that the *de novo* rates reported here should fairly estimate *WT* rates, the inability to directly measure mCG gain (and loss) rates in *WT* heterochromatin is a limitation of our study. Another limitation is that we do not consider possible contributions of active DNA demethylation to the loss rates, partly because mCG losses in *ddm1* and *h1ddm1* heterochromatin should be dominated by inefficient MET1 maintenance, and because our modeling approach cannot distinguish losses caused by maintenance failure and active demethylation. Furthermore, the source of mCG losses would not affect any of our conclusions about the mechanism and rate of mCG gains. Nonetheless, the contribution of active demethylation to mCG inheritance in heterochromatin remains to be investigated.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Leaf DNA isolation and bisulfite library prep
  - Sperm DNA isolation and bisulfite library prep
  - Leaf RNA isolation and library preparation
  - Sequencing
  - Short read mapping and quantification
  - Visualization of experimental data
  - Plant leaf area measurements
  - Analysis of data from Osakabe et al
  - Transposable element annotations
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Mathematical model for TE methylation dynamics: Constructing the recursion relations
  - Mathematical model for TE methylation dynamics: steady state solutions
  - Mathematical model for TE methylation dynamics: Time dependent solutions
  - Mathematical model for TE methylation dynamics: gain and loss rates per cell cycle
  - Relationship between the mathematical model and bisulfite sequencing data from leaf tissue
  - Mathematical model for TE methylation dynamics: Numerical fits
  - Mathematical model for TE methylation dynamics: Limitations of the model
  - Statistical analysis of empirical data

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112132.

ACKNOWLEDGMENTS

The authors would like to thank Jasper Rine for advice and mentorship to D.B.L., Lesley Philips, Timothy Wells, Sophie Able, and Christina Wistrom for support with plant growth, and Bhagyshree Jamge and Frédéric Berger for help with analysis of *ddm1* × *WT* RNA-sequencing data. This work was supported by BBSRC Institute Strategic Program GEN (BB/P013511/1) to X.F., M.H., and D.Z., a European Research Council grant MaintainMeth (725746) to D.Z., and a postdoctoral fellowship from the Helen Hay Whitney Foundation to D.B.L.

AUTHOR CONTRIBUTIONS

D.B.L. conceived the research plan, performed experiments and analysis, and wrote the paper. A.B. developed the mathematical model, performed analysis, and wrote the paper. S.H. performed sperm isolation and library preparation. J. Choi contributed to RNA-seq analysis. E.H. contributed to DNA methylation analysis. J. Colicchio measured plant phenotypes. I.A. measured plant phenotypes. X.F. supervised experiments on sperm DNA methylation. M.H. developed the mathematical model, supervised analysis, and wrote the paper. D.Z. conceived the research plan, supervised experiments, and wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 4, 2022
Revised: November 10, 2022
Accepted: February 1, 2023
Published: February 22, 2023

REFERENCES

1. Kim, M., and Costello, J. (2017). DNA methylation: an epigenetic mark of cellular memory. Exp. Mol. Med. *49*, e322. https://doi.org/10.1038/emm.2017.10.
2. Zhang, H., Lang, Z., and Zhu, J.-K. (2018). Dynamics and function of DNA methylation in plants. Nat. Rev. Mol. Cell Biol. *19*, 489–506. https://doi.org/10.1038/s41580-018-0016-z.
3. Nuñez, J.K., Chen, J., Pommier, G.C., Cogan, J.Z., Replogle, J.M., Adriaens, C., Ramadoss, G.N., Shi, Q., Hung, K.L., Samelson, A.J., et al. (2021). Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. Cell *184*, 2503–2519.e17. https://doi.org/10.1016/J.CELL.2021.03.025.

4. Williams, B.P., and Gehring, M. (2020). Principles of epigenetic homeostasis shared between flowering plants and mammals. Trends Genet. *36*, 751–763. https://doi.org/10.1016/J.TIG.2020.06.019.

5. Baylin, S.B., and Jones, P.A. (2016). Epigenetic determinants of cancer. Cold Spring Harb. Perspect. Biol. *8*, a019505. https://doi.org/10.1101/cshperspect.a019505.

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

7. Lloyd, J.P.B., and Lister, R. (2021). Epigenome plasticity in plants. Nat. Rev. Genet. *23*, 55–68. https://doi.org/10.1038/s41576-021-00407-y.

8. Ong-Abdullah, M., Ordway, J.M., Jiang, N., Ooi, S.-E., Kok, S.-Y., Sarpan, N., Azimi, N., Hashim, A.T., Ishak, Z., Rosli, S.K., et al. (2015). Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. Nature *525*, 533–537. https://doi.org/10.1038/nature15365.

9. Schmitz, R.J., Lewis, Z.A., and Goll, M.G. (2019). DNA methylation: shared and divergent features across eukaryotes. Trends Genet. *35*, 818–827. https://doi.org/10.1016/J.TIG.2019.07.007.

10. Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. Science *187*, 226–232. https://doi.org/10.1126/SCIENCE.187.4173.226.

11. Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. Cytogenet. Cell Genet. *14*, 9–25. https://doi.org/10.1159/000130315.

12. Jeltsch, A. (2006). On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. Epigenetics *1*, 63–66. https://doi.org/10.4161/EPI.1.2.2767.

13. Gowher, H., and Jeltsch, A. (2018). Mammalian DNA methyltransferases: new discoveries and open questions. Biochem. Soc. Trans. *46*, 1191–1202. https://doi.org/10.1042/BST20170574.

14. Catania, S., Dumesic, P.A., Pimentel, H., Nasif, A., Stoddard, C.I., Burke, J.E., Diedrich, J.K., Cook, S., Shea, T., Geinger, E., et al. (2020). Evolutionary persistence of DNA methylation for millions of years after ancient loss of a de novo methyltransferase. Cell *180*, 263–277.e20. https://doi.org/10.1016/J.CELL.2019.12.012.

15. Haerter, J.O., Lövkvist, C., Dodd, I.B., and Sneppen, K. (2014). Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states. Nucleic Acids Res. *42*, 2235–2244. https://doi.org/10.1093/NAR/GKT1235.

16. Wang, Q., Yu, G., Ming, X., Xia, W., Xu, X., Zhang, Y., Zhang, W., Li, Y., Huang, C., Xie, H., et al. (2020). Imprecise DNMT1 activity coupled with neighbor-guided correction enables robust yet flexible epigenetic inheritance. Nat. Genet. *52*, 828–839. https://doi.org/10.1038/s41588-020-0661-y.

17. Haggerty, C., Kretzmer, H., Riemenschneider, C., Kumar, A.S., Mattei, A.L., Bailly, N., Gottfreund, J., Giesselmann, P., Weigert, R., Brändl, B., et al. (2021). Dnmt1 has de novo activity targeted to transposable elements. Nat. Struct. Mol. Biol. *28*, 594–603. https://doi.org/10.1038/s41594-021-00603-8.

18. Lövkvist, C., Dodd, I.B., Sneppen, K., and Haerter, J.O. (2016). DNA methylation in human epigenomes depends on local topology of CpG sites. Nucleic Acids Res. *44*, 5123–5132. https://doi.org/10.1093/NAR/GKW124.

19. Huff, J.T., and Zilberman, D. (2014). Dnmt1-Independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. Cell *156*, 1286–1297. https://doi.org/10.1016/j.cell.2014.01.029.

20. Dumesic, P.A., Stoddard, C.I., Catania, S., Narlikar, G.J., and Madhani, H.D. (2020). ATP hydrolysis by the SNF2 domain of Dnmt5 is coupled to both specific recognition and modification of hemimethylated DNA. Mol. Cell *79*, 127–139.e4. https://doi.org/10.1016/J.MOLCEL.2020.04.029.

