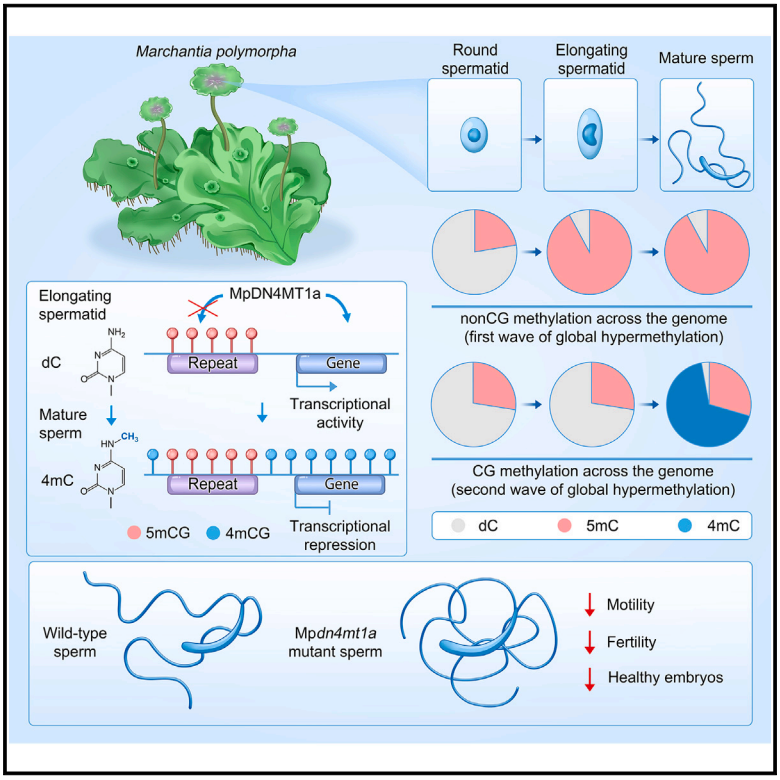
Cell

Article

Extensive N4 cytosine methylation is essential for *Marchantia* sperm function

Graphical abstract



Highlights

- Extensive 5-methylcytosine reprogramming occurs during *Marchantia* spermatogenesis
- N4-methylcytosine (4mC) is installed globally at CG sites in *Marchantia* sperm
- 4mC requires a methyltransferase specifically expressed during late spermiogenesis
- 4mC regulates sperm mRNA content, motility, and fertility

Authors

James Walker, Jingyi Zhang, Yalin Liu, ..., Liam Dolan, Keiji Nakajima, Xiaoqi Feng

Correspondence

Xiaoqi.Feng@ista.ac.at

In brief

Global N4 cytosine methylation in *Marchantia polymorpha* sperm regulates gene expression and promotes sperm fertility.

Walker et al., 2025, Cell *188*, 2890–2906
May 29, 2025
https://doi.org/10.1016/j.cell.2025.03.014

CellPress

CellPress
OPEN ACCESS

Cell

Article

Extensive N4 cytosine methylation is essential for *Marchantia* sperm function

James Walker,[1,7,8] Jingyi Zhang,[1,2,7] Yalin Liu,[1,3,7] Shujuan Xu,[4,7] Yiming Yu,[4] Martin Vickers,[1] Weizhi Ouyang,[4] Judit Tálas,[1] Liam Dolan,[5] Keiji Nakajima,[6] and Xiaoqi Feng[1,4,9,10,*]

[1]Department of Cell and Developmental Biology, John Innes Centre, Norwich NR4 7UH, UK
[2]Guangdong Basic Research Center of Excellence for Precise Breeding of Future Crops, College of Agriculture, South China Agricultural University, Guangzhou 510642, China
[3]Guangdong Provincial Key Laboratory for the Development Biology and Environmental Adaptation of Agricultural Organisms, College of Life Sciences, South China Agricultural University, Guangzhou 510642, China
[4]Institute of Science and Technology Austria, Klosterneuburg 3400, Austria
[5]Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, Vienna 1030, Austria
[6]Nara Institute of Science and Technology, Nara 630-0192, Japan
[7]These authors contributed equally
[8]Present address: The Salk Institute for Biological Studies, La Jolla, CA 92037, USA
[9]X (formerly Twitter): @XiaoqiFeng_m
[10]Lead contact
*Correspondence: Xiaoqi.Feng@ista.ac.at
https://doi.org/10.1016/j.cell.2025.03.014

SUMMARY

N4-methylcytosine (4mC) is an important DNA modification in prokaryotes, but its relevance and even its presence in eukaryotes have been mysterious. Here we show that spermatogenesis in the liverwort *Marchantia polymorpha* involves two waves of extensive DNA methylation reprogramming. First, 5-methylcytosine (5mC) expands from transposons to the entire genome. Notably, the second wave installs 4mC throughout genic regions, covering over 50% of CG sites in sperm. 4mC requires a methyltransferase (MpDN4MT1a) that is specifically expressed during late spermiogenesis. Deletion of Mp*DN4MT1a* alters the sperm transcriptome, causes sperm swimming and fertility defects, and impairs post-fertilization development. Our results reveal extensive 4mC in a eukaryote, identify a family of eukaryotic methyltransferases, and elucidate the biological functions of 4mC in reproductive development, thereby expanding the repertoire of functional eukaryotic DNA modifications.

INTRODUCTION

In prokaryotes, DNA is methylated in three forms: N6-methyladenosine (6mA), N4-methylcytosine (4mC), and 5-methylcytosine (5mC).[1,2] This methylation is catalyzed by methyltransferases that transfer a methyl group from S-adenosyl-methionine to the appropriate positions on target bases.[3] Through highly variable target recognition domains, the DNA methyltransferases (Dnmts) recognize specific sequence motifs, such as GATC and CCWGG, that are targeted by DNA adenine methyltransferase (Dam) and DNA cytosine methyltransferase (Dcm), respectively, in *Escherichia coli*.[4] This sequence specificity results in a relatively sparse amount of methylation across the genome (e.g., ~1.5%–2% of adenines and ~0.5%–0.8% of cytosines are methylated in *E. coli*).[4–6] Despite this scarcity, DNA methylation in prokaryotes has been found to play a range of key roles, including host-genome defense and regulation of transcription.[1,2]

In unicellular eukaryotes, 6mA has been reported with similarly low abundance (up to 0.4% of adenines in *Chlamydomonas reinhardtii* and 0.66% in *Tetrahymena thermophila*).[5] The existence of 6mA in multicellular eukaryotes is a subject of dispute, as the very low levels detected (~0.0001%–0.1%) have been ascribed to bacterial contamination.[5,7] The existence of 4mC in eukaryotic DNA is also disputed due to the amounts detected (typically <1% of cytosines) falling below technical false positive rates.[6,8,9] By contrast, 5mC is prevalent across eukaryotes, except that it has been independently lost in several lineages, including the model organisms *Drosophila melanogaster*, *Caenorhabditis elegans*, *Schizosaccharomyces pombe*, and *Saccharomyces cerevisiae*.[10] Eukaryotic 5mC is catalyzed by Dnmts that are highly conserved.[10–12] The genome distribution of 5mC can be global or mosaic. In vertebrates, 5mC occurs throughout the whole genome in the CG dinucleotide context, covering, for example, more than 80% of CGs in human tissues.[13] Such methylation is catalyzed *de novo* by Dnmt3 and maintained during DNA replication by Dnmt1.[14,15] By contrast, flowering plants have mosaic 5mC restricted to transposable elements (TEs) and some genes, occurring in all

sequence contexts (CG, CHG, and CHH; H = A, C, or T).[16,17] *De novo* methylation in all sequence contexts is catalyzed by the plant homolog of Dnmt3, DOMAINS REARRANGED METHYLTRANSFERASEs (DRMs), via the RNA-directed DNA methylation pathway.[18] CG and CHG/H methylation is maintained during cell divisions by a plant homolog of Dnmt1 (called METHYLTRANSFERASE 1 [MET1]) and two plant-specific CHROMOMETHYLASEs (CMT3/2), respectively.[19–21]

5mC plays numerous fundamental roles during animal and plant development.[22–24] 5mC exerts homeostatic functions in animal and plant genomes, with its disruption often causing developmental abnormalities and diseases, such as cancer.[13,16,25–27] During specific developmental processes or stress responses, 5mC undergoes local or global changes to promote transcriptional alterations.[13,16,24] One of the most drastic 5mC reprogramming events occurs during mammalian germline development, when primordial germ cells undergo genome-wide demethylation and remethylation.[28,29] Such reprogramming is essential for promoting totipotency and reproductive functions.[27,28] In flowering plants, functional 5mC reprogramming also occurs during spermatogenesis.[30] However, such reprogramming does not occur at a global scale but manifests in the hypermethylation of hundreds of specific loci. 5mC reprogramming during flowering plant spermatogenesis regulates gene expression, thereby promoting reproductive function.[31]

