RdDM-independent de novo and heterochromatin DNA methylation by plant CMT and DNMT3 orthologs

Rafael Yaari1, Aviva Katz1, Katherine Domb1, Keith D. Harris1, Assaf Zemach1 & Nir Ohad1,2

To properly regulate the genome, cytosine methylation is established by animal DNA methyltransferase 3s (DNMT3s). While altered DNMT3 homologs, *Domains rearranged methyltransferases* (DRMs), have been shown to establish methylation via the RNA directed DNA methylation (RdDM) pathway, the role of true-plant DNMT3 orthologs remains elusive. Here, we profile de novo (RPS transgene) and genomic methylation in the basal plant, *Physcomitrella patens*, mutated in each of its *PpDNMTs*. We show that *PpDNMT3b* mediates CG and CHH de novo methylation, independently of *PpDRMs*. Complementary de novo CHG methylation is specifically mediated by the CHROMOMETHYLASE, *PpCMT*. Intragenomically, *PpDNMT3b* functions preferentially within heterochromatin and is affected by *PpCMT*. In comparison, *PpDRMs* target active-euchromatic transposons. Overall, our data resolve how DNA methylation in plants can be established in heterochromatin independently of RdDM; suggest that DRMs have emerged to target euchromatin; and link DNMT3 loss in angiosperms to the initiation of heterochromatic CHH methylation by CMT2.
NA methylation, the addition of a methyl group to a cytosine base, is a prominent epigenetic modification in many eukaryotes. It is catalyzed by distinct DNA methyltransferase (DNMT) families of proteins that share a conserved methyl-transferase domain (MTD). In plants, DNMTs evolved to methylate cytosines located in specific contexts (CG, CHG, and CHH; H=A, C, or T), distinct genetic elements (e.g., transposons and genes), various chromatin configurations (hetero-chromatin and eu-chromatin), as well as to establish methylation de novo at unmethylated sites or to maintain methylation upon DNA replication. Plants encode four types of DNMTs: Methyltransferase 1 (MET1), DNA methyltransferase 3 (DNMT3), chromomethylase (CMT), and domain rearranged methyltransferase (DRM). MET1s are homologs of mammalian DNMT1 and maintain de novo CG methylases and in specific tissues also at CH sites. DNMT3s are ancient DNMTs that exist in animals, plants, and other eukaryotes. Mammalian DNMT3s function primarily as de novo CG methylases and in specific tissues also at CH sites. However, despite their significant role in mammals, non-animal DNMT3s have not been investigated thus far. DNMT3s were overlooked in plants probably due to their deficiency in angiosperms (flowering plants) and the discovery of their close homologs, DRMs, which function in de novo methylation. DRMs are plant specific DNMTs with a rearranged DNMT3-MTD. Angiosperm DRMs are a part of the RNA directed DNA methylation (RdDM) pathway that utilizes small RNA to establish de novo methylation within euchromatic transposons, that is enriched with active histone marks such as H3K4me3 and depleted of repressive marks as H3K9me2. The function of plant DNMTs was comprehensively investigated in Arabidopsis thaliana and partially explored in a few additional angiosperms, of which lack DNMT3 in their genomes.

Here, we investigate both de novo (transgene) and maintenance (whole genome) DNA methylation activities in the early divergent land plant, Physcomitrella patens (P. patens), which encodes four types of plant DNMTs, including two DNMT3s. These experiments reveal unique biosynthetic methylation mechanisms of plant DNMTs, which postulate the selection forces leading to the appearance and disappearance of specific methylation pathways during plant evolution.

**Results**

**Plant DNMT3s are evolutionary distinct from DRMs.** P. patens encodes two DNMT3s, designated here as PpDNMT3a and PpDNMT3b, which are composed of a DNMT3-type N-terminal MTD and a C-terminal domain of unknown function (DUF3444). Our genome and transcriptome searches revealed that this protein organization is conserved among non-flowering streptophytes DNMT3s (Supplementary Fig. 1). The existence of two full-length DNMT3 homologs (Supplementary Fig. 1) in two distantly-related gymnosperm subclasses that were separated around 300 million years ago implies upon the persistence of DNMT3 in gymnosperms. We did not detect DNMT3 in any available angiosperm genomes or transcriptomes, supporting the notion that DNMT3 completely disappeared from this plant lineage. Phylogenetic analysis of the MTD showed that plant DNMT3 form a monophyletic clade together with animal DNMT3 which is separated from the DRM clade (Fig. 1a), suggesting the functional conservation of DNMT3s among plants and animals and/or functional speciation between plant DNMT3 and DRM proteins. Additionally, while DRM paralogs are common among plant evolution, they diverged into distinct orthologs only in seed plants, e.g., DRM2 and DRM3 in angiosperm (Fig. 1a), implying on further functional diversification of DRMs in this plant lineage. Paralogs of plant DNMT3s are also common, however based on our evolutionary analysis, these duplications did not evolve into conserved DNMT3 ortholog families across multiple species (Fig. 1a). Of note, PpDNMT3a and PpDNMT3b are not orthologs of mammalian DNMT3a and PpDNMT3b, respectively (Fig. 1a). Similarly, PpDRM1 and PpDRM2 are not orthologs of angiosperm DRM1 and DRM2, respectively (Fig. 1a). In summary, while DRMs are commonly considered as the plant homologs of euarotic DNMT3, here we show that DRMs are evolutionary distinct from DNMT3, and that true DNMT3 plant homologs exist throughout the plant kingdom, except in angiosperm.