21. Tirot, L., Jullien, P.E., and Ingouff, M. (2021). Evolution of CG methylation maintenance machinery in plants. Epigenomes *5*, 19. https://doi.org/10.3390/EPIGENOMES5030019.

22. Seisenberger, S., Peat, J.R., Hore, T.A., Santos, F., Dean, W., and Reik, W. (2013). Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. Philos. Trans. R. Soc. Lond. B Biol. Sci. *368*, 20110330. https://doi.org/10.1098/RSTB.2011.0330.

23. Tucci, V., Isles, A.R., Kelsey, G., and Ferguson-Smith, A.C.; Erice Imprinting Group (2019). Genomic imprinting and physiological processes in mammals. Cell *176*, 952–965. https://doi.org/10.1016/j.cell.2019.01.043.

24. Kazachenka, A., Bertozzi, T.M., Sjoberg-Herrera, M.K., Walker, N., Gardner, J., Gunning, R., Pahita, E., Adams, S., Adams, D., and Ferguson-Smith, A.C. (2018). Identification, characterization, and heritability of murine metastable epialleles: implications for non-genetic inheritance. Cell *175*, 1717. https://doi.org/10.1016/J.CELL.2018.11.017.

25. Quadrana, L., and Colot, V. (2016). Plant transgenerational epigenetics. Annu. Rev. Genet. *50*, 467–491. https://doi.org/10.1146/ANNUREV-GENET-120215-035254.

26. Becker, C., Hagmann, J., Müller, J., Koenig, D., Stegle, O., Borgwardt, K., and Weigel, D. (2011). Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature *480*, 245–249. https://doi.org/10.1038/nature10555.

27. Schmitz, R.J., Schultz, M.D., Lewsey, M.G., O'Malley, R.C., Urich, M.A., Libiger, O., Schork, N.J., and Ecker, J.R. (2011). Transgenerational epigenetic instability is a source of novel methylation variants. Science *334*, 369–373. https://doi.org/10.1126/science.1212959.

28. Van Der Graaf, A., Wardenaar, R., Neumann, D.A., Taudt, A., Shaw, R.G., Jansen, R.C., Schmitz, R.J., Colomé-Tatché, M., and Johannes, F. (2015). Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. Proc. Natl. Acad. Sci. USA *112*, 6676–6681. https://doi.org/10.1073/pnas.1424254112.

29. Watson, J.M., Platzer, A., Kazda, A., Akimcheva, S., Valuchova, S., Nizhynska, V., Nordborg, M., and Riha, K. (2016). Germline replications and somatic mutation accumulation are independent of vegetative life span in Arabidopsis. Proc. Natl. Acad. Sci. USA *113*, 12226–12231. https://doi.org/10.1073/pnas.1609686113.

30. Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature *452*, 215–219. https://doi.org/10.1038/nature06745.

31. Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell *133*, 523–536. https://doi.org/10.1016/j.cell.2008.03.029.

32. Du, J., Johnson, L.M., Jacobsen, S.E., and Patel, D.J. (2015). DNA methylation pathways and their crosstalk with histone methylation. Nat. Rev. Mol. Cell Biol. *16*, 519–532. https://doi.org/10.1038/nrm4043.

33. Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. *11*, 204–220. https://doi.org/10.1038/nrg2719.

34. Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D. (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell *153*, 193–205. https://doi.org/10.1016/j.cell.2013.02.033.

35. Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat. Struct. Mol. Biol. *21*, 64–72. https://doi.org/10.1038/nsmb.2735.

36. Rajakumara, E., Law, J.A., Simanshu, D.K., Voigt, P., Johnson, L.M., Reinberg, D., Patel, D.J., and Jacobsen, S.E. (2011). A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. Genes Dev. *25*, 137–152. https://doi.org/10.1101/gad.1980311.

37. Du, J., Zhong, X., Bernatavichute, Y.V., Stroud, H., Feng, S., Caro, E., Vashisht, A.A., Terragni, J., Chin, H.G., Tu, A., et al. (2012). Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell *151*, 167–180. https://doi.org/10.1016/j.cell.2012.07.034.

38. Sequeira-Mendes, J., Aragüez, I., Peiró, R., Mendez-Giraldez, R., Zhang, X., Jacobsen, S.E., Bastolla, U., and Gutierrez, C. (2014). The functional topography of the Arabidopsis genome is organized in a reduced number of linear motifs of chromatin states. Plant Cell *26*, 2351–2366. https://doi.org/10.1105/tpc.114.124578.

39. Zhang, Y., Harris, C.J., 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. Proc. Natl. Acad. Sci. USA *115*, E1069–E1074. https://doi.org/10.1073/pnas.1716300115.

40. Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat. Rev. Genet. *15*, 394–408. https://doi.org/10.1038/nrg3683.

41. Pélissier, T., Thalmeir, S., Kempe, D., Sänger, H.L., and Wassenegger, M. (1999). Heavy de novo methylation at symmetrical and non-symmetrical sites is a hallmark of RNA-directed DNA methylation. Nucleic Acids Res. *27*, 1625–1634. https://doi.org/10.1093/NAR/27.7.1625.

42. Cuerda-Gil, D., and Slotkin, R.K. (2016). Non-canonical RNA-directed DNA methylation. Nat. Plants *2*, 16163. https://doi.org/10.1038/nplants.2016.163.

43. Chan, S.W.-L., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science *303*, 1336. https://doi.org/10.1126/science.1095989.

44. 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. Science *323*, 1600–1604.

45. To, T.K., Yamasaki, C., Oda, S., Tominaga, S., Kobayashi, A., Tarutani, Y., and Kakutani, T. (2022). Local and global crosstalk among heterochromatin marks drives DNA methylome patterning in Arabidopsis. Nat. Commun. *13*, 861. https://doi.org/10.1038/s41467-022-28468-5.

46. Hazarika, R.R., Serra, M., Zhang, Z., Zhang, Y., Schmitz, R.J., and Johannes, F. (2022). Molecular properties of epimutation hotspots. Nat. Plants *8*, 146–156. https://doi.org/10.1038/s41477-021-01086-7.

47. Lyons, D.B., and Zilberman, D. (2017). DDM1 and lsh remodelers allow methylation of DNA wrapped in nucleosomes. Elife *6*, e30674. https://doi.org/10.7554/eLife.30674.

48. Choi, J., Lyons, D.B., and Zilberman, D. (2021). Histone H1 prevents non-CG methylation-mediated small RNA biogenesis in Arabidopsis heterochromatin. Elife *10*, e72676. https://doi.org/10.7554/ELIFE.72676.

49. Choi, J., Lyons, D.B., Kim, M.Y., Moore, J.D., and Zilberman, D. (2020). DNA methylation and histone H1 jointly repress transposable elements and aberrant intragenic transcripts. Mol. Cell *77*, 310–323.e7. https://doi.org/10.1016/j.molcel.2019.10.011.

50. Hsieh, P.-H., He, S., Buttress, T., Gao, H., Couchman, M., Fischer, R.L., Zilberman, D., and Feng, X. (2016). Arabidopsis male sexual lineage exhibits more robust maintenance of CG methylation than somatic tissues. Proc. Natl. Acad. Sci. USA *113*, 15132–15137. https://doi.org/10.1073/pnas.1619074114.

51. Park, K., Kim, M.Y., Vickers, M., Park, J.S., Hyun, Y., Okamoto, T., Zilberman, D., Fischer, R.L., Feng, X., Choi, Y., and Scholten, S. (2016). DNA demethylation is initiated in the central cells of Arabidopsis and rice. Proc. Natl. Acad. Sci. USA *113*, 15138–15143. https://doi.org/10.1073/PNAS.1619047114.

52. He, S., Vickers, M., Zhang, J., and Feng, X. (2019). Natural depletion of H1 in sex cells causes DNA demethylation, heterochromatin decondensation and transposon activation. Elife *8*, e42530. https://doi.org/10.7554/ELIFE.42530.001.