Marchantia polymorpha is an early-diverging land plant that shared a last common ancestor with flowering plants approximately 450 mya.[32] The vegetative thallus (a leaf-like tissue) displays similar 5mC patterns to flowering plants, with methylation in the CG, CHG, and CHH contexts associated with TEs[33,34] (16% of the genome is methylated; Figure 1A). However, unlike flowering plants, in which some genes are CG methylated, *Marchantia* vegetative tissues have no methylation in genes, similar to other early-diverging land plant species, *Physcomitrium patens* and *Selaginella moellendorffii*.[10,33,35] In contrast to the mosaic methylation in vegetative tissues, global cytosine methylation (covering 97% of the genome; Figure 1A) in all sequence contexts has been detected in *Marchantia* sperm.[34] The switch from mosaic methylation patterns in vegetative thallus to the global pattern in sperm suggests that genome-scale DNA methylation reprogramming occurs during male reproductive development. The high methylation level in the sperm was reported to be 5mC, which was postulated to be catalyzed by 5mC methyltransferases highly expressed in the antheridiophore (structure bearing the male organ of *Marchantia*).[34] However, cytosine methylation was detected by bisulfite sequencing (BS-seq),[34] which detects methylated cytosines due to their resistance to deamination by sodium bisulfite, whereas non-methylated cytosines are deaminated into uracils and read as thymines upon sequencing.[36] As 5mC is not the only methylated cytosine that is resistant to bisulfite conversion,[37] the nature of cytosine methylation and the mechanisms underlying this global methylation reprogramming remain unclear.

Here, we investigate DNA methylation dynamics and mechanisms during *Marchantia* spermatogenesis. We identify two temporal waves of extensive DNA methylation expansion from TEs to the non-repetitive parts of the genome. The first, occurring throughout spermiogenesis, installs 5mC in non-CG contexts across the genome and is dependent on canonical eukaryotic 5mC methyltransferases (MpDNMT3b and MpCMTa). The second wave, which installs extensive CG methylation at genic regions, occurs at the last stage of spermiogenesis. At this particular stage, we discover the specific expression of two putative 4mC methyltransferases (MpDN4MT1a and b). Using a combination of independent technologies, we confirm that this second wave of methylation is indeed 4mC and covers over 50% of genomic CGs. 4mC is dependent on the catalytic activity of MpDN4MT1a and is required for transcriptional suppression during the final stage of spermiogenesis, sperm motility and fertility, and post-fertilization development. In summary, our work establishes the presence of 4mC in a eukaryote and illustrates the importance of this abundant mark for eukaryotic chromatin regulation that is essential for development.

RESULTS

Two waves of extensive DNA methylation reprogramming during spermiogenesis

To decipher the mechanisms of cytosine methylation reprogramming in *Marchantia* sperm, we isolated mature sperm as well as individual antheridia (the male organs analogous to testes) during late spermatogenesis from *M. polymorpha* (subsp. ruderalis). Each antheridium was dissected in half, with one half precisely staged using microscopy based on reported developmental features, and the other half subjected to simultaneous RNA sequencing (RNA-seq) and BS-seq (Tables S1A and S1B). This method allowed us to categorize each antheridium into early (containing spermatocytes), middle (early spermiogenesis, containing mostly round spermatids), or late (late spermiogenesis, when DNA compaction and sperm tail formation occur, containing mostly elongating spermatids) stages (Figures 1B and S1A). Consequently, each antheridium provided paired information on its transcriptome and DNA methylome (Figure S1B). The accuracy of our developmental staging was further validated by examining the expression of genes with known transcription patterns[38] (Figure S1C).

BS-seq of developing antheridia halves and mature sperm showed that cytosine methylation at TEs increases considerably throughout spermatogenesis in all sequence contexts, from 38% average fractional methylation in early stage to 72% in mature sperm (27% in thallus; Figures 1C and S1D). This is most evident in the male sex chromosome (chromosome V) because of the high density of TEs on this chromosome[39] (Figure 1D). We also observed two temporal waves of methylation expansion into the non-repetitive parts of the genome. The first wave involves non-CG (CHG and CHH) methylation, which expands throughout spermiogenesis from covering 35% of the genome in round spermatids (middle stage) to 80% in elongating spermatids (late stage) and ultimately to 96% in mature sperm (Figures 1C–1E and S1D). The second wave manifests in the expansion of CG methylation primarily at the last stage of spermiogenesis (late stage onward; Figures 1C–1E and S1D). This event culminates in CG

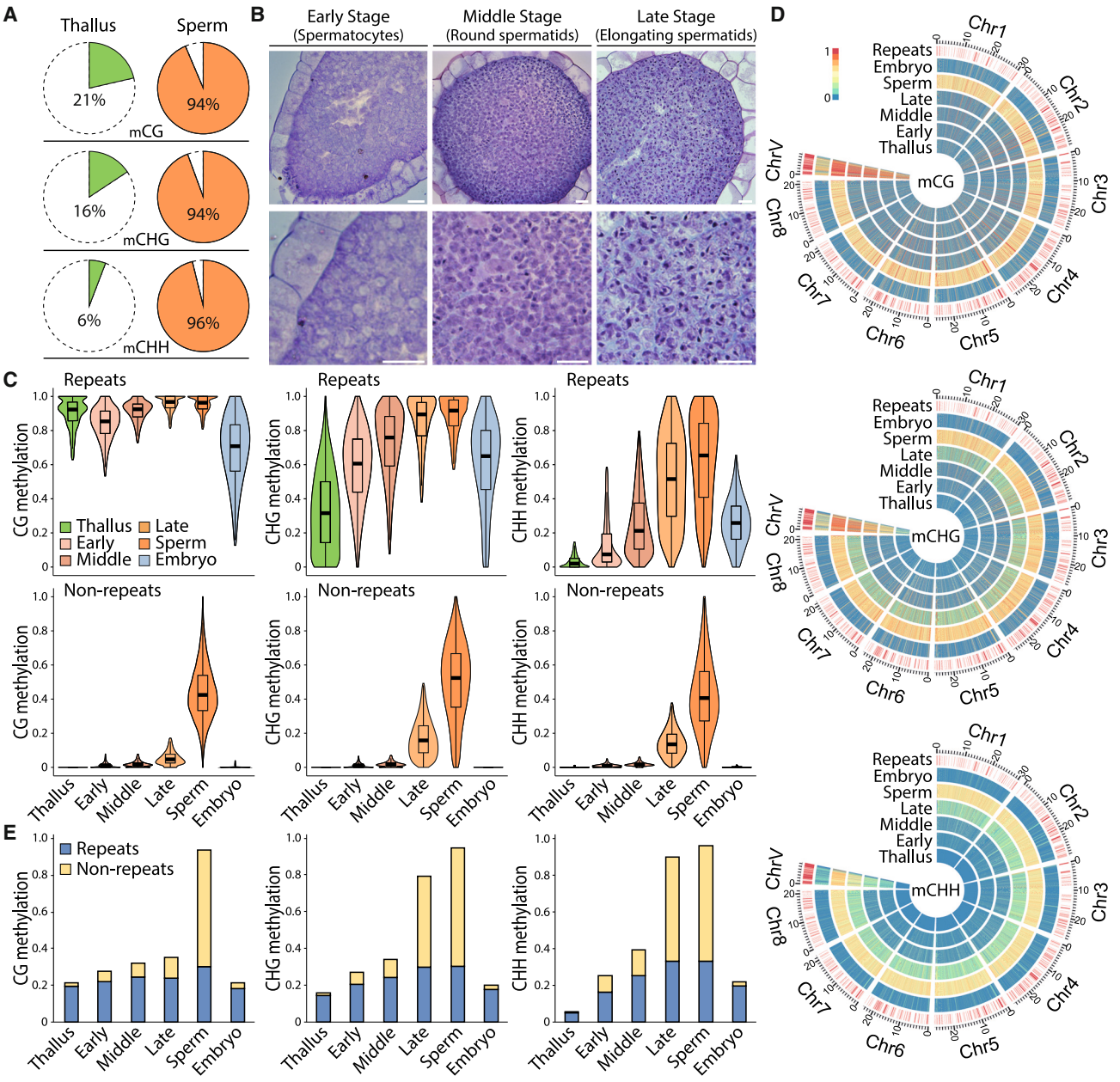


Figure 1. Two waves of DNA methylation reprogramming occur during *Marchantia* sperm development

(A) Pie charts illustrating the percentage of 100-bp windows across the *Marchantia* genome with >0.2 CG methylation, >0.1 CHG methylation, or >0.05 CHH methylation in thallus and sperm.

(B) Transverse sections of developing antheridia stained with toluidine blue. Scale bars, 20 μm.

(C) Violin plots showing methylation for 100-bp windows associated with repeats or non-repeats for thallus, antheridia (of early, middle, and late stages), sperm, and embryo.