**De novo methylation is dependent on PpDNMT3b and PpCMT.** Profiling genomic methylation in DNMT mutants refers mainly to DNA methylation maintenance activities. To evaluate the activity of P. patens DNMTs in de novo methylation, we profiled the methyleomes of P. patens DNMT deletion mutant plants, namely met, cmt, dnmt3a, dnmt3b, drm1, and drm2 single deletion mutants, as well as in drm1/drm2 (drm1drm2) and dnmt3a/dnmt3b (dnmt3ab) double deletion mutants (Supplementary Fig. 3). All single and double DRM and DNMT3 mutants were viable and developed similarly to wild type (WT) (Supplementary Fig. 4). Genomic methylation averages clearly showed that CG, CHG, and CHH sites were virtually eliminated and specifically disrupted in met, cmt, and dnmt3b mutants, respectively (Fig. 1b). More precisely, met mutant lost 93% of CG methylation, cmt mutant lost 97% of CHG methylation, and dnmtb mutant lost 95% of CHH methylation (Fig. 1b). These complete and specific hypomethylation in P. patens DNMT mutants led us to conclude that CG, CHG, and CHH contexts in P. patens are directly and primarily methylated by PpMET, PpCMT, and PpDNMT3b, respectively.
we found that RPS is regularly methylated in *P. patens rdr2* mutant plants. Altogether, these context-specific RPS methylation phenotypes in each of the mutants suggest that de novo methylation in *P. patens* can be mediated by *DNMT3b* at CG and CHH sites and by *CMT* at CHG sites without the involvement of DRMs or the canonical RdDM pathway. The reduction of CG methylation in RPS DNA in *met* T1 plants suggests that de novo CG methylation of RPS is also relied on PpMET. This assumption is in confirmation with published de novo methylation activity of mammalian DNMT1 and *Arabidopsis* MET1 either by themselves or in cooperation with de novo methylases. Alternatively, CG hypomethylation in *met* mutant could suggest that CG methylation in RPS is dependent on PpMET maintenance activity within just a few rounds of somatic cell generations.

**PpDNMT3b and PpCMT affect genomic CG methylation.** The near-complete elimination of CG methylation in the *met* genome (Fig. 1b) suggests that unlike animal DNMT3, PpDNMT3s do not have a role in maintaining genomic CG methylation. However, by focusing on transposable elements (TEs), we found a consistent decrease of 13% in CG methylation in both single *dnmt3b* and double *dnmt3ab* mutants (Fig. 2a), suggesting that DNMT3b is partially involved in maintaining the CG methylome. Further dissection of CG methylation based on their neighboring 5' nucleotides, i.e., NCG sites (N = any nucleotide), revealed that ACG sites are preferentially hypomethylated in *dnmt3b* and *dnmt3ab* (Fig. 2b). In association with the particular ACG hypomethylation in *dnmt3b* plants, we found that in *met* mutant ACG sites exhibit the highest residual CG methylation levels (Fig. 2c).

Among the four NCG sites, CCGs had the lowest CG-hypomethylated effect in *dnmt3b* mutant (Fig. 2b). CCG is one form of CHG for which we have previously showed that its methylation (mCCG) in the entire *Arabidopsis* genome and a couple of examined sequences in *P. patens*, is dependent on the methylation of the internal CG site (mCCG) maintained by MET1 genes. Here, we extended this observation to the entire *P. patens* genome by showing that CHG methylation, specifically at CCG sites, was diminished in the *met* mutant (Fig. 2e). This contributed to a 13% reduction in CHG methylation at TE sequences (Fig. 2d). Interestingly, we found that the reciprocal effect also exists, i.e., CmCG dependency on mCCG. Out of the four NmCG methylation contexts, CmCG was particularly reduced in the *cmt* mutant (Fig. 2b), while in *met* mutant CmCG residual level was second to ACG (Fig. 2c). Accordingly, along with their de novo methylation activities these results demonstrate the ability of PpCMT and PpDNMT3b in establishing CG methylation at genomic CCG and DCGs (D = A, G, or T) sites, respectively.