53. Calarco, J.P., Borges, F., Donoghue, M.T.A., Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijó, J.A., Becker, J.D., and Martienssen, R.A. (2012). Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell *151*, 194–205. https://doi.org/10.1016/J.CELL.2012.09.001.

54. Tsukahara, S., Kobayashi, A., Kawabe, A., Mathieu, O., Miura, A., and Kakutani, T. (2009). Bursts of retrotransposition reproduced in Arabidopsis. Nature *461*, 423–426. https://doi.org/10.1038/nature08351.

55. Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., and Kakutani, T. (2001). Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature *411*, 212–214. https://doi.org/10.1038/35075612.

56. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295. https://doi.org/10.1038/NBT.3122.

57. Panda, K., and Slotkin, R.K. (2020). Long-read cDNA sequencing enables a "gene-like" transcript annotation of transposable elements. Plant Cell *32*, 2687–2698. https://doi.org/10.1105/TPC.20.00115.

58. Pimentel, H., Bray, N.L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. Nat. Methods *14*, 687–690. https://doi.org/10.1038/nmeth.4324.

59. Osakabe, A., Jamge, B., Axelsson, E., Montgomery, S.A., Akimcheva, S., Kuehn, A.L., Pisupati, R., Lorković, Z.J., Yelagandula, R., Kakutani, T., and Berger, F. (2021). The chromatin remodeler DDM1 prevents transposon mobility through deposition of histone variant H2A.W. Nat. Cell Biol. *23*, 391–400. https://doi.org/10.1038/S41556-021-00658-1.

60. Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W.L., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., and Ecker, J.R. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. Cell *126*, 1189–1201. https://doi.org/10.1016/j.cell.2006.08.003.

61. Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat. Genet. *39*, 61–69. https://doi.org/10.1038/ng1929.

62. Soppe, W.J.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. EMBO J. *21*, 6549–6559.

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

64. Kato, M., Takashima, K., and Kakutani, T. (2004). Epigenetic control of CACTA transposon mobility in Arabidopsis thaliana. Genetics *168*, 961–969. https://doi.org/10.1534/genetics.104.029637.

65. Bourguet, P., Picard, C.L., Yelagandula, R., Pélissier, T., Lorković, Z.J., Feng, S., Pouch-Pélissier, M.N., Schmücker, A., Jacobsen, S.E., Berger, F., and Mathieu, O. (2021). The histone variant H2A.W and linker histone H1 co-regulate heterochromatin accessibility and DNA methylation. Nat. Commun. *12*, 2683. https://doi.org/10.1038/s41467-021-22993-5.

66. Kakutani, T., Jeddeloh, J.A., and Richards, E.J. (1995). Characterization of an Arabidopsis thaliana DNA hypomethylation mutant. Nucleic Acids Res. *23*, 130–137. https://doi.org/10.1093/NAR/23.1.130.

67. Kakutani, T., Jeddeloh, J.A., Flowers, S.K., Munakata, K., and Richards, E.J. (1996). Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. Proc. Natl. Acad. Sci. USA *93*, 12406–12411. https://doi.org/10.1073/PNAS.93.22.12406.

68. Ito, T., Tarutani, Y., To, T.K., Kassam, M., Duvernois-Berthet, E., Cortijo, S., Takashima, K., Saze, H., Toyoda, A., Fujiyama, A., et al. (2015). Genome-wide negative feedback drives transgenerational DNA methylation dynamics in Arabidopsis. PLoS Genet. *11*, e1005154.

69. Papareddy, R.K., Páldi, K., Smolka, A.D., Hüther, P., Becker, C., and Nodine, M.D. (2021). Repression of CHROMOMETHYLASE 3 prevents

epigenetic collateral damage in Arabidopsis. Elife *10*, e69396. https://doi.org/10.7554/ELIFE.69396.

70. Liu, Z.-W., Shao, C.-R., Zhang, C.-J., Zhou, J.-X., Zhang, S.-W., Li, L., Chen, S., Huang, H.-W., Cai, T., and He, X.-J. (2014). The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. PLoS Genet. *10*, e1003948. https://doi.org/10.1371/journal.pgen.1003948.

71. Johnson, L.M., Du, J., Hale, C.J., Bischof, S., Feng, S., Chodavarapu, R.K., Zhong, X., Marson, G., Pellegrini, M., Segal, D.J., et al. (2014). SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature *507*, 124–128. https://doi.org/10.1038/NATURE12931.

72. Papareddy, R.K., Páldi, K., Paulraj, S., Kao, P., Lutzmayer, S., and Nodine, M.D. (2020). Chromatin regulates expression of small RNAs to help maintain transposon methylome homeostasis in Arabidopsis. Genome Biol. *21*, 251. https://doi.org/10.1186/s13059-020-02163-4.

73. Krueger, F., and Franke, A. (2012). DNA methylome analysis using short bisulfite sequencing data. Nat. Methods *9*, 145–151. https://doi.org/10.1038/nmeth.1828.

74. To, T.K., and Kakutani, T. (2022). Crosstalk among pathways to generate DNA methylome. Curr. Opin. Plant Biol. *68*, 102248. https://doi.org/10.1016/J.PBI.2022.102248.

75. Erdmann, R.M., and Picard, C.L. (2020). RNA-Directed DNA methylation. PLoS Genet. *16*, e1009034. https://doi.org/10.1371/JOURNAL.PGEN.1009034.

76. Long, J., Walker, J., She, W., Aldridge, B., Gao, H., Deans, S., Vickers, M., and Feng, X. (2021). Nurse cell-derived small RNAs define paternal epigenetic inheritance in Arabidopsis. Science *373*, eabh0556. https://doi.org/10.1126/science.abh0556.

77. Bouyer, D., Kramdi, A., Kassam, M., Heese, M., Schnittger, A., Roudier, F., and Colot, V. (2017). DNA methylation dynamics during early plant life. Genome Biol. *18*, 179. https://doi.org/10.1186/S13059-017-1313-0/FIGURES/4.

78. Kawakatsu, T., Nery, J.R., Castanon, R., and Ecker, J.R. (2017). Dynamic DNA methylation reconfiguration during seed development and germination. Genome Biol. *18*, 171. https://doi.org/10.1186/S13059-017-1251-X/FIGURES/8.

79. Zhou, M., Coruh, C., Xu, G., Martins, L.M., Bourbousse, C., Lambolez, A., and Law, J.A. (2022). The CLASSY family controls tissue-specific DNA methylation patterns in Arabidopsis. Nat. Commun. *13*, 244. https://doi.org/10.1038/s41467-021-27690-x.

80. Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., et al. (2012). Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science *337*, 1360–1364. https://doi.org/10.1126/SCIENCE.1224839.

81. Shahryary, Y., Symeonidi, A., Hazarika, R.R., Denkena, J., Mubeen, T., Hofmeister, B., Van Gurp, T., Colomé-Tatché, M., Verhoeven, K.J.F., Tuskan, G., et al. (2020). AlphaBeta: computational inference of epimutation rates and spectra from high-throughput DNA methylation data in plants. Genome Biol. *21*, 260. https://doi.org/10.1186/S13059-020-02161-6.

82. Kooke, R., Johannes, F., Wardenaar, R., Becker, F., Etcheverry, M., Colot, V., Vreugdenhil, D., and Keurentjes, J.J.B. (2015). Epigenetic basis of morphological variation and phenotypic plasticity in Arabidopsis thaliana. Plant Cell *27*, 337–348. https://doi.org/10.1105/TPC.114.133025.

83. 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. https://doi.org/10.1371/JOURNAL.PGEN.1000530.

84. Jeddeloh, J.A., Stokes, T.L., and Richards, E.J. (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat. Genet. *22*, 94–97. https://doi.org/10.1038/8803.