(D) Heatmaps displaying methylation for 10-kb windows across the *Marchantia* Tak-1 chromosomes (Chr1–8, autosomes; ChrV, male sex chromosome) in the tissues shown in (C). Red bars indicate 10-kb windows covered by >95% TEs.

(E) Bar graphs depicting percentages of 100-bp genomic windows with evident methylation (as in A) in various tissues. Blue and yellow show windows associated with repeats and non-repeats, respectively. Early, middle, and late represent early, middle, and late stage, respectively (C–E).

See also Figure S1.

methylation occupying 94% of the genome in mature sperm, compared with 35% in late stage and 21% in thallus (Figure 1E). Methylation expansion at this scale has only been previously reported in the mammalian germ line, where DNA methylation is re-established genome-wide following global demethylation in primordial germ cells.[28,29]

In mammals, parental DNA methylation is globally erased and reprogrammed following fertilization.[27] The absence of global methylation in *Marchantia* thallus (Figures 1C–1E and S1D) and sporophytes (the diploid generation)[34] suggests a parallel loss of paternal methylation in *Marchantia* sperm after fertilization. To scrutinize this, we isolated *Marchantia* embryos and examined their methylation patterns post-fertilization. Methylation patterns in the embryo are similar to those of thallus, with an absence of both CG and non-CG methylation at genic regions (Figures 1C–1E and S1D). Additionally, the enhanced CG methylation observed at TEs in sperm is significantly reduced in the embryo, with methylation levels even lower than in the thallus (Figures 1C and S1D). These results indicate that sperm hypermethylation is fully reversed during embryo development.

Non-CG methylation reprogramming is induced by 5mC methyltransferases MpDNMT3b and MpCMTa

To understand the mechanism underlying non-CG methylation reprogramming, we examined the transcriptional dynamics of Dnmts during spermatogenesis. We found a substantial increase in the expression of Mp*DNMT3b* and Mp*CMTa* during spermiogenesis (Figure 2A; Table S2). Orthologs of these two genes are involved in CHH and CHG methylation of TEs in *P. patens*.[40,41] Furthermore, their temporal expression patterns correlate with the timing of TE methylation reinforcement and non-CG methylation expansion (Figures 1C–1E, 2A, and S1D), suggesting the likely involvement of Mp*DNMT3b* and Mp*CMTa* in these events. To mitigate the potential adverse effect of null mutations on development, as observed in the *cmt* mutants of *P. patens* that failed to produce antheridia, we generated knockdown mutants for Mp*DNMT3b* and Mp*CMTa* (Figure S1E). Both Mp*dnmt3b* and Mp*cmta* knockdown mutants produce antheridia and viable sperm. BS-seq of these mutant sperm shows a substantial reduction in CHG and CHH methylation throughout the genome compared with wild-type, whereas CG methylation remains largely unchanged (Figures 2B–2D and S1F). Therefore, MpDNMT3b and MpCMTa mediate non-CG methylation reprogramming in sperm.

Genes with homology to N4-cytosine methyltransferases are specifically expressed during late spermiogenesis

The timing of CG methylation expansion suggests that it is caused by enzyme(s) that are specifically expressed at the last stage of sperm maturation. Since none of the known *Marchantia* methyltransferases—including Mp*MET*—has such an expression pattern (Table S2), we examined all genes expressed specifically during late spermiogenesis. Among the most highly transcribed were two tandemly duplicated genes (Mp6g18330 and Mp6g18340; Figure 3A; Table S2) that cluster with prokaryotic N4 cytosine methyltransferases in a phylogenetic tree (Figure 3B) and contain TSPPY sequences characteristic of the N4 cytosine methyltransferase catalytic domain[42,43] (Figure S2). The arrangement of motifs in these proteins also reflects prokaryotic 4mC methyltransferases rather than 5mC methyltransferases[42,44] (Figure S2).

These genes are not prokaryotic contaminants, as the comparison of RNA-seq reads and genome annotations confirmed the existence of introns (Figure S3A), a key feature that distinguishes eukaryotic genes from prokaryotic genes. Furthermore, homologs of these genes are present in the genomes of all published liverwort species (Figure S3B). These include two other *M. polymorpha* subspecies, subsp. *polymorpha and* subsp. *montivagans*; two other *Marchantia* species, *M. paleacea* and *M. inflexa*; and in *Lunularia cruciata* (Figure S3B), which is a member of a distinct taxonomic order (Lunulariales) and last shared a common ancestor with the Marchantiales order approximately 228 mya.[45] No homologs of Mp6g18330 and Mp6g18340 were detected in other published plant genomes (259 in total),[46] suggesting these genes are specific to liverworts. Horizontal gene transfer from bacteria to liverworts is common,[32] and these two genes are likely to have originated through such an event. Another putative 4mC methyltransferase found in rotifers, N4CMT, is also thought to have arisen through horizontal gene transfer from bacteria.[9] Indeed, N4CMT and MpDN4MT1s cluster with phage and bacterial 4mC methyltransferases, respectively (Figure 3B). This, together with the lack of 4mC methyltransferases in other plant and animal lineages, indicates that these enzymes were acquired independently.

Transcripts of both genes are absent in the thallus, as well as published datasets from other tissues.[32] Their expression is only detected in late-stage antheridia, containing elongating spermatids, and in the mature sperm (Figure 3A; Table S2). Mp6g18330- and Mp6g18340-citrine fusion proteins expressed under native promoters confirm the specific expression in elongating spermatids (Figures S3C–S3J). This expression pattern suggests a potential contribution to the observed global CG methylation expansion during late spermiogenesis, as revealed by BS-seq. Given that both 4mC and 5mC resist deamination and can be read as methylated cytosines by BS-seq,[1,47] we further hypothesized that the global expansion of CG methylation occurs at the N4 position of cytosine, catalyzed by Mp6g18330 and Mp6g18340. Henceforth, these genes will be referred to as *M. polymorpha DNA N4 CYTOSINE METHYLTRANSFERASE1a* (Mp*DN4MT1a*) and Mp*DN4MT1b*.

Extensive 4mC is present in *Marchantia* sperm and requires MpDN4MT1a

To test our hypothesis that MpDN4MT1a and MpDN4MT1b catalyze 4mC in sperm, we performed a DNA dot blot immunoassay using an anti-4mC antibody. To address concerns regarding antibody specificity and bacterial DNA contamination,[5] we included unmethylated PCR product and *Marchantia* thallus DNA as negative controls, along with a PCR product synthesized with N4-methyl-dCTP as a positive control. While both negative controls yielded no signal, 4mC was detected in sperm DNA at a similar intensity to that in the positive control (Figure 3C).

To further validate the presence of 4mC using an independent, non-antibody-based method, we quantified modified nucleotides in thallus and mature sperm DNA using liquid chromatography-mass spectrometry (LC-MS was performed at two independent facilities to ensure repeatability; Figure 3D), a highly accurate technique for detecting and measuring DNA and RNA modifications.[48] Our LC-MS experiment accurately quantified