Non-CG methylation by mammalian DNMT3 is targeted preferentially to CW sites (W = A or T), such as CAC and CAG. Herein we found CHH methylation (mediated by PpDNMT3b) to be preferentially targeted to CWH sites (Supplementary Fig. 5), suggesting for functional conservation of CW methylation between mammalian and moss DNMT3s. However, the particular regulation of CHG methylation...
TEs in the corresponding to 1.5 times the IQR.

Box plot of NCG methylation difference in TEs between WT and indicated mutants (N = any nucleotide). 

Averaged genomic CHG methylation level in WT and DNMT mutants separated to CWG (i.e., CAG or CTG) and CCG sequences, which are mostly transcriptionally silenced and are controlled by PpDNMTs similarly to the way TE methylation is controlled solely by PpCMT.

PpDNMT3b mediates heterochromatic-mCHH and affected by PpCMT. DNA methylation in P. patens is specifically targeted to TEs (Supplementary Fig. 6) and segregated away from genes. Only about 0.5% of the methylated cytosines reside within genomic sequences, which are mostly transcriptionally silenced and are controlled by PpDNMTs similarly to the way TE methylation is regulated by PpDNMTs (Supplementary Fig. 6). In agreement, DNA methylation in P. patens is positively associated with heterochromatin (i.e., H3K9me2) and negatively associated with euchromatin (e.g., H3K4me3) marks (Fig. 3a). Further showed that similarly to Arabidopsis, long TEs in P. patens tend to be more heterochromatic, whereas short TEs are more euchromatic. Consistent with the relationship with heterochromatin, we found DNA methylation level to associate with TE size, i.e., to accumulate at relatively longer TEs (Fig. 3c). These correlations of DNA methylation with heterochromatin, together with the complete or near complete elimination of CG, CHG, and CHH methylation in met mutant. Overall, our results suggest that PpMET, PpCMT and PpDNMT3b function preferentially at heterochromatin, and particularly enriched for GC nucleotides and H3K9me2, and particularly increased within low GC and H3K9me2 TE regions (Fig. 3f, g). When focusing on short TEs (<500 bps), we found that hypermethylation and hypo-methylation in cmt background continued to associate with euchromatic and hetero-chromatic regions, respectively (Supplementary Fig. 7b), suggesting that the chromatin structure, rather than TE size, determines the CHH methylation effect in cmt mutant. 

PpDRMs target transcribed-euchromatic TEs. Neither single (drm1 or drm2) nor double (drm1drm2) mutant showed reduction of global genomic methylation (Fig. 1b and Supplementary Fig. 8a). Similar to drm mutants, no effect on methylation was recently reported for P. patens rdr2 mutant, which we validated here while substantially expanding our analysis to a larger genomic portion (80 vs. 20%; Supplementary Fig. 8a and Supplementary Table 2). These results, together with the complete CG, CHG, and CHH hypo-methylation in met mutant, imply a trivial methylation activity of DRMs and RDR2 in P. patens.

As opposed to a global methylation phenotype, we next checked for a localized methylation effect in drm mutants within statistically supported differentially methylated regions (DMRs)
separated into distinct chromatin configurations. While hypo-methylated DMRs were not significantly enriched over hyper-methylated DMRs in either drm or rdr2 mutant (Supplementary Fig. 8b), we found that CHH-DMRs of drm1drm2 double mutant were particularly hypo-methylated within genomic regions enriched for siRNA, low GC content, low histone H3 abundance, high H3K4me3, short TEs, and long-terminal-repeat (LTR) regions of retrotransposons (Fig. 4a – top panel). Single mutants drm1 and drm2 CHH-DMRs were mostly hyper-methylated and did not associate with any chromatin or DNA features (Fig. 4a lower panels), thus implying for an unrelated noise, which is a common feature of asymmetric methylation. Under this assumption, the particular hypomethylation effect in drm1drm2 (Fig. 4a top panel) suggests for some functional redundancy between DRM1 and DRM2. Intriguingly, we found CHH-DMRs of single and double drm mutants, as well as of rdr2 to be gradually hypo-methylated within a small number of windows (≤ 5610) of expressed TEs (Fig. 4a right panels and Supplementary Fig. 8c, d), which were also abundant in H3K4me3 and depleted of H3K9me2 (Supplementary Fig. 8e). Overall, these results associate PpDRMs methylation activity with RDR2 generated siRNA, as well as with actively-transcribed euchromatic TE sequences, both of which are signatures of RdDM activity in angiosperm.

The weak genomic methylation activity of DRMs in P. patens could be explained by the exceptionally high efficiency of PpCMT and PpDNMT3b. PpCMT targets CHG methylation as strongly as PpMET targets CG methylation (Figs. 2a, d and 4e), and PpDNMT3b targets CHH methylation with more than twice the level of CHH methylation in Arabidopsis (Fig. 4e)41. Consequently, together with their ability to de novo methylate DNA, it is possible that PpCMT and PpDNMT3b target and maintain non-CG methylation even within euchromatic regions that have a weak heterochromatic signal.