85. Bourguet, P., Yelagandula, R., To, T.K., Osakabe, A., Alishe, A., Lu, R.J.-H., Kakutani, T., Chen, P.-Y., and Berger, F. (2022). The histone variant H2A.W cooperates with chromatin modifications and linker histone H1 to maintain transcriptional silencing of transposons in Arabidopsis. Preprint at bioRxiv. https://doi.org/10.1101/2022.05.31.493688.

86. Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics *10*, 232. https://doi.org/10.1186/1471-2105-10-232/COMMENTS.

87. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572. https://doi.org/10.1093/BIOINFORMATICS/BTR167.

88. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

89. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527. https://doi.org/10.1038/nbt.3519.

90. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915. https://doi.org/10.1038/S41587-019-0201-4.

91. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

92. Wickham, H. (2009). Ggplot2 : Elegant Graphics for Data Analysis (Springer).

93. Collins, T.J. (2007). ImageJ for microscopy. Biotechniques *43*, S25–S30. https://doi.org/10.2144/000112517.

94. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
Epitect bisulfite conversion kit	Qiagen	E7335S
NEB Next II library preparation kit	NEB	E7645
Scriptseq RNA-seq kit	Epicentre	SSV21124
Ribo-zero kit	Illumina	MRZPL1224
NEB next indexing primers	NEB	E7335S
charge switch genomic DNA columns	Thermo-Fisher	CS11203
Epitect fast bisulfite conversion kit	Qiagen	59802
Deposited data		
ChIP-seq data	(Choi et al.)[49]	GSE179796
met1 and *h1met1* bs-seq data	(Choi et al.)[49]	GSE122394
Bisulfite and RNA-sequencing	This study	GSE197718
h1 bs-seq and nucleosome data	(Lyons and Zilberman)[47]	GSE96994
H2A.W ChIP, *ddm1* x Col-0 bs-seq, and RNA-seq	(Osakabe et al.)[59]	GSE150436
ddm1 G9 (F10) bs-seq data	(Ito et al.)[68]	DRS019614
Experimental models: Organisms/strains		
h1.1 tDNA mutant Arabidopsis thaliana	ABRC	SALK_128430C
h1.2 tDNA mutant Arabidopsis thaliana	GABI-KAT	GABI_406H11
ddm1.10 tDNA mutant Arabidopsis thaliana	ABRC	SALK_093009.30.45.x
cmt2-3 tDNA mutant Arabidopsis thaliana	ABRC	SALK_012874
cmt3-11 tDNA mutant Arabidopsis thaliana	ABRC	SALK 148381
drm2-2 tDNA mutant Arabidopsis thaliana	ABRC	SALK 150863
Software and algorithms		
bsmap	(Xi and Li)[86]	N/A
bismark	(Krueger and Andrews)[87]	N/A
bedtools	(Quinlan and Hall)[88]	N/A
kallisto	(Bray et al.)[89]	N/A
sleuth	(Pimentel et al.)[58]	N/A
hisat2	(Kim et al.)[90]	N/A
IGV	(Robinson et al.)[91]	N/A
Stringtie	(Pertea et al.)[56]	N/A
TE methylation simulation code	This study	https://doi.org/10.5281/zenodo.7308320

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Daniel Zilberman (daniel.zilberman@ist.ac.at).

Materials availability

Mutant plants generated for this study are available by request to the lead contact.

Data and code availability

- Sequencing data have been deposited at the NCBI and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- All original code has been deposited at github, accessible at https://github.com/BriffaAKR/h1ddm1_TEmeth_modelling and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The *h1ddm1* line was previously described.[47] DDM1 is a potent regulator of DNA methylation and its removal initiates the decay of mCG in *h1ddm1*. Therefore, to generate *h1ddm1* we self-fertilized *h1*/+;*ddm1*/+ and isolated *h1*;*ddm1*. To generate compound *h1ddm1* methyltransferase mutants, we crossed *h1/h1*;*ddm1*/+ (abbreviated *h1*;*ddm1*/+) to homozygous null *drm2-2* (SALK_150863C), *cmt2-3* (SALK_012874) and *cmt3-11* (SALK_148381C) mutants. We then isolated *h1*;*drm2*;*ddm1*/+ and *h1*;*cmt2*;*ddm1*/+ and self-fertilized these to generate each respective F2 *drm2* and *cmt2* compound mutant. For *cmt3* we were unable to obtain *h1*;*cmt3*;*ddm1*/+ from the initial cross. We did find *h1*;*ddm1*;*cmt3*/+ frequently however, and therefore used this to generate *h1ddm1cmt3*, but as a result were unable to generate an equivalent F2 for this line. *Arabidopsis* seeds were sown on soil, stratified for 4 days at 4 deg. C and transferred to controlled environment chambers where they were grown in 16h light/8hr dark at 20 deg. C until tissue harvest.

METHOD DETAILS

Leaf DNA isolation and bisulfite library prep

DNA was isolated by pulverizing ~0.5 g of flash frozen rosette leaves of 4 week post-germination plants. ~100 mg of resulting powder was used to extract genomic DNA (gDNA) with the DNeasy plant mini kit (Qiagen cat. no. 69104) per manufacturer's instructions. gDNA was subsequently sonicated with the Bioruptor Pico (Diagenode) to ~250 bp median fragment length using 10 cycles of 30 s on and off. Agencourt Ampure beads (referred to as ''beads'' henceforth, cat. no. A63881) were then used at 2X volume to purify the sheared DNA. Following ligation of methylated Truseq sequencing adapters (Illumina) to sheared DNA, bisulfite conversion of DNA was carried out according to manufacturer's protocol (Qiagen Epitect Kit, cat. no. 59104) except without using carrier RNA. DNA was purified twice with 1.2X beads and converted a second time to ensure complete bisulfite conversion of unmethylated cytosine. Libraries were constructed using NEBnext kits (NEB cat. no. E7645) or Nugen/Tecan Ovation Ultralow (cat. no. 0344NB-08) following the manufacturer's instructions. NEB next indexing primers (cat. no. E7335S) were used for generating multiplexed libraries during PCR amplification.

Sperm DNA isolation and bisulfite library prep

Open flowers of F4 or F5 plants (which is here denoted as F4 or F5 sperm, respectively) were collected for pollen isolation in Galbraith buffer (45 mM MgCl2, 30 mM sodium citrate, 20 mM MOPS, 1% Triton X-100, pH7.0) by vortexing at 2000 rpm for 3 min. Flower parts were removed by straining through a 40 micron filter. Pollen grains were obtained by centrifugation at 2600 g for 5 min and broken down with glass beads (Sigma). The lysate was then transferred to a 40 micron cell strainer and the flow-through was centrifuged at 800 g for 10 min at 4°C. The pellet was re-suspended in Galbraith buffer and stained with SYBR Green for fluorescence-activated cell sorting (FACS). Vegetative nuclei and sperm nuclei (SN) were separated and collected based on size and fluorescence intensity. DNA was extracted from SN with ChargeSwitch gDNA Micro Tissue Kit (ThermoFisher cat. no. CS11203) following the manufacturer's instructions. Briefly, SN were lysed in lysis buffer with proteinase K provided by the kit in a 55°C water bath overnight. RNA was digested with RNase A at room temperature for 30 min, then DNA was captured with magnetic beads provided, washed, and eluted. After quantification by fluorometry, about 10–20 ng of DNA was used for bisulfite-sequencing library preparation with Ovation Ultralow Methyl-Seq DR Multiplex System (Nugen part no. 0336) following the manufacturer's instructions. Briefly, DNA was sonicated with Bioruptor sonicator, end repaired, ligated to adapters which contain indexes, and bisulfite converted twice with EpiTect Fast DNA Bisulfite Kit (Qiagen 59802) following the manufacturer's instructions. After PCR amplification, libraries were purified with Agencourt RNAClean XP Beads.