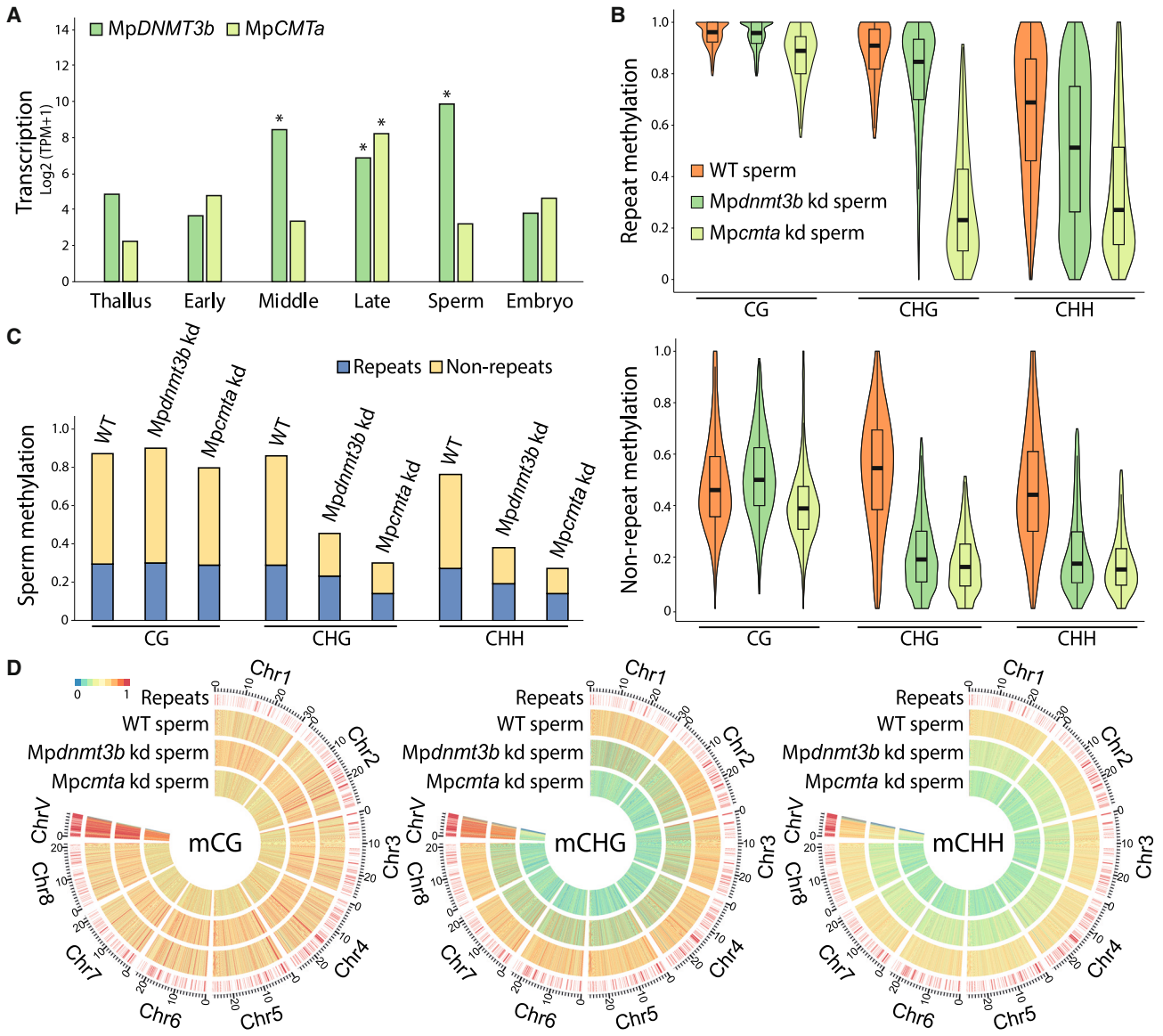


Figure 2. Non-CG methylation reprogramming in sperm requires Mp*DNMT3b* and Mp*CMTa*

(A) Transcript levels of Mp*DNMT3b* and Mp*CMTa* in thallus; early, middle, and late stage antheridia; mature sperm; and embryo. TPM, transcripts per million. *$p < 0.0346$, one-tailed *t* test demonstrating significantly higher expression compared with early stage.

(B) Violin plots showing CG, CHG, or CHH methylation at repeats or non-repeats, as in Figure 1C, in the sperm of the wild-type (WT), the Mp*dnmt3b* knockdown (KD) mutant, and the Mp*cmta* KD mutant.

(C) Bar graphs illustrating the percentage of 100-bp genomic windows with >0.3 methylation in CG, CHG, or CHH context in WT, Mp*dnmt3b* KD, and Mp*cmta* KD sperm. Blue and yellow show windows associated with repeats and non-repeats, respectively.

(D) Heatmaps as in Figure 1D showing methylation for WT, Mp*dnmt3b* KD, and Mp*cmta* KD sperm.

See also Figure S1.

the 4mC content (89%) in the N4-methyl-dCTP-synthesized PCR product (Figure 3E). Additionally, it detected no 4mC and 7.6% 5mC in thallus DNA, consistent with our BS-seq results (Figures 3D and 3E). Mature sperm exhibited much higher levels of 5mC (31.0%), reflecting hypermethylation observed by BS-seq (Figure 3E). However, this is substantially less than the 52.7% cytosine methylation measured by BS-seq, consistent with the presence of 4mC (Figure 3E). Indeed, LC-MS revealed that 4mC comprises a striking 14.8% of the cytosines in sperm (Figures 3D and 3E).

To test the link between MpDN4MT1s and 4mC, we generated two independent Mp*dn4mt1a* and Mp*dn4mt1b* double-knockout mutants (simplified as Mp*dn4mt1-1 and Mpdn4mt1-2*) via CRISPR-Cas9-mediated deletion (Figure S3K) and performed dot blot immunoassays and LC-MS on isolated sperm DNA. As expected, 4mC is undetectable in both mutants

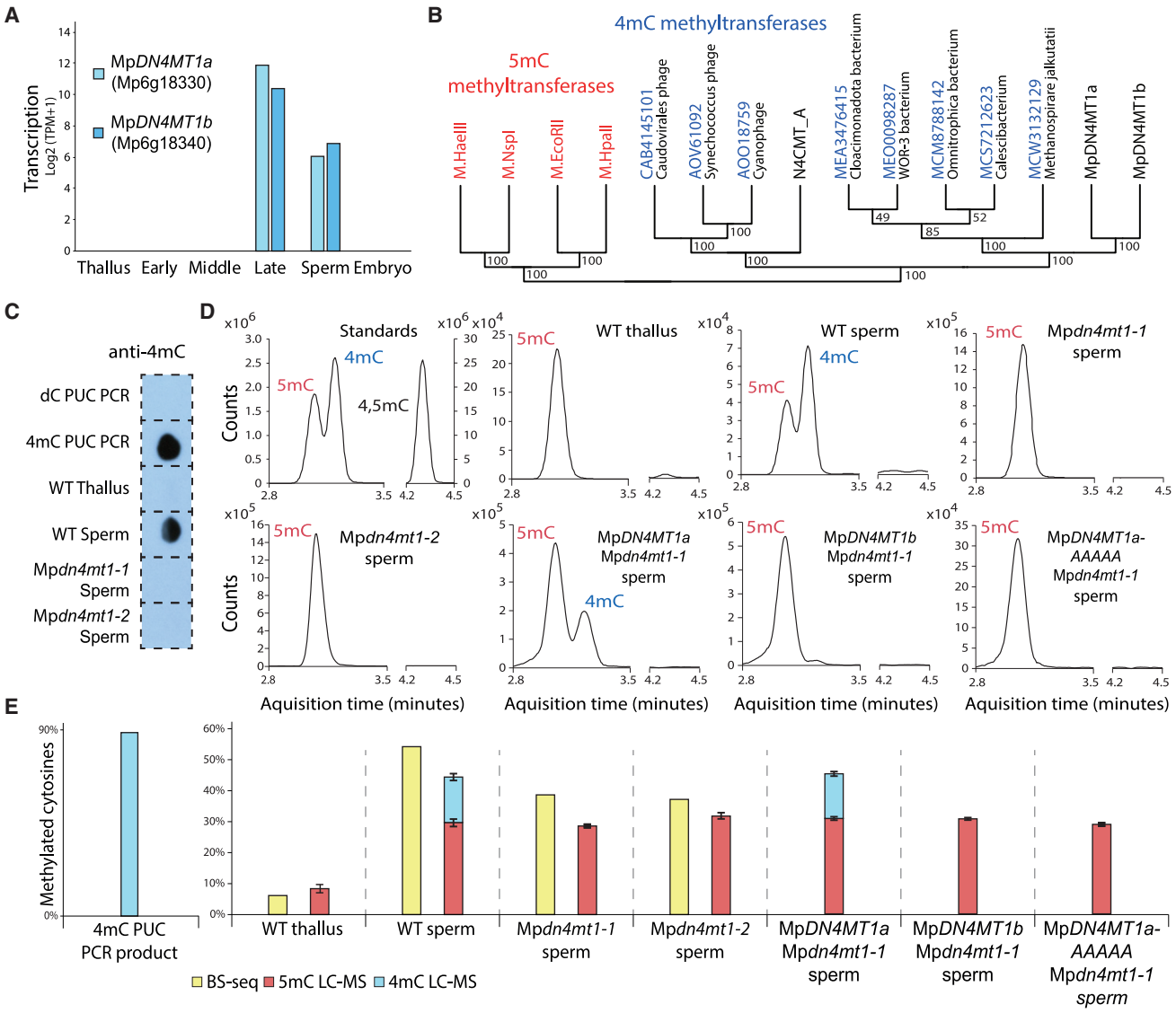


Figure 3. MpDN4MT1a is required for extensive 4mC in sperm

(A) Transcript levels of Mp*DN4MT1a* and Mp*DN4MT1b* in thallus; early, middle, and late stage antheridia; mature sperm; and embryo. TPM, transcripts per million.

(B) Cladogram of orthologous protein sequences representing prokaryotic 5mC (red) and 4mC (blue) methyltransferases with MpDN4MT1s and N4CMT_A in black. Bootstrap values are indicated.

(C) DNA dot blot immunoassay with anti-4mC antibody of dCTP PCR product, 4mdCTP PCR product, WT thallus, WT sperm, Mp*dn4mt1-1* sperm, and Mp*dn4mt1-2* sperm.