In support of a trivial role for RdDM in P. patens, we found siRNA in P. patens to overlap with only 5% of methylated TEs, in comparison to 65% in Arabidopsis (Fig. 4b). Moreover, similarly to Arabidopsis, we found siRNA in P. patens to be enriched

Fig. 3 PpCMT and PpDNMT3 methylate heterochromatin. a Pearson correlation coefficients between CG/CHG/CHH methylation, GC content, and indicated histone modifications of TEs in 50 bp windows. b Box plots showing GC content, H3K9me2, and H3K4me3 levels in 50 bp windows within five quantile TE sizes. c Box plots of average DNA methylation in 50 bp windows of WT protonema over five quantiles of TE sizes. d Box plots of percent-methylation-change between WT and indicated mutants 50 bp windows with a minimum 10% methylation in either of the samples, over TE size. e Patterns of TE CHH methylation in WT and indicated mutants as described in Fig. 2a. f Box plots showing the distribution of percent-methylation-change per 50 bp windows between WT and cmt mutant over H3K9me2, GC content, and TE size quantiles. g CHH methylation level (red WT, blue mutant), CHH methylation difference (cmt minus WT), H3K9me2, and gene/TE annotations of a representative region from Chromosome 1:459,000–702,000. Genes and TEs oriented 5′ to 3′ and 3′ to 5′ are shown above and below the line, respectively. Open black box marks a cmt hypo-methylated region enriched for H3K9me2.
within long-heterochromatic TEs (Fig. 4c). In Arabidopsis, RdDM functions mostly in euchromatic TEs, while heterochromatic siRNAs are hardly involved in maintaining DNA methylation\footnote{18,42,43}. If the same is true in P. patens, then the exceptionally low abundance of siRNA in euchromatic TEs (0.9\%) could further explain the minor role of PpDRMs in genomic methylation.

In addition to actively transcribed TEs, another source for euchromatic TEs could be those located in gene promoters\footnote{18,44}. Notably, we found that the frequency of TE integration within the first 200 bp upstream to transcription start site (TSS) of genes, was lower by up to 2.3 times in P. patens than in Arabidopsis (Fig. 4d). This result is counterintuitive, considering that the P. patens genome contains eight times more TEs than that of Arabidopsis, which are also spread more evenly along the chromosomes in comparison to the centric concentration of TEs in Arabidopsis\footnote{39}. Hence, the particular depletion of TEs in P. patens from gene promoters, which are known to be the main target of DRMs and RdDM in angiosperms\footnote{18,44–46}, could contribute for the weak genomic methylation effect of PpDRMs and RDR2 in P. patens.

**Discussion**

To date, functional analyses of plant DNMTs were focused primarily on Arabidopsis and a few additional angiosperms. P. patens is a basal land plant that diverged from angiosperms about 400 million years ago\footnote{47} and encodes homologs of all four plant DNMT protein families\footnote{14}, including DNMT3 which has been lost during angiosperms evolution. Thus, our comprehensive analysis of the entire PpDNMT proteins under de novo and homeostasis methylation conditions allowed us to reveal their function, as well as to infer on the evolutionary mechanisms of DNA methylation in plants (Fig. 5).
Mammalian DNMT3s function primarily as de novo methylases of CG sites and in some tissues also of CH sites\(^{20}\). We show here that PpDNMT3s are required for de novo methylation of CG and CHH sites (Fig. 1b). As PpDNMT3b is the first non-animal DNMT3 to be functionally characterized, our results imply that de novo methylation of CG and non-CG sites is an ancient feature of eukaryotic DNMT3 that predates the divergence of plant and animal DNMT3s. Additionally, our data demonstrate the ability of DNMT3 to specialized in their hosts, such as the preference of mammalian DNMT3 towards CG sites and that of moss DNMT3 towards CHH sites. Conservation and diversification between mammalian and moss DNMT3s would provide the basis for further structure-function interactions of eukaryotic DNMT3. The narrow overlap of siRNA with DNA methylation (Fig. 4b) and the trivial methylation effect in mammalian DNMT3 towards CG sites and in some tissues also of CH sites\(^{20}\). We demonstrate that de novo methylation of PpDNMT3 does not involve the RdDM pathway. In comparison, the association between PpDRMs and RdDM in Arabidopsis (angiosperms) and P. patens (basal/DNMT3-encoding plants), respectively. b Schematic illustration of the evolution of plant DNMTs and their function based on previous and our studies. Backbone of phylogenetic tree is inspired by https://langdalelab.com/
is involved in establishing CHH methylation that is subsequently maintained by PpDNMT3b. This hypothesis is supported by the ability of Arabidopsis CMTs to establish CHH methylation in vitro, and by the residual of CWA methylation in Ppdnmt3b mutants (Supplementary Fig. 5) that resembles the preference of some angiosperm CMTs toward such CHH subcontext.