Leaf RNA isolation and library preparation

Leaves (as above) from the F3 generation were flash frozen in liquid nitrogen, pulverized with mortar and pestle on dry ice, and the resulting material was subjected to vortexing in Trizol (Invitrogen, cat. no. 15596–026). Chloroform was then added at one-fifth the total volume and further vortexing was carried out until the solution appeared homogeneous. RNA was subsequently pelleted in ice-cold isopropanol. The resuspended RNA was subjected to rRNA removal with Ribo-zero plant kit (Illumina, MRZPL1224) according to the manufacturer's protocol. 50ng ribo-depleted RNA was used for library preparation with the Scriptseq kit (Epicentre, cat. no. SSV21124) following the manufacturer's protocol but with the following modifications: the RNA fragmentation step was extended to 10 min, and the temperature was increased to 90°C.

Sequencing

For *h1ddm1* bisulfite libraries and some control and compound *h1ddm1* methyltransferase libraries, we used the Illumina HiSeq 2500 at the QB3 Vincent Coates Genomic Sequencing Lab at UC Berkeley. *h1ddm1* compound mutants were mostly sequenced on the NextSeq 500 (Illumina) at the John Innes Center. Sperm bisulfite libraries were sequenced at the Bauer Core Facility at Harvard University with Illumina HiSeq 2000.

Short read mapping and quantification

Bisulfite libraries were mapped to the genome with BSMAP[86] for all analyses, except single-read analyses, for which we used Bismark.[87] BSMAP output was converted to per-base methylation scores with BSMAP's methratio.py script. RNA was mapped to the genome with Hisat2,[90] using default settings except "dta" was on. Transcripts were assembled using Stringtie[56] using Araport11 (March 2021 release; available as Araport11_GTF_genes_transposons.Mar202021.gtf at www.aradibopsis.org) as a guide, as we noticed that *ddm1* mutant transcripts added in this release[57] were also present in our *ddm1* data. For subsequent RNA quantification, we used our merged assembled transcripts to generate indices for use with kallisto pseudomapping[89] using the following settings for the quant program: –single –fr-stranded -b 100 -l 320 -s 30.

Visualization of experimental data

Methratio.py output was converted to GFF using awk commands and further processed and analyzed using commands outlined in https://github.com/dblyons/modeling_h1ddm1/.

Kallisto output was fed into the R environment for processing and fitting with the Sleuth software package,[58] which was used to perform likelihood ratio tests to derive q-values (i.e. FDR-adjusted p values) for each TE per genotype comparison, as outlined here: https://rawgit.com/pachterlab/sleuth/master/inst/doc/intro.html, and as specifically invoked here: https://github.com/dblyons/modeling_h1ddm1/blob/main/kallisto.R.

The ends-analysis.pl script (https://zilbermanlab.net/dzlab-tools-1-5-81-linux-tar/) was used to generate enrichment score matrices of mapped bisulfite data around genomic features of interest with the following settings invoked: -b 50 -d 2500 -s 6. These matrices were imported to R (http://www.R-project.org) for further processing and visualization using the code at https://github.com/dblyons/modeling_h1ddm1/blob/main/plot_ends.R. Both base functions and the ggplot2 library[92] were used for all plots except for heatmaps in Figures 1, 2 and S1, and genome browser images. Heatmaps were generated with pheatmap (https://cran.r-project.org/web/packages/pheatmap/index.html). Built-in R function 'cor' and 'cor.test' were used for calculating Pearson and Spearmann correlations. Genome tracks are screenshots of our data were displayed in Integrated Genome Viewer.[91] For RNA-seq, we used bedtools genomecov[88] with scaling on to convert Hisat2 bam output to scaled browser tracks with units in reads per million mapped (RPM). For determining MET1-dependency of mCHH, we used the same method as in ref. [48] such that: TE mCHH > =5% in WT and > =5% in *met1* is MET1-independent, while TEs fitting the following criteria are MET1-dependent: TE mCHH > =5% in WT and <2% in *met1*.

Plant leaf area measurements

Flats of the four different genotypes of plants were arranged on a platform to align images. Each plant was segmented manually using ImageJ,[93] and then analyzed using ImageJ's mask feature to isolate leaves; the measure feature was used to measure all visible leaf area. Photos were taken using a Canon Rebel T3i.

Analysis of data from Osakabe et al

To determine whether DNA methylation changes occur at the promoters of TE genes that are significantly repressed in the *ddm1* x Col-0 hybrid, we used the expression values provided in source data for Figure 5 of ref. [59] to group TE genes as downregulated (inactive) or similarly expressed (neutral) in the *ddm1* x Col-0 hybrid F1. For mapping of methylation at these groups of TEs shown in Figure 3, we calculated the expected methylation in the hybrid as the mean of the parental lineages (Col-0 + *ddm1*)/2 and used SRR13354021 as the observed F1 hybrid methylation data (Col-0 x *ddm1*). These data were processed with ends_analysis.pl, as outlined above. For comparison of *ddm1* H2A.W with expressed TEs, we used H2A.W ChIP-seq processed data from GSE150436, an analysis which essentially replicated the findings shown in Figure 3C.[59]

Transposable element annotations

Heterochromatic TEs used for methylome analysis here are the same as in ref. [49]. For *ddm1* mutant transcriptional analysis shown in Figure 3, we use the assembled transcripts from our analysis outlined above, which closely matches that of the March 2021 Araport11 gene and transposon release. This merged transcript annotation, as well as genome coordinate files of the modeled TEs, and upregulated transcripts in *ddm1* are available at www.github.com/dblyons/annotations_2022.

QUANTIFICATION AND STATISTICAL ANALYSIS

Mathematical model for TE methylation dynamics: Constructing the recursion relations

We model the transgenerational methylation dynamics of CG sites within TEs. Each CG site is defined to be in one of three states: fully unmethylated, hemi-methylated or fully methylated. Recursion relations are used to express the methylation status of the TE at the end of each cell cycle such that:

$U^{(n)}$ = fraction of sites unmethylated at the end of the n^{th} cell cycle

$H^{(n)}$ = fraction of sites hemi-methylated at the end of the n^{th} cell cycle

$M^{(n)}$ = fraction of sites methylated at the end of the n^{th} cell cycle where $U^{(n)} + H^{(n)} + M^{(n)} = 1$.

As each CG site contains two cytosines, our model relates to the experimentally measured methylation level, $\langle mC \rangle$, via $\langle mC \rangle = \frac{2M^{(n)} + H^{(n)}}{2M^{(n)} + 2H^{(n)} + 2U^{(n)}} = \frac{2M^{(n)} + H^{(n)}}{2(M^{(n)} + H^{(n)} + U^{(n)})}$.

To derive the recursion relations, we consider three key processes during the cell cycle: a) replication, b) maintenance methylation and c) *de novo* methylation and assign a probability for the various possible transitions between states. We do not explicitly include active demethylation in the model (discussed further in limitations of the model). In *met1* mutants, maintenance is completely compromised, whereas as in the various *h1ddm1* mutants, it is only partially compromised. We therefore consider these two cases separately beginning with the *h1ddm1* mutants.

Replication occurs in a well-defined and relatively narrow time-window of the cell cycle. For the purposes of our model, we define replication to be the start of a new cell cycle, as this is the point that new DNA is synthesized. Instead of modeling the top and bottom strands of DNA specifically, we make the standard assumption that 50% of hemi-methylated sites become unmethylated upon replication.

a) Replication:

$$P(M \rightarrow H) = 1$$

$$P(H \rightarrow H) = \frac{1}{2}$$

$$P(H \rightarrow U) = \frac{1}{2}$$

$$P(U \rightarrow U) = 1.$$

After replication, maintenance methylation by MET1 occurs semi-conservatively such that $H \rightarrow M$. We assume that maintenance occurs both rapidly and very efficiently after replication. Indeed, for all various *h1ddm1* mutants, we find efficient maintenance to hold self-consistently (see below). We therefore re-write r_{maint}, the contribution from the MET1-maintenance pathway, as $r_{maint} = 1 - \epsilon$ where $0 \leq \epsilon \ll 1$ represents the probability of maintenance failure. This gives:

b) Maintenance:

$$P(M \rightarrow M) = 1$$

$$P(H \rightarrow M) = 1 - \epsilon$$

$$P(H \rightarrow H) = \epsilon$$

$$P(U \rightarrow U) = 1.$$

Finally, we consider the *de novo* pathways and define $P(U \rightarrow H) = 2\, r_{novo} = \delta$, where the factor of 2 is included as there are two possible cytosine targets in each CG site. In the mutants that we study, $0 < \delta \ll 1$ is also found to hold self-consistently (see below).