(D) LC-MS peaks of methylated deoxycytidine (5mC, 4mC, and 4,5mC) standards and DNA isolated from WT thallus, WT sperm, Mp*dn4mt1-1* mutant sperm, and Mp*dn4mt1-2* mutant sperm, as well as Mp*d4mt1-1* sperm with either WT Mp*DN4MT1a* (Mp*DN4MT1a* reintro), WT Mp*DN4MT1b* (Mp*DN4MT1b* reintro), or a catalytic mutant Mp*DN4MT1a* (Mp*DN4MT1a* AAAAA reintro) reintroduced with their native promoters.

(E) Percentages of methylated cytosines detected by BS-seq or LC-MS in the 4mdCTP PCR product, WT thallus DNA, and sperm DNA samples described in (D). LC-MS results were obtained from two independent facilities for WT thallus, WT sperm, Mp*dn4mt1-1* sperm, and Mp*dn4mt1-2*. Error bars are depicted for these and represent standard errors.

See also Figures S2 and S3.

(Figures 3C–3E). To confirm that the absence of 4mC is indeed caused by the deletion of MpDN4MT1s, we reintroduced each gene under its native promoter into Mp*dn4mt1-1* plants and examined sperm DNA methylation using LC-MS. The 4mC level in Mp*dn4mt1-1* mutant sperm is fully restored by Mp*DN4MT1a* expression but not by Mp*DN4MT1b*, showing that MpDN4MT1a (and not MpDN4MT1b) is required for 4mC methylation (Figures 3D and 3E). To further test whether MpDN4MT1a directly catalyzes 4mC in the sperm, we introduced a Mp*DN4MT1a* gene with its catalytic domain mutated (TSPPY to AAAAA) using its native promoter into Mp*dn4mt1-1* mutant plants. Unlike native Mp*DN4MT1a*,

mutated Mp*DN4MT1a* lacking a functional catalytic domain fails to complement 4mC in Mp*dn4mt1-1* mutant sperm (Figures 3D and 3E). These results demonstrate that MpDN4MT1a is a bona fide N4 cytosine methyltransferase that catalyzes extensive 4mC in *Marchantia* sperm.

4mC occurs in the CG context and is enriched for genic regions

BS-seq utilizes the resistance conveyed by the methyl group to bisulfite-mediated deamination of cytosine to uracil.[36] Unlike 5mC, which is almost completely resistant to bisulfite conversion, 4mC is only partially resistant, with roughly 43%–55% of 4mC remaining unmodified in a typical treatment.[47,49] To investigate the context and distribution of N4 cytosine methylation within the sperm genome, we performed BS-seq with one round of bisulfite treatment and compared it with data obtained from two treatments, as we expect 4mC to be progressively deaminated while 5mC remains unmodified. This comparison revealed a sensitivity of CG methylation to the second treatment in non-repetitive genomic regions (16% methylation loss, similar to the 14% methylation reduction of a 4mC spike-in control) but minimal sensitivity of CG methylation in repeats (1.8%) and similarly negligible sensitivity in non-CG contexts throughout the genome (1.7%; Figures S4A and S4B). Consistently, sequencing of Mp*dn4mt1-1* knockout sperm with one bisulfite treatment shows a near-complete loss of CG methylation at non-repetitive genic regions, whereas non-CG methylation is largely unaltered (Figures 4A–4C and S5A). The specific sensitivity of CG methylation in genic regions to bisulfite conversion and its dependence on Mp*DN4MT1* indicate that 4mC occurs in the CG context at genic regions in sperm.

To further validate our BS-seq result and characterize the distribution of 4mC in the sperm genome, we performed PacBio single-molecule, real-time sequencing (SMRT-seq; Table S3A), which has been shown to distinguish between 4mC and 5mC in prokaryotes.[50,51] Given the high (0.5%–7%) false discovery rates reported for 4mC detection by SMRT-seq in eukaryotic DNA, and the fact that the analysis tools were initially designed for use in prokaryotes, great caution is required to infer the presence of 4mC with this method.[6,48] Consequently, we inferred 4mC by comparing wild-type sperm with *Mpdn4mt1-1* mutant sperm, expecting that most false signals would be similarly produced in both samples. In the wild-type, 18% of cytosines were called as N4 methylated, in contrast to 4% in the Mp*dn4mt1-1* mutant (Table S3B). This is comparable to our LC-MS results and consistent with the dependence of 4mC on MpDN4MT1a (Figure 3E). SMRT-seq data also confirm that MpDN4MT1a preferentially targets CG dinucleotides, with 63% of total CG sites being called as 4mC methylated (80% outside of TEs), compared with 8% of CHG sites and 8% of CHH sites (Figures 4A–4C and S5A; Table S3B). This preference is lost in the Mp*dn4mt1-1* mutant, with 6% 4mC called at CG sites, 5% at CHG sites, and 4% at CHH sites (Figure S5A; Table S3B). The apparent false discovery rate of about 4% indicates that N4 methylation of non-CG cytosines is rare, whereas methylation of CG sites in genes and other non-repetitive sequences surpasses 75% (Table S3B).

To explore the underlying foundation of the ~4% false discovery rate, we performed SMRT-seq on whole genome amplified wild-type and Mp*dn4mt1-1* mutant sperm DNA. 0.4% CG, 0.7% CHG, and 1.1% CHH in wild-type, as well as 0.4% CG, 0.6% CHG, and 1.0% CHH in Mp*dn4mt1*, are subsequently called 4mC (Table S3C). False 4mC calls are associated with young LINE-1 transposable elements in human whole genome amplified samples and are predicted to be caused by DNA secondary structures that affect DNA polymerase kinetics in SMRT-seq.[52] Likewise, we detect an enrichment of 4mC calls in *Marchantia* repeats in the amplified libraries (4mC calls from 2.9% to 3.1% of cytosines in repeats compared with 0.7%–0.8% in non-repetitive regions; Table S3C), which explains the over-representation of 4mC calls for repetitive regions in *Marchantia* native libraries (Table S3B). We reasoned that the remaining false calls in the native libraries are likely because of the extensive 5mC in *Marchantia* sperm, as 5mC can affect DNA polymerase kinetics and cause false positive 4mC calling.[48,53] Accordingly, 12.9% of CGs in repeats are called 4mC in the Mp*dn4mt1-1* native library, compared with 4.4% in genic regions, which have substantially less 5mC (Figures 4A and 4C; Table S3B). Thus, a small amount of over-represented 4mC calls occur over TEs, while the majority of extensive 4mCGs called in genic regions in wild-type sperm are bona fide modifications, a conclusion that is consistent with BS-seq results (Figures 4A–4C).

To corroborate the presence of genic 4mCG found by BS-seq and SMRT-seq, we employed two additional methods of 4mC sequencing. First, 4mC-Tet-assisted bisulfite sequencing (4mC-TAB-seq) was performed, utilizing ten-eleven translocation methylcytosine dioxygenase (TET) enzymes to deaminate all cytosines except for 4mC (Table S1B). As 4mC is only partially resistant to bisulfite treatment,[47,49] 4mC-TAB-seq cannot determine absolute 4mC levels. Additionally, TETs only deaminate cytosines in the symmetrical CG context with high efficiency[54,55]; therefore, 4mC-TAB-seq can be utilized to distinguish between low and high 4mCG methylation. As a result, we find 4mCG detected by 4mC-TAB-seq is enriched for genic regions in wild-type sperm and correlates with 4mCG called by BS-seq and SMRT-seq, while 4mCG is abolished in Mp*dn4mt1-1* sperm (Figures S5A and S5B; Table S1B). Second, we performed Apolipoprotein B mRNA Editing Catalytic Polypeptide-like (APOBEC)-mediated deamination sequencing (4mC-AMD-seq), a bisulfite-independent method that deaminates unmethylated C and 5mC but not 4mC through the action of an enzyme so that only the latter is read as C[56] (Table S1B). Consistently, 4mC-AMD-seq of wild-type and Mp*dn4mt1-1* sperm demonstrates that 4mC is confined to the CG context at genic regions and is dependent on the presence of *MpDN4MT1* (Figures 4A–4C, S5A, and S5B). Our results obtained from LC-MS, BS-seq, 4mC-TAB-seq, and 4mC-AMD-seq unambiguously establish the presence of extensive 4mCG in genic regions of *Marchantia* sperm.