DRMs most likely evolved from plant-DNMT3s (Fig. 1a). Additionally, thus far there is not a single plant species (including early-diverged ones) that encode a DRM as its only DNMT besides MET1. Therefore, our findings of de novo methylation by PpCMT and PpDNMT3b, suggest that de novo methylation in early diverged plants was dependent on DNMT3 and/or CMTs (Fig. 3b). Assuming that basal-DRMs inherited their ability for de novo methylation from their ancestral plant-DNMT3s, one could ask what was the selective force for the appearance of DRMs was, considering that plants already had the machineries for de novo methylation of CG and non-CG sites (Fig. 5). We propose that DRM’s ability to mediate methylation to a particular chromatin region played an important role in their evolution. CMT and DNMT3 evolved to target H3K9me2-heterochromatin, which consequently induced the emergence of DRMs to function in euchromatin, which is depleted of H3K9me2 and enriched for H3K4me3. H3K4 methylation was shown to inhibit mammalian DNMT3s, and PpDNMT3’s CHH methylation activity is negatively associated with H3K4 methylation (Fig. 3a–d). Accordingly, the association of PpDRMs CHH methylation at H3K4me3-euchromatric sites (Fig. 4a and Supplementary Fig. 8e), suggests that DRMs evolved to target a specific type of nucleosomal DNA (i.e., containing H3K4me) that cannot be targeted directly by DNMT3. H3K4me3-associated euchromatric TEIs include actively-transcribed TEs, as well as TEs that are located near genes. Such genetic elements are required to be silenced for the benefit of the host. For that reason, DRMs and the RdDM pathway probably persisted along plant evolution, even in species such as P. patens with highly efficient DNMTs that could mediate methylation even to sequences with low heterochromatic signal (Fig. 3b–d). In seed plants, de novo and maintenance of genomic methylation by RdDM is evidently enhanced, possibly in response to bursts of particular TE families that escaped silencing or integrated next to genes (Fig. 4d). The profound role of RdDM in angiosperms is associated with gaining additional genetic components that could have expanded or enhanced various steps in the pathway, such as siRNA biogenesis and methylation activity. For example, the appearance of SHH1 in angiosperms, which binds H3K9me2 and required for siRNA biogenesis, may have expanded the activity of RdDM in these species towards more heterochromatic elements. In addition to the developments in RdDM pathway, flowering plants also evolved a new CMT protein clade, that is CMT2, which targets heterochromatic CHH methylation via direct binding to H3K9me2. Consequently, the enhancement of de novo and genomic methylation activity by DRMs and heterochromatic CHH methylation by CMT2 could have compensated for the primary activities of DNMT3s, which allow their complete extinction in angiosperm (Fig. 5b).

Methods

Biological materials. All mutant plants were generated in the background of ‘Gransden 2004’ strain of P. patens \(^{[23,24]}\) and were propagated on BCD or BCDAT media\(^{[25]}\) at 25 °C under a 16 h light and 8 h dark regime. De novo methylation of the regions of the deleted genes were amplified with KOD hot start DNA polymerase (Novagen), cloned into the pTZ21 vector (Fermentas) and sequenced to validate their integrity. Next, the 5′ and 3′ fragments were subcloned into either the pMBL5 vector (GenBank: DQ228103.1) or the pMBLS Nos Hyg vector \(^{[26]}\).

Constructs were introduced into protoplasts via PEI-mediated transformation as described using 15 µg of plasmid restricted to linearize the construct 72 h after regeneration, transformants were selected on BCDAT medium containing 25 µg/ml hygromycin (Duchefa) or 25 µg/ml G418 (calbiochem). Resistant plants were further tested by tissue PCR to verify correct integration of the construct into the genome. PCR reactions mixed between the insertion and the sequence flanking the deleted fragment at both the 5′ and 3′ ends (primers listed in Supplementary Table 2). In addition, loss of the endogenous targeted loci was correlated with lack of amplification of the targeted sequence as compared to a positive control. ΔPdmd2 and ΔPdmdmt3a single deletion mutant protoplasts were used to generate ΔPdmd1ΔΔPdmd2 and ΔPdmdmt3aΔPdmdmt3b double deletion mutant lines, respectively, as described above.