We neglect the possibility that *de novo* methylation could occur twice at the same CG site (i.e., a $U \rightarrow M$ transition) within a single cell cycle, as this process is of order δ^2. We also neglect the possibility that a *de novo* event could follow a maintenance failure, as this composite process is of order $\epsilon\delta$. The above assumptions are consistent with *de novo* methylation happening either concurrently with maintenance, or as an extended process throughout the cell cycle.

c) *De novo*:

$$P(M \rightarrow M) = 1$$

$$P(H \rightarrow H) = 1$$

$$P(U \rightarrow H) = \delta$$

$$P(U \rightarrow U) = 1 - \delta.$$

Combining the above three processes then generates:

$$\begin{aligned} M^{(n+1)} = {} & M^{(n)} \times P(M \rightarrow H)_{rep.} \times P(H \rightarrow M)_{maint.} \times P(M \rightarrow M)_{novo} \\ & + H^{(n)} \times P(H \rightarrow H)_{rep.} \times P(H \rightarrow M)_{maint.} \times P(M \rightarrow M)_{novo} \end{aligned}$$

$$\begin{aligned} H^{(n+1)} = {} & M^{(n)} \times P(M \rightarrow H)_{rep.} \times P(H \rightarrow H)_{maint.} \times P(H \rightarrow H)_{novo} \\ & + H^{(n)} \times P(H \rightarrow H)_{rep.} \times P(H \rightarrow H)_{maint.} \times P(H \rightarrow H)_{novo} \\ & + H^{(n)} \times P(H \rightarrow U)_{rep.} \times P(U \rightarrow U)_{maint.} \times P(U \rightarrow H)_{novo} \\ & + U^{(n)} \times P(U \rightarrow U)_{rep.} \times P(U \rightarrow U)_{maint.} \times P(U \rightarrow H)_{novo} \end{aligned}$$

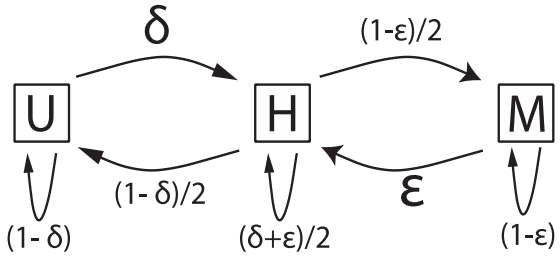
$$\begin{aligned} U^{(n+1)} = {} & H^{(n)} \times P(H \rightarrow U)_{rep.} \times P(U \rightarrow U)_{maint.} \times P(U \rightarrow U)_{novo} \\ & + U^{(n)} \times P(U \rightarrow U)_{rep.} \times P(U \rightarrow U)_{maint.} \times P(U \rightarrow U)_{novo}, \end{aligned}$$

leading to the recursion relations (see figure below):

$$M^{(n+1)} = (1 - \epsilon)\left(M^{(n)} + H^{(n)} \Big/ 2\right)$$

$$H^{(n+1)} = \epsilon\left(M^{(n)} + H^{(n)} \Big/ 2\right) + \delta\left(H^{(n)} \Big/ 2 + U^{(n)}\right)$$

$$U^{(n+1)} = (1 - \delta)\left(H^{(n)} \Big/ 2 + U^{(n)}\right).$$



Using $U^{(n)} + H^{(n)} + M^{(n)} = 1$ to eliminate $H^{(n)}$ gives:

$$M^{(n+1)} = (1 - \epsilon)\left(1 + M^{(n)} - U^{(n)}\right) \Big/ 2$$

$$U^{(n+1)} = (1 - \delta)\left(1 - M^{(n)} + U^{(n)}\right) \Big/ 2.$$

This pair of equations can be rescaled and added to provide:

$$\frac{M^{(n)}}{(1 - \epsilon)} + \frac{U^{(n)}}{1 - \delta} = 1,$$

which allows the recursion relations to be expressed in terms of a single variable only, where we retain only terms linear in ϵ and δ:

$$M^{(n+1)} = \frac{\delta}{2} + \left(1 - \frac{\delta + \epsilon}{2}\right)M^{(n)} \qquad \text{(Equation 1)}$$

$$U^{(n+1)} = \frac{\epsilon}{2} + \left(1 - \frac{\delta + \epsilon}{2}\right)U^{(n)}.$$

Next, we consider the *met1* mutation. The above derivation is unchanged for replication. We assume complete maintenance failure, so that $\epsilon = 1$. As previously, we neglect processes of order δ^2 within a single cell cycle, and later find that $0 < \delta \ll 1$ also holds self-consistently for *met1* mutants. Applying these modifications provides the following recursion relations for $n > 0$:

$$M^{(n)} = 0$$

$$H^{(n+1)} = \frac{H^{(n)}}{2} + \delta\left(\frac{H^{(n)}}{2} + U^{(n)}\right)$$

$$U^{(n+1)} = (1 - \delta)\left(H^{(n)} \Big/ 2 + U^{(n)}\right).$$

Using $H^{(n)} + U^{(n)} = 1$ then gives:

$$H^{(n+1)} = \frac{1 - \delta}{2}H^{(n)} + \delta$$

$$U^{(n+1)} = \frac{1 - \delta}{2}U^{(n)} + \frac{1 - \delta}{2}.$$

Mathematical model for TE methylation dynamics: steady state solutions

At steady state, we define $M^{(n+1)} = M^{(n)} = M^*$, and equivalently for H and U. For the various *h1ddm1* mutants, we then find that to lowest order

$$M^* = \frac{\delta}{\delta + \epsilon}$$

$$H^* = 0$$

$$U^* = \frac{\epsilon}{\delta + \epsilon}.$$

For all mutants considered here, except for *met1*, the values we extract from the data indeed give $\epsilon, \delta \ll 1$, thereby self-consistently justifying the assumptions made in our derivation. Hence,

$$\langle mC \rangle^* = \frac{H^* + 2M^*}{2(M^* + H^* + U^*)} = M^* = \frac{\delta}{\delta + \epsilon}, \quad \text{(Equation 2)}$$

Whereas, for *met1* mutants, to lowest order:

$$M^* = 0$$

$$H^* = 2\delta$$

$$U^* = 1 - 2\delta,$$

for which,

$$\langle mC \rangle^* = \frac{H^* + 2M^*}{2(M^* + H^* + U^*)} = \frac{H^*}{2} = \delta. \quad \text{(Equation 3)}$$

Mathematical model for TE methylation dynamics: Time dependent solutions

Using a standard approach to solve difference equations, the recursion relation in Eq. (1) can be solved to express $M^{(n)}$ as a function δ, ϵ and the initial fraction of methylation, M^0:

$$M^{(n)} = \left(M^0 - \frac{\delta}{\delta+\epsilon}\right) e^{-\left|\ln\left(1-\frac{\delta+\epsilon}{2}\right)\right| n} + \frac{\delta}{\delta+\epsilon} = \left(M^0 - \frac{\delta}{\delta+\epsilon}\right) e^{-\left(\frac{\delta+\epsilon}{2}\right) n} + \frac{\delta}{\delta+\epsilon},$$

for $0 < \delta \ll 1$ and $0 < \epsilon \ll 1$, and therefore where we again retain only leading order terms. To allow comparison with the experimentally measured methylation level, the above exponential decay must be expressed as a function of the number of plant generations, x, and the number of cell cycles through the germline per plant generation, such that $n = n_{cc}x$. The experimentally measured methylation level then becomes:

$$\langle mC(x) \rangle = M^{(n)} = \left(M^0 - \frac{\delta}{\delta+\epsilon}\right) e^{-\left(\frac{\delta+\epsilon}{2}\right) n_{cc} x} + \frac{\delta}{\delta+\epsilon} \quad \text{(Equation 4)}$$

with the free parameters: δ, the *de novo* methylation rate for each unmethylated CG site per cell cycle, ϵ, the maintenance failure rate for each CG site per cell cycle, and M^0, the initial methylation level before the system was perturbed away from steady state.