A notable feature of our SMRT-seq, 4mC-TAB-seq, and 4mC-AMD-seq data is the scarce (if any) N4 methylation of CG sites in TEs (Figures 4A–4C and S5A; Table S3B). TEs are extensively 5mCG methylated (Figure 4A), and our result could be explained by the failure of these methods to detect N4 methylation at cytosines methylated at both 4th and 5th positions. Alternatively, existing 5mC at CG sites could prevent N4 methylation by MpDN4MT1a during the last stage of sperm maturation. To distinguish these possibilities, we explored the existence of

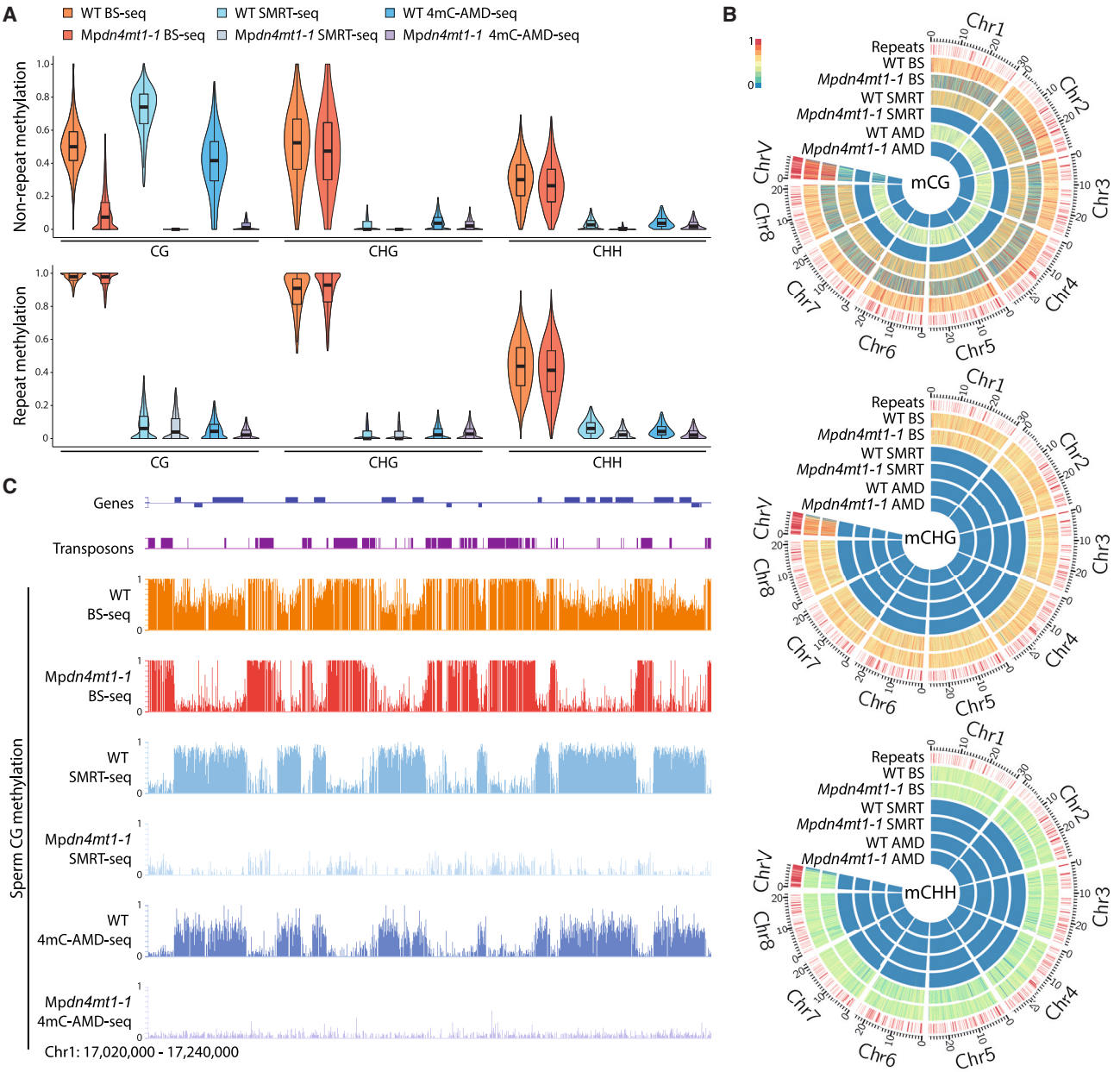


Figure 4. 4mC occurs across genic regions in the CG context

(A) Violin plots as in Figure 1C showing methylation of 100-bp genomic windows detected by BS-seq (one bisulfite treatment) in WT and Mp*dn4mt1-1* sperm, and 4mC detected by SMRT-seq and 4mC-AMD-seq in WT sperm and Mp*dn4mt1-1* sperm, associated with either repeats or non-repeats.

(B) Heatmaps as in Figure 1D, depicting methylation in WT and Mp*dn4mt1-1* sperm detected by BS-seq (one bisulfite treatment), SMRT-seq, and 4mC-AMD-seq.

(C) Snapshots of CG methylation detected by BS-seq (one bisulfite treatment) of WT and Mp*dn4mt1-1* sperm, as well as SMRT-seq and 4mC-AMD-seq, over genes and transposons.

See also Figures S4 and S5.

cytosines that are both C5 and N4 methylated (4,5mC) in *Marchantia* sperm by LC-MS, utilizing a synthesized 4,5mC standard. This experiment failed to detect any 4,5mC (Figure 3D), indicating that MpDN4MT1a does not methylate 5mC to form 4,5mC. In agreement with this, 5mCG was successfully predicted specifically in TEs in both wild-type and Mp*dn4mt1-1* sperm SMRT-seq data (Figures S5C and S5D). Thus, genic regions in *Marchantia* sperm are distinguished by N4 methylation at CG sites, whereas TEs are marked by C5 CG methylation. During spermiogenesis, 5mCG methylation rises to a saturated level of methylation (median at 0.97) in elongating spermatids, before MpDN4MT1a catalyzes 4mC (Figure 1C). Therefore, our data suggest a straightforward model for MpDN4MT1a recruitment genome-wide, where the saturated 5mCG at TEs occludes

4mC, while the remaining unmethylated CGs, at non-repetitive regions, become marked with 4mC.

4mC facilitates transcriptional shutdown in the sperm

The global nature of 4mC in *Marchantia* sperm suggests its involvement in genome-wide functions. As 4mC is deposited at the final stage of spermiogenesis, when global compaction of chromatin occurs in the sperm, we hypothesized that 4mC may play a role in facilitating this process. To test this hypothesis, we examined chromatin accessibility across the wild-type and Mp*dn4mt1-1* mutant sperm genomes using assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq; Table S1A). We observed a substantial increase in chromatin accessibility in Mp*dn4mt1-1* mutant sperm compared with wild-type sperm across all genic regions, where 4mC is prevalent (Figures 5A, 5B, and S6A). By contrast, the chromatin accessibility at transposons, where there is no 4mC, is unaffected by the loss of Mp*DN4MT1* (Figure 5A). These data indicate that 4mC suppresses chromatin accessibility in the sperm.

As chromatin accessibility is closely linked with transcription,[57,58] we next explored the effect of 4mC on transcription by performing RNA-seq on Mp*dn4mt1-1* mutant and wild-type sperm (Table S1A). *Marchantia* sperm has a protamine-based chromatin and is known to be transcriptionally inactive.[59,60] Consistently, a previous study reported an undetectable amount of RNA in *Marchantia* sperm.[34] However, since it is common for sperm to contain a low level of leftover mRNA,[61] we adapted a single-cell RNA-seq protocol to sequence mRNA extracted from the Mp*dn4mt1-1* mutant and wild-type sperm. Consistent with the genome-wide alteration of chromatin accessibility, a comparison of the mutant and wild-type RNA-seq data revealed global alterations in the sperm transcriptome (Figures 5C, S6B, and S6C; Table S4A).

In wild-type sperm, we found that chromatin accessibility over gene bodies is poorly correlated with gene transcription (Pearson's $R^2 = 0.24$; Figures 5D and 5E). As increased chromatin accessibility over gene bodies is a signature of active transcription,[62,63] this result aligns with the understanding that within mature protamine-packaged sperm, the majority of transcripts are not transcribed *de novo* but instead are residual from earlier spermiogenesis.[61,64,65] However, in Mp*dn4mt1-1* mutant sperm, chromatin accessibility over gene bodies and gene transcription exhibit a strong positive correlation (Pearson's $R^2 = 0.74$; Figures 5D and 5E), indicating active transcription in the absence of 4mC. Therefore, our data suggest that 4mC facilitates the transcriptional shutdown that occurs during the last stage of sperm maturation. Consistently, analysis of the transcripts that are differentially abundant between Mp*dn4mt1-1* mutant and wild-type sperm shows that their levels are more similar between Mp*dn4mt1-1* mutant sperm and elongating spermatids (the stage before 4mC deposition) than between wild-type sperm and elongating spermatids (Figures S6D and S6E; Table S4A), indicating that 4mC mediates the final transition to a mature sperm transcriptome.