Generation of RPS transgenic lines. The RPS transgene was introduced into the genome of WT and mutant plants via non-homologous recombination. To this end, a pMBL5 + Zeo vector was constructed by subcloning the Zeocin resistance cassette (Sh ble gene) from pRT101-Zeo \(^{[33]}\), replacing the G418 resistance cassette (nptII gene) of the pMBL5 vector (GenBank: DQ228103.1). The RPS fragment was subcloned from the p35 GUS/RPS vector \(^{[34]}\) into pMBL5 + Zeo vector. Both the RPS and Zeocin resistance sequences were ligated in the final pMBL5 + RPS construct to ensure integrity. Following transformation (as described above) and selection on BCDAT medium containing 50 µg/ml Zeocin (Invigro), resistant plants were tested to verify insertion of the construct into the genome by tissue PCR amplifying an internal transgene sequence spanning both the RPS sequence and the selection cassette (primers listed in Supplementary Table 3).

Validation of RPS absence in P. patens genome and sRNAome. The RPS sequence (GenBank: X92381.1) was used for homology search (blat) in the P. patens v3.0 genome. Additionally, it was used to search for corresponding small RNAs by NCBI SRA-Blast \(^{[60]}\) using small RNA-seq data of P. patens prototema (SRX247005-SRX247008 and SRX327325-SRX27330).

Published genomic data. Data for sRNA were derived from \(^{[35]}\), for mRNA from \(^{[36]}\), and for histone modifications from \(^{[37]}\).

Bisulfite sequencing of the RPS transgene. A fragment of RPS was PCR amplified from bisulfite treated genomic DNA, extracted from prototema tissue, using primers RPS-top-R-new and RPS-top-F (primers listed in Supplementary Table 3) and KAPA HiFi Uracil polymerase (kappa biosysstems), then cloned into pET11.2 (Thermo Fisher Scientific). The methylation status of RPS forward strand of individual clones was determined by Sanger sequencing.

Phylogenetic analysis. PpDNMT3b and PpDNMT3c protein sequences were used to search for homologs by blastp vs. NCBI Non-redundant protein database and by blast vs. the 1000 plants (1kp) transcriptome database. Alignment of selected DNMT3, DRM and DNMT1 MTD protein sequences was performed using MUSCLE v3.8.3. The motif order was rearranged in DRM sequences to match the linear organization of canonical DNMTs. Protein accessions are listed in Supplementary Table 2. MTDs of animal and plant DNMT1/MT1 homologs were added as outgroup. The phylogenetic tree was constructed by IQ-TREE v1.6.4 using default parameters and illustrated by FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

BS-seq library preparation. Around 0.5 µg of genomic DNA from prototema tissue was extracted, sheared (by sonication), end repaired (10 µl T4 DNA ligase buffer (NEB B0205S), 4 µl 10 mM dNTP mix, 1 µl T4 DNA polymerase (NEB M0203S), 1 µl Klenow DNA polymerase (NEB M0215S), 1 µl T4 PNK (NEB M0201S), 1 µl 10X PNK buffer (NEB), 10 µl 1X TadP buffer (NEB), 1 µl TadP and 1 µl Klenow exo minus (NEB M0212S), water to 50 µl), and ligation to methylated-adapter (25 µl quick ligation buffer (NEB), 1 µl 10 mM preannealed bs-seq-adapter (Supplementary Table 2), 1 µl quick ligation ligase (NEB M2002S), water to 50 µl). Adaptor-ligated libraries were subjected to two sequential treatments of bisulfite conversion using the EpI-Tect Bisulfite kit (Qiagen). Bisulfite-converted libraries were amplified by PCR (2.5 U of ExTaq DNA polymerase, Takara Bio), 5 µl of 10X PCR master mix (Promega), 5 µl of each primer, and 12 µl DEPC water. The PCR products were purified with the solid-phase reversible immobilization method using AMP-Pure beads (Beckman Coulter) and quantified with Bioanalyzer (Agilent).

Deep sequencing was performed on Illumina Hi-Seq 2000.

Published on December 16, 2019 by NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-09496-0 | www.nature.com/naturecommunications
BS-seq data analysis. BS-seq reads were processed by converting all the Cs in the ‘forward’ reads to Ts, and all the Cs in the ‘reverse’ reads to As. Converted reads were aligned to the converted reference scaffold using Bowtie168. Methylation level for individual cytosines along the chromosomes was calculated by counting the number Cs divided by the number of (C + T) single-c file on GSE118153. Genomic methylation averages (Fig. 1b, Fig. 2e, Supplementary Fig. 4, and Supplementary Table 3) were calculated by averaging methylation of the entire single-c file separated to distinct sequence contexts (e.g., CG, CHG, or CHH) and nuclear vs. organelle chromosomes. Single-c file were further used to plot methylation patterns in TE (Fig. 2a, Fig. 2d, and Fig. 3e), as well as fractional methylation within a 50 bp sliding window (‘50’ file on GSE118153) that were used in downstream analyses (e.g., Fig. 2b, c, Fig. 3a, c, d, and Fig. 4a). TE frequency meta-analysis. The abundance of TEs near TSSs of P. patens and A. thaliana genes was assessed using publicly available genes and TEs annotations and TE frequency meta-analysis were aligned to the converted reference scaffold using Bowtie168. Methylation level a Supplementary Information provided as a Source Data corresponding author on reasonable request. The source data underlying Fig. 1c is
Reporting Summary
Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The data supporting the findings of this study are available within the article and its Supplementary Information. BS-seq data has been deposited in the Gene Expression Omnibus (GEO) database under the accession number GSE118153. The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request. The source data underlying Fig. 1c is provided as a Source Data file. A reporting summary for this Article is available as a Supplementary Information file.