Mathematical model for TE methylation dynamics: gain and loss rates per cell cycle

For the full $M \rightarrow U$ transition (i.e., a loss) to occur, a minimum of two replication steps must occur. Therefore, to calculate the first order (in δ and ϵ) contribution to the overall loss (and gain) rates, we follow a site over two full cell cycles. This is still a per cell cycle rate, as the change is begun in the stated cell cycle, although the process is not completed until the next cell cycle:

$$\begin{aligned} P(M \rightarrow U) &= P(M \rightarrow H)_{replication} \times P(H \rightarrow H)_{maintenance} \times P(H \rightarrow H)_{de\ novo} \times P(H \rightarrow U)_{replication} \times P(U \rightarrow U)_{maintenance} \times P(U \rightarrow U)_{de\ novo} \\ &= (1)(\epsilon)(1)(1/2)(1)(1-\delta) \\ &\approx \epsilon/2, \end{aligned}$$

and similarly,

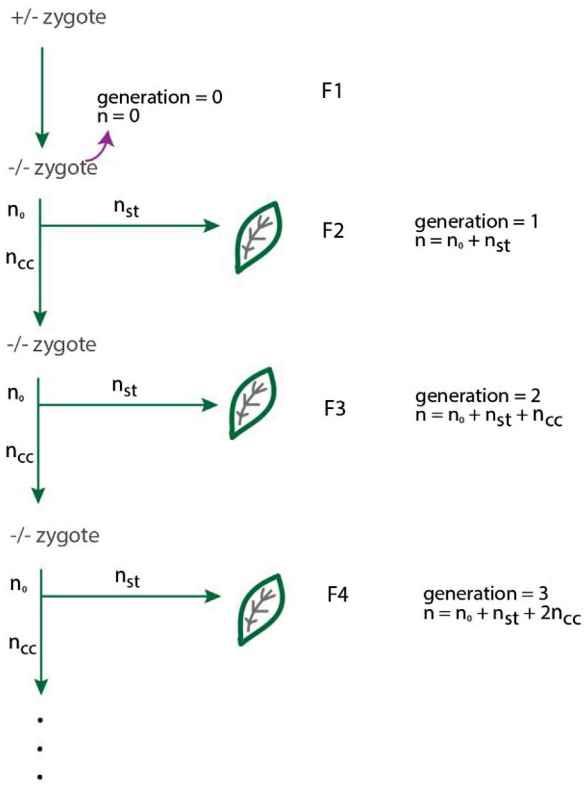
$$\begin{aligned} P(U \rightarrow M) &= P(U \rightarrow U)_{replication} \times P(U \rightarrow U)_{maintenance} \times P(U \rightarrow H)_{de\ novo} \times P(H \rightarrow H)_{replication} \times P(H \rightarrow M)_{maintenance} \times P(M \rightarrow M)_{de\ novo} \\ &= (1)(1)(\delta)(1/2)(1-\epsilon)(1) \\ &\approx \delta/2, \end{aligned}$$

as illustrated in Figure 5A. Note that these rates differ from those derived in the recursion relations above, as the latter are strictly over a single cell cycle.

Relationship between the mathematical model and bisulfite sequencing data from leaf tissue

Our model is applied to individual heterochromatic TEs with a length >2 kb (n = 1910) and a *WT* methylation level $M_{WT} > 0.65$ for both replicates (leading to the exclusion of only 27 TEs). This results in 1883 TEs which we fit using the model.

The transgenerational bisulfite sequencing data of leaf tissues show decreasing methylation levels as a function of time for the following mutants: *h1ddm1*, *h1ddm1cmt2*, *h1ddm1cmt3* and *h1ddm1drm2*. We fit these methylation timeseries to the decaying exponential in Eq. (4) (the numerical details are explained in the following section). Consequently, we need to know the effective timepoints at which the leaf-tissue sequencing data is obtained, as described and illustrated schematically below.



We assume that the heterozygous plant methylation level is at its steady state and define $n = 0$ to occur when the homozygous offspring is formed, coinciding with (we assume) an abrupt potential change in the values of δ and ϵ. The changed δ and ϵ values will now correspond to a different steady-state methylation level. The model predicts that the methylation level of the mutant plant will decay exponentially with time toward this new value. To fit the experimental data to this exponential decay, we first need to know the approximate number of replication events between data points.

Following Watson et al.,[29] we assume a constant number of cell divisions through the cell lineage leading to the generative lineage ("germline"), $n_{cc} = 34$. We also assume a constant number of somatic cell divisions to produce the sequenced leaf tissue: n_{st}, this value, however, is unknown. Under these assumptions, for plants of generation F2 and later, there is a uniform number of cell divisions between each generation of $\Delta n = n_{cc}$. The heterozygous plant, however, is more problematic as it is assumed to stay at its original steady-state methylation level up until the formation of the homozygous zygote. The F2 generation is therefore separated by $\Delta n = n_0 + n_{st}$, where n_0 is the number of replications before the leaf tissue branches from the germline, both values of which are unknown.

The data points for early generations are particularly valuable as the methylation level changes most rapidly between these. We treat biological replicates of the same generation as multiple data points with the same time-value. As an initial test of the model, we first fit to the methylation level averaged over all the selected TEs (Figure 5B). Here all generations from F2 onwards are included. We find excellent fits to both the *h1ddm1* and the *h1ddm1drm2* datasets, along with a good fit for *h1ddm1cmt2*. However, in the latter case, the replicates have a greater spread, increasing the fit uncertainty. Furthermore, for *h1ddm1cmt3*, the important F2 generation is not available significantly reducing our confidence in this fit.

To partly overcome this difficulty, we use our other datasets to generate an extra effective datapoint corresponding to the *WT* plant. As demonstrated by the fit to the average mCG level for *h1ddm1* F2 to F11 replicates (fit without *WT*: Figure 5B solid line), the mCG level clearly decays from a high initial value at $x = 0$ ($\sigma_{fit} = 0.011$). By projecting the *WT* leaf methylation level onto the *h1ddm1* fit without WT, we find (coincidentally) that the *WT* leaf methylation level corresponds to $x = 0$ (to within the fit uncertainty). For each individual TE, therefore, we approximate the initial methylation level (at $x = 0$) by the two *WT* leaf replicates and confirm that the fit to average methylation levels when including these two extra effective datapoints (fit with *WT*: Figure 5B dashed line) is highly

consistent to the fit without *WT*. We then fit to the methylation timeseries for individual TEs to generate distributions of δ and ϵ values for each mutant (described in detail below). We confirmed that fits excluding the *WT* leaf replicates at $x = 0$ provide comparable distributions, however, manual inspection revealed these to provide less reliable estimates of the decay constant than the fits with the *WT* leaf replicates included (due to some fits generating anomalously low initial methylation levels and consequently unreliable values for the decay constant). All presented fits, therefore, include the *WT* replicates at $x = 0$ unless otherwise stated. For *met1* mutants the methylation decay is too rapid to be accessible in the transgenerational data. Only the steady-state methylation level is available, which corresponds directly to δ (see Eq. (3)).