To investigate how 4mC facilitates transcription shutdown in mature sperm, we closely examined DNA methylation at genic regions. Wild-type CG methylation (4mC) is indiscriminately deposited across genes regardless of their expression in the sperm (Figures 5B and 5F). By contrast, CHG and CHH methylation (5mC) is excluded from genes with higher transcript levels in Mp*dn4mt1-1* mutant sperm, particularly over transcription start sites (TSSs; Figures 5B and 5F), resembling 5mC absence from active mammalian promoters.[13] Therefore, 4mC is the predominant form of methylation over the TSS of these genes. As chromatin accessibility at TSS is critical for gene transcription,[57,58] this suggests that 4mC acts as a main repressive mark at genes that do not gain substantial 5mC during sperm maturation. In support of this hypothesis, we found that genes whose TSS has low 5mC gain more accessibility in Mp*dn4mt1-1* mutant sperm, compared with genes that have sufficient TSS 5mC (Figure 5G). Moreover, genes with low TSS 5mC are significantly over-represented in genes that show higher transcript abundance in Mp*dn4mt1-1* mutant compared with wild-type sperm, in comparison to those that show lower abundance ($p < 4.80e{-}33$; Fisher's exact test) or remain unchanged ($p < 4.09e{-}91$; Fisher's exact test). These data show that 4mC is especially important for the suppression of genes that gain less 5mC during earlier stages of spermiogenesis.

4mC is required for sperm function and post-fertilization development

The effect of 4mC on sperm chromatin accessibility and transcription suggests that it may be important for sperm function. We performed a Gene Ontology analysis on the genes significantly under-represented in the Mp*dn4mt1-1* mutant sperm compared with wild-type sperm. This analysis revealed the highest enrichment for genes relating to "cilium- or flagella-dependent cell motility" (Table S4B), including *DYNEIN LIGHT CHAIN 7* (Mp*LC7*),[38,66] *CENTRIN1 and 3* (Mp*CEN1/3*),[67] and the centriolar components *BASAL BODY PROTEIN 10* (Mp*BLD10*)[68] and *SPINDLE ASSEMBLY ABNORMAL PROTEIN 6* (Mp*SAS6*),[69] which are known to be involved in flagella formation and function (Figure 5C). The mutation of *LC7* was reported to cause a slow and jerky swimming pattern in *Chlamydomonas*,[66] while Mp*bld10* mutants in *Marchantia* produce largely immotile sperm.[68]

In contrast to transcripts that are upregulated and exhibit increased chromatin accessibility in Mp*dn4mt1-1* mutant sperm, the chromatin accessibility of these under-represented genes remains low (Figures 5B and S6F). These genes also show significantly higher 5mC levels at their TSS (Figures 5B and S6F). This suggests that these cell-motility-related genes are likely transcribed before sperm maturation, with their mRNA persisting into the mature sperm. Indeed, in elongating spermatids, these genes are transcribed and exhibit low 5mC at the TSS (Figure 5B; Table S4A). As our data suggest 4mC facilitates global transcriptional shutdown in the sperm, the under-representation of these gene transcripts in Mp*dn4mt1-1* mutant sperm likely results from dilution caused by large-scale transcriptional activation. It is plausible that the dilution of these transcripts leads to reduced functional proteins due to competition for translational machinery. Consistent with this idea, the inhibition of CEN1 translation is known to disrupt flagellum formation during *Marsilea* spermiogenesis.[67]

Consistent with the reported functions of the genes under-represented in Mp*dn4mt1-1* mutant sperm, we found that Mp*dn4mt1-1* (and Mp*dn4mt1-2*) mutant sperm exhibit significantly less directional and slower swimming compared with

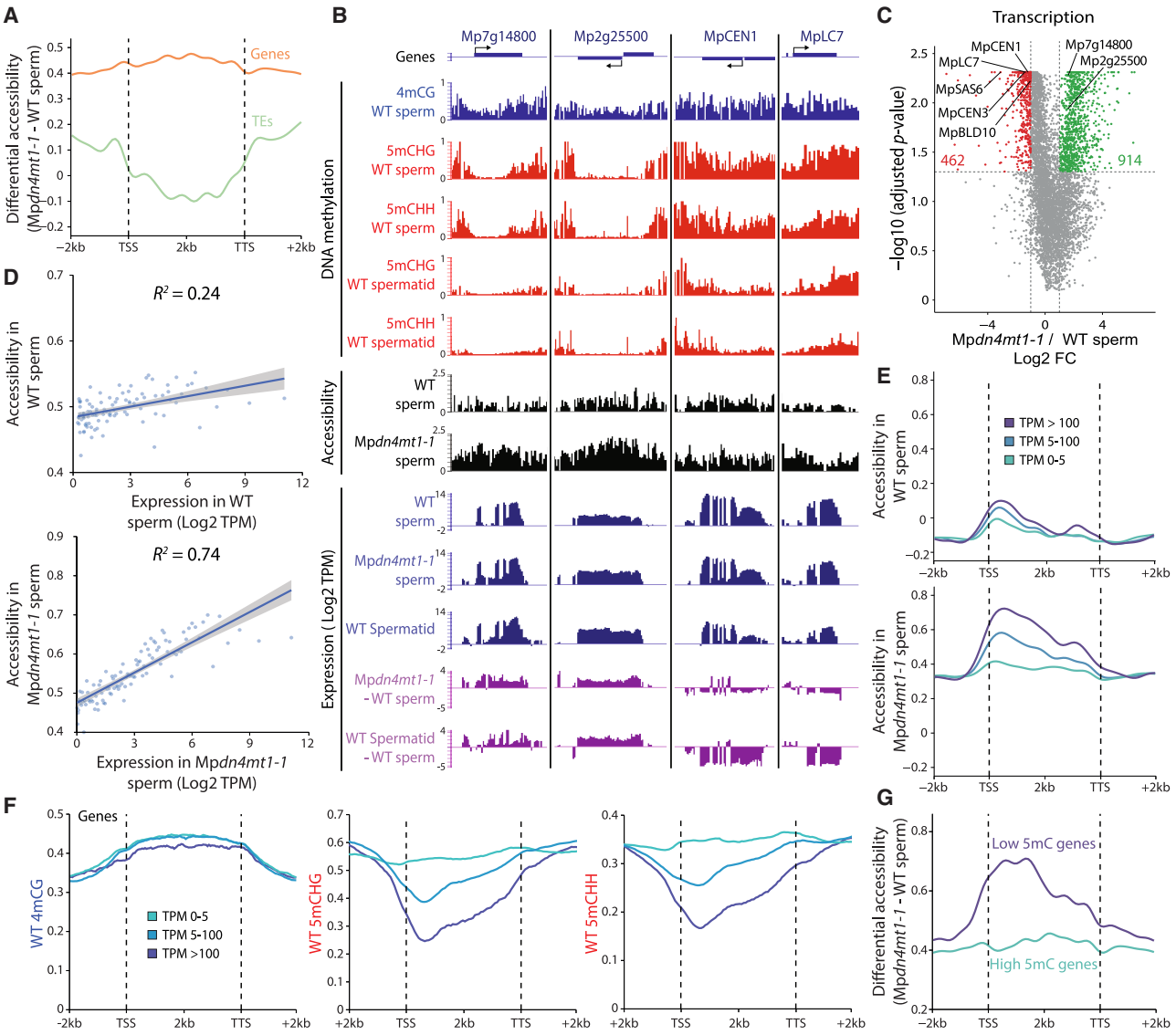


Figure 5. 4mC suppresses chromatin accessibility and transcription in sperm

(A) Profiles of the differential accessibility (measured by ATAC-seq) between Mp*dn4mt1-1* mutant sperm and WT sperm for genes and transposable elements (TEs). Genes and TEs were aligned at the 5′ and 3′ ends, and average differential accessibility levels for each 50-bp interval are plotted. Similar plots are used in (E)–(G).

(B) Snapshots of methylation (4mC and 5mC measured by 4mC-AMD-seq and BS-seq, respectively), WT and Mp*dn4mt1-1* sperm accessibility (measured by ATAC-seq), and expression patterns of WT sperm, Mp*dn4mt1-1* sperm, elongating WT spermatid, Mp*dn4mt1-1* sperm minus WT sperm, and elongating WT spermatid minus WT sperm (log_2 TPM). Arrows indicate the direction of transcription.

(C) Volcano plot illustrating transcripts that are over-represented (green) or under-represented (red) in Mp*dn4mt1-1* sperm compared with WT sperm. Vertical dashed lines mark 2-fold changes (FCs), and the horizontal dashed line shows $p = 0.05$.

(D) The relationship between gene expression (log_2 TPM) and chromatin accessibility (over 2 kb downstream of gene TSS). Each point represents one of 100 percentiles of gene expression (genes with TPM = 0 are excluded). The shaded area representing the 95% confidence interval. R^2 represents the proportion of the total variation in the dependent variable that is explained by the independent variable.

(E) Chromatin accessibility profiles for WT and Mp*dn4mt1-1* sperm over genes with 0–5, 5–100, or >100 TPM in Mp*dn4mt1-1* sperm.