Received: 3 November 2018 Accepted: 13 March 2019 Published online: 08 April 2019

References
1. Goll, M. G. & Estor, T. H. Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. 74, 481–514 (2005).
2. Feng, S. et al. Conservation and divergence of methylation patterning in plants and animals. Proc. Natl Acad. Sci. USA 107, 8689–8694 (2010).
3. Niedererbuah, C. et al. Widespread natural variation of DNA methylation within angiosperms. Genome Biol. 17, 194 (2016).
4. Du, J., Johnson, L. M., Jacobsen, S. E. & Patel, D. J. DNA methylation pathways and their crosstalk with histone methylation. Nat. Rev. Mol. Cell Biol. 16, 519–532 (2015).
5. Zemach, A., McDannell, J. I., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328, 916–919 (2010).
6. Cedar, H. & Bergman, Y. Programming of DNA methylation patterns. Annu. Rev. Biochem. 81, 97–117 (2012).
7. Jurkowska, R. Z. & Jetlsch, A. Enzymology of mammalian DNA methyltransferases. Adv. Exp. Med. Biol. 945, 87–122 (2016).
8. Wendte, J. M. & Schmitz, R. J. Specifications of targeting heterochromatin modifications in plants. Mol. Plant 11, 381–387 (2018).
9. Sotelo-Silveira, M., Chávez Montes, R. A., Sotelo-Silveira, J. R., Marsch-Martínez, N. & de Fèter, S. Entering the next dimension: plant genomes in 3D. Trends Plant. Sci. 23, 598–612 (2018).
10. Song, X. & Cao, X. Context and complexity: analyzing methylation in trinucleotide sequences. Trends Plant. Sci. 22, 351–353 (2017).
11. Liu, C. et al. Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution. Genome Res. 26, 1057–1068 (2016).
12. Feng, W. & Michaels, S. D. Accessing the inaccessible: the organization, transcription, replication, and repair of heterochromatin in plants. Annu. Rev. Genet. 49, 439–459 (2015).
13. Springer, N. M., Lisch, D. & Li, Q. Creating order from chaos: epigenome dynamics in plants with complex genomes. Plant Cell 28, 314–325 (2016).
14. Malik, G., Dangwal, M., Kapoor, S. & Kapoor, M. Role of DNA methylation in growth and differentiation in Physcomitrella patens and characterization of PpHmt1, a cytosine DNA methyltransferase. Plant Cell 27, 4081–4094 (2012).
15. Bewick, A. J. et al. The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. Genome Biol. 18, 65 (2017).
16. Du, J. et al. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell 151, 167–180 (2012).
17. Stroud, H. et al. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat. Struct. Mol. Biol. 21, 64–72 (2013).
18. Zemach, A. et al. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell 153, 193–205 (2013).
19. Huff, I. T. & Zilberman, D. Translational-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. Cell 156, 1286–1297 (2014).
20. He, Y. & Ecker, J. R. Non-CG methylation in the human genome. Annu. Rev. Genom. Hum. Genet. 16, 55–77 (2015).
21. Cao, X. et al. Conserved plant genes with similarity to mammalian de novo DNA methyltransferases. Proc. Natl Acad. Sci. USA 97, 4979–4984 (2000).
22. Tanigawa, M., Hardcastle, T. J., Lewis, M. G. & Regulation of genome-wide DNA methylation by mobile small RNAs. New Phytol. 217, 540–546 (2018).
23. Matzke, M. A. & Mosher, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat. Rev. 15, 394–408 (2014).
24. Caider-Gid, D. & Slotkin, R. K. Non-canonical RNA-directed DNA methylation. Nat. Plants 2, 16163 (2016).
25. Underwood, C. J., Henderson, I. R. & Martienssen, R. A. Genetic and epigenetic variation of transposable elements in Arabidopsis. Curr. Opin. Plant. Biol. 36, 135–141 (2017).
26. Daccord, N. et al. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. Nat. Genet. 49, 1099–1106 (2017).
27. Schmid, M. W. et al. Extensive epigenetic reprogramming during the life cycle of Marchantia polymorpha. Genome Biol. 19, 9 (2018).
28. Richards, C. L. et al. Ecological plant epigenetics: evidence from model and non-model species, and the way forward. Ecol. Lett. 20, 1576–1590 (2017).
29. Anderson, S. N. et al. Subtle perturbations of the maize methylome reveal genes and transposons silenced by chromomethylase or RNA-directed DNA methylation pathways. G3 (Bethesda). 8, 1921–1932 (2018).
30. Noy-Malka, C. et al. A single CMT methyltransferase homolog is involved in CHG DNA methylation and development of Physcomitrella patens. Plant Mol. Biol. 84, 719–735 (2014).