Mathematical model for TE methylation dynamics: Numerical fits

The methylation timeseries, $\langle mC\rangle(x)$, are fit using *curve_fit* from the *optimize* library in *scipy*.[94] First, we performed fits using three free parameters: M_0, M^* (both as defined previously), and B, the decay constant:

$$\langle mC(x)\rangle = (M_0 - M^*)e^{-Bx} + M^*$$

and the total fit uncertainty, σ_{fit}, is estimated using the sum of the diagonal of the covariance matrix:

$$\sigma_{fit} = \sigma_{M_0}^2 + {\sigma_B}^2 + {\sigma_{M^*}}^2.$$

The fit parameters were constrained to $0 < M_0 < 1, 0 < M^* < 1$ and $B > 0$, while using initial values of: $M_0 = 0.8, B = 1$ and $M^* = 0.3$. We do not fit to δ and ϵ directly, as the numerical routine is more reliable when fitting to parameters with an order of magnitude close to one. As $B = n_{cc}(\delta + \epsilon)/2$, we reject any fits for which $B > n_{cc}$ as this corresponds to an unphysical value for δ or ϵ. The majority of TEs produce good fits and a small minority give exceptionally poor fits. To be conservative, we reject all fits with $\sigma_{fit} > 1$. A reliable estimate of the decay constant requires an appreciable decrease in the methylation level with time. We therefore exclude TEs with $\langle mCG\rangle_{\{F2,F3\}} \leq \lambda M^*$ for *h1ddm1*, *h1ddm1cmt2* and *h1ddm1cmt3* or $\langle mCG\rangle_{\{F3,F4\}} \leq \lambda M^*$ for *h1ddm1cmt3* using a threshold of $\lambda = 1.05$, where $\langle mCG\rangle_{\{F2,F3\}}$ is the average methylation level of all F2 and F3 replicates for a particular mutant. Upon manual inspection of the remaining fits, we also excluded nine further TEs. Finally, we selected the subset of TEs with acceptable fits for all four mutants (amounting to 1617 TEs, over 85% of the initial set that we attempted to fit).

The values of δ and ϵ are then given by:

$$\delta = \frac{2BM^*}{n_{cc}}$$

$$\epsilon = \frac{2B(1 - M^*)}{n_{cc}}.$$

For a function $f(x_1, x_2, x_3)$ the uncertainty in f is given by

$$\sigma_f^2 = \boldsymbol{v}^T \bullet \boldsymbol{\sigma} \bullet \boldsymbol{v}, \qquad \text{(Equation 5)}$$

where $\boldsymbol{\sigma}$ is the covariance matrix for x_1, x_2 and x_3 and $\boldsymbol{v}$ is a vector with elements $v_i = \frac{\partial f}{\partial x_i}$. The uncertainties in the fitted values of δ and ϵ are therefore:

$$\sigma_\delta = \frac{2BM^*}{n_{cc}}\sqrt{\left(\frac{\sigma_B}{B}\right)^2 + \left(\frac{\sigma_{M^*}}{M^*}\right)^2 + \frac{2\sigma_{BM^*}}{BM^*}}$$

$$\sigma_\epsilon = \frac{2B(1 - M^*)}{n_{cc}}\sqrt{\left(\frac{\sigma_B}{B}\right)^2 + \left(\frac{\sigma_{M^*}}{1 - M^*}\right)^2 - \frac{2\sigma_{BM^*}}{B(1 - M^*)}},$$

where σ_{BM^*} is the covariance of B and M^*.

Two-sided paired t-tests were performed for all mutant pairings of ϵ-distributions (p value $< 1\times10^{-170}$ for all pairings) and, similarly, for all mutant pairings of δ-distributions p value $< 1\times10^{-200}$ for all pairings). Finally, we note that the presence of the small peak at low δ-values (Figure 7F) is likely an artifact arising for TEs with a small decay constant. Due to their slow approach to steady-state, the fit algorithm may underestimate the precise value of M^*, and hence δ.

Mathematical model for TE methylation dynamics: Limitations of the model

We have assumed that the *de novo* rate, δ, is independent of the existing methylation level. This may not always be the case. For example, a decrease in RdDM targeting as more methylation is lost is possible, as appears to be the case for a small subset of TEs (Figures S6F-S6H). Alternatively, there may be an analogous cooperative *de novo* pathway to that proposed for clustered CG sites as found in CG islands of mammalian genomes,[18] in which case δ would represent the average *de novo* rate across the whole

TE. Due to the cooperativity, the effective δ could then reduce across generations as the overall methylation level declines. In the case of small variations of the *de novo* rate over successive generations, the model will still produce a good fit to a decaying exponential with effective rates for both δ and ϵ. A much larger variation in δ as a function of time, however, would likely result in a clearly visibly non-exponential decay of the methylation level, which is not observed in our data. Similarly, the loss rate, ϵ, could potentially vary with time. For example, if (as discussed above), RdDM is upregulated during embryonic development, this would generate an epsilon varying periodically as a function of the developmental stage. Our model would then simplify these dynamics with a developmentally-averaged value of ϵ.

A further issue is that the values of δ and ϵ could each be different in the germline versus in somatic tissue (e.g., in the leaf). Our model assumes that these values are the same. Experimental data indicates that F4 and F5 mCG levels in the *h1ddm1* leaf and sperm are similar (Figure 2A), which supports similar values of δ and ϵ in the germline and somatic tissue.

The model also assumes that all CG sites within a given TE are equivalent and all exhibit the same dynamical turnover of methylation, parameterized through δ and ϵ. However, as is apparent from the per-CG site methylation histograms (Figures 1F, S1I), a minority of CG sites are always highly methylated or always exhibit a near-absence of methylation, configurations that are essentially excluded within our model. Future model developments may need to better incorporate this heterogeneity.

We do not explicitly include active demethylation in the model. $M \rightarrow H$ transitions are assumed to be dominated by passive methylation loss at replication, followed by a maintenance failure. However, the $M \rightarrow H$ transition rate, ϵ, could also include a direct contribution from active demethylation. Secondly, there is the possibility that a demethylase acts on a hemi-methylated H resulting in a $H \rightarrow U$ transition. However, all complete $M \rightarrow U$ transitions begin with an initial $M \rightarrow H$ (via replicative loss followed by maintenance failure or via a demethylase), which is of order ϵ and which we find self-consistently to be small compared to the 100% rate of M → H transitions caused by DNA replication. Hence, provided the active demethylation $H \rightarrow U$ rate is also similarly small (also true self-consistently if active demethylation of H sites occurs at a similar rate to that at M sites), then this contribution will be of second order and can thus be neglected. Consequently, even though the model does not explicitly contain active demethylation, it can be included through a simple redefinition of the parameter ϵ to encompass contributions from both maintenance failure and active demethylation. All our results and in particular our estimates of the *de novo* gain rate, δ are therefore unaffected by the explicit omission of active demethylation.

Statistical analysis of empirical data

Spearman's correlation and associated p values were calculated with cor.test in base R (Figure S1G). Pearson's correlation coefficients were calculated with base R function cor (Figures 1D and 1G, S1J, 2E, 3G, 3J, 5C-5D, S5A, S5C-S5D, 7E, 7I-7L, S7A-S7C). Likelihood ratio tests were used via Kallisto/Sleuth[58,89] to derive q-values (i.e. FDR-adjusted p values) for each expression data shown in Figures 3A–3G. Significance of plant phenotypes was assessed with two-sample unpaired t-tests performed in R using the t.test base package (Figure 4D). Change in gene mCHG hypermethylation was also assessed with a paired t test (Figure 4A). TE hypomethylation in the CG context of *h1ddm1cmt3* was calculated using Fisher's exact test (Figures 6C and S6D). For additional statistical analysis shown in Figures 6D, 6G, S6E, and 7G we used the paired t.test function in base R.