31. Yaari, R. et al. DNA METHYLTRANSFERASE 1 is involved in mCG and mCCG DNA methylation reference scaffold and is essential for sporophyte development in Physcomitrella patens. Plant Mol. Biol. 88, 375–400 (2015).
32. Gentry, M. & Meyer, P. An 11bp region with stem formation potential is essential for de novo DNA methylation of the RPS element. PLoS ONE 8, e63652 (2013).
33. Muller, A., Marins, M., Kamisugi, Y. & Meyer, P. Analysis of hypermethylation in the RPS element suggests a function signal for short inverted repeats de novo methylation. Plant Mol. Biol. 48, 383–399 (2002).
49. Frost, J. M. et al. FACT complex is required for DNA demethylation at promoters of Arabidopsis thaliana. Proc. Natl Acad. Sci. USA 110, 2828–2836 (2013).
50. Zhang, X. et al. DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. Nat. Struct. Mol. Biol. 19, 870–875 (2012).
51. Stroud, H., Greenberg, M. & Feng, S. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell 152, 352–364 (2013).
52. Tan, F. et al. Analysis of chromatin regulators reveals specific features of rice DNA methylation pathways. Plant Physiol. (2016). https://doi.org/10.1104/pp.16.00393
53. Li, Q. et al. RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. Proc. Natl Acad. Sci. USA 112, 2869–2874 (2015).
54. Rensing, S. A. et al. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319, 64–69 (2008).
55. Bewick, A. J. et al. On the origin and evolutionary consequences of gene body DNA methylation. Proc. Natl Acad. Sci. USA 113, 9111–9116 (2016).
56. Frost, J. M. et al. FACT complex is required for DNA demethylation at heterochromatic regions during reproduction in Arabidopsis. Proc. Natl Acad. Sci. USA 112, E4720–E4729 (2015).
57. Gouli, Q. & Baulcombe, D. C. DNA methylation signatures of the plant chromomethyltransferases. PLoS Genet. 12, 1–17 (2016).
58. Ibarra, C. A. et al. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Sci. (80.-.) 337, 1360–1364 (2012).
59. Roudier, F. et al. Integrative epigenomic mapping delineates four main chromatin states in Arabidopsis. EMBO J. 30, 1928–1938 (2011).
60. Matzke, M. A., Kanno, T. & Matzke, A. J. M. RNA-directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants. Annu. Rev. Plant. Biol. 66, 243–267 (2015).
61. Ma, L. et al. Angiosperms are unique among land plant lineages in the occurrence of key genes in the RNA-directed DNA methylation (RdDM) pathway. Genome Biol. Evol. 7, 2648–2662 (2015).
62. Law, J. A., Vashishth, A. A., Wohlschlegel, J. A. & Jacobsen, S. E. SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. PLoS Genet. 7, e1002195 (2011).
63. Ashton, N. W. & Cove, D. J. The isolation and preliminary characterisation of auxotrophic and analogue resistant mutants of the moss, Physcomitrella patens. Mol. Gen. Genet. MGG 154, 87–95 (1977).
64. Nishiyama, T., Higashiyama, Y., Sakakibara, I., Kato, M. & Hasebe, M. Tagged mutagenesis and gene-trap in the moss, Physcomitrella patens by shuttle mutagenesis. DNA Res. 7, 9–17 (2000).
65. Frank, W., Decker, E. L. & Reski, R. Molecular tools to study Physcomitrella patens. Plant. Biol. 7, 220–227 (2005).
66. Zimmer, A. D. et al. Reannotated and extended community resources for the genome of the non-seed plant Physcomitrella patens provide insights into the evolution of plant gene structures and functions. BMC Genom. 14, 498 (2013).
67. Leinonen, R., Sugawara, H., Shumway, M. & Collaboration, I. N. S. D. The sequence read archive. Nucleic Acids Res. 39, D19–D21 (2011).
68. Johnson, M. et al. NCBI BLAST: a better web interface. Nucleic Acids Res. 36, W5–W9 (2008).
69. Wickett, N. J. et al. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl Acad. Sci. USA 111, E4859–E4868 (2014).
70. Matasci, N. et al. Data access for the 1,000 Plants (1KP) project. Gigascience 3, 17 (2014).
71. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797 (2004).
72. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522 (2018).
73. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274 (2015).
74. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jeromin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589 (2017).
75. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).