(F) 4mCG (WT sperm 4mC-AMD-seq) and 5mCHG/5mCHH (WT sperm BS-seq) over genes with 0–5, 5–100, or >100 TPM in Mp*dn4mt1-1* sperm.

(G) Differential chromatin accessibility between Mp*dn4mt1-1* and WT sperm over genes with low or high non-CG 5mC methylation. CHG and CHH methylation lower than 0.3 and 0.15, respectively, over 2 kb downstream of gene TSS, were regarded as low non-CG methylation, and the rest of the genes were regarded as having high non-CG methylation.

See also Figure S6.

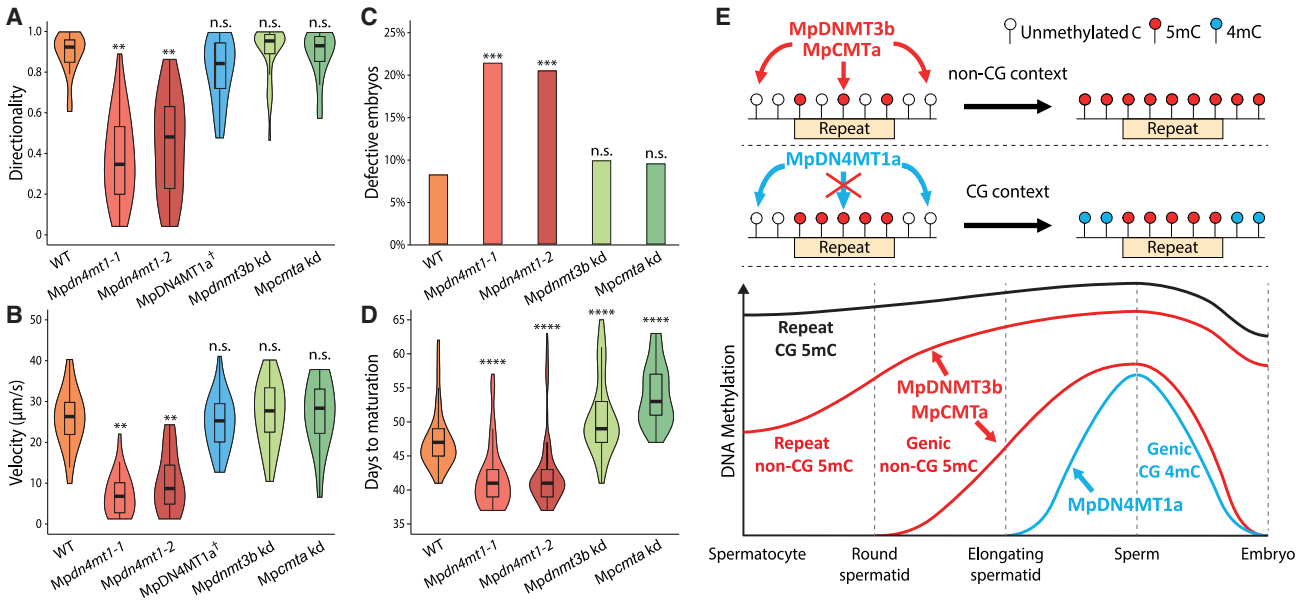


Figure 6. 4mC is important for sperm motility and post-fertilization development

(A and B) Directionality and velocity of motile sperm. $n = 30$ (WT), 28 (Mp*dn4mt1-1*), 31 (Mp*dn4mt1-2*), 31 (Mp*DN4MT1a* reintroduced into Mp*dn4mt1-1*; denoted by a dagger symbol), 31 (Mp*dnmt3b* KD), and 33 (Mp*cmta* KD). p values (Kolmogorov-Smirnov test) between indicated genotypes and WT are shown: $^{**}p < 3.8e{-}11$, n.s., not significant.

(C) Percentage of defective embryos observed after fertilization with WT ($n = 416$), Mp*dn4mt1-1* ($n = 210$), Mp*dn4mt1-2* ($n = 355$), Mp*dnmt3b* KD ($n = 193$), and Mp*cmta* KD ($n = 390$) sperm. p values (Fisher's exact test) between indicated genotypes and WT are displayed: $^{***}p < 0.0001$, n.s. not significant.

(D) Days to maturation (the release of spores) for viable embryos fertilized as in (C). $^{****}p < 1.2e{-}15$, Kolmogorov-Smirnov test. n.s., not significant.

(E) Temporal patterns and mechanisms of methylation reprogramming during *Marchantia* sperm development. First, 5mC mediated by MpDNMT3b and MpCMTa in non-CG contexts is reinforced over repeats and expands into genic regions during spermatogenesis. Subsequently, during the final stage of spermiogenesis, MpDN4MT1a establishes 4mC in the CG context across the genome, except in TEs covered by 5mC.

See also Figure S6.

wild-type sperm (Figures 6A, 6B, and S6G; Data S1A–S1E). Complementation of the Mp*dn4mt1-1* mutant with wild-type MpDN4MT1a successfully rescued the swimming phenotypes (Figures 6A and 6B; Data S1F). Notably, the effect on swimming behavior is specific to 4mC, as Mp*dnmt3b* and Mp*cmta* knockdown mutant sperm cells swim with comparable velocity and directionality to wild-type (Figures 6A and 6B; Data S1G and S1H). As fertilization in *Marchantia* is achieved by active swimming of sperm from male to female plants,[70] this motility defect would cause wild-type sperm to outcompete 4mC mutant sperm in natural environments.

We examined the impact of 4mC on sperm fertility by applying isolated wild-type and Mp*dn4mt1-1* mutant sperm cells (mixed together with equal numbers) directly to archegoniophores (female-organ-bearing structures) of female wild-type plants. Genotyping of resulting embryos showed a significant 91% reduction of fertility in Mp*dn4mt1-1* mutant sperm compared with wild type (among 809 embryos, 68 and 741, respectively, are of Mp*dn4mt1-1/+* heterozygous and wild-type genotypes; $p < 1.0e{-}5$, Fisher's exact test). Among the resulting embryos, a significantly higher proportion (20.5%–21.4%) of those produced from Mp*dn4mt1-1* (and similarly Mp*dn4mt1-2*) mutant sperm show defects during later embryo development than those produced from wild-type sperm ($p < 0.0001$; Fisher's exact test; Figures 6C and S6H). Among the ~80% Mp*dn4mt1-1/+* heterozygous embryos that can complete embryo development, maturation occurred about 6 days faster than wild-type embryos (Figure 6D). To test whether the requirement for post-fertilization development is specific to 4mC, we examined embryos fertilized with Mp*dnmt3b* or Mp*cmta* knockdown sperm. In contrast to the effect of 4mC, reduced 5mC in the sperm did not increase defective embryos (Figure 6C), although the resulting embryos took longer to mature (Figure 6D). Together, our results demonstrate that 4mC has important roles in sperm function, fertility, and post-fertilization development that are distinct from those of 5mC.

DISCUSSION

4mC has been extensively characterized in prokaryotes. However, whether 4mC has any function—or even exists—in eukaryotes is unclear. This ambiguity is attributed to the detected amounts—typically <1% of cytosines—falling below technical false positive rates (frequently associated with antibody-based and long-read sequencing methods), lack of evidence from multiple independent detection methods,[8,9] and complication of bacterial contamination.[6] These issues are also implicated in the detection of 6mA in multicellular eukaryotes.[5,7,48] In this study, using multiple independent methods, we demonstrate that late spermiogenesis in *Marchantia* involves extensive N4

cytosine methylation. This methylation is catalyzed by a methyltransferase, MpDN4MT1a, which is specifically expressed in the sperm (Figure 6E). The requirement of MpDN4MT1a for normal sperm mRNA abundance, swimming behavior, and embryo development demonstrates the biological importance of 4mC. Overall, our results unambiguously establish 4mC as a functional DNA modification in eukaryotes.

The closest homologs of MpDN4MT1a are bacterial N4 cytosine methyltransferases (Figure S2), and no homologs are present in available genome sequences of plant species outside liverworts. This, given the known cases of horizontal gene transfer between liverworts and bacteria,[32] suggests that MpDN4MT1a arose via such an event. Correspondingly, analyses of gene transfers in land plants have identified MpDN4MT1a.[71] However, MpDN4MT1a has clearly been adapted for eukaryotic genomic functions. In prokaryotes, N4 cytosine methyltransferases typically target a sequence motif as an essential part of restriction-modification systems that distinguish self from foreign DNA.[2] However, MpDN4MT1a targets CG dinucleotides. The high frequency of CG dinucleotides allows extensive 4mC methylation in the genome, encompassing 15% of total cytosines. This level of 4mC methylation far exceeds that observed in prokaryotes (typically <1%) or 6mA in prokaryotes or unicellular eukaryotes (typically <2%).

We found that 5mC reprogramming precedes MpDN4MT1a expression and 4mC reprogramming (Figure 6E), and there is no 4,5mC in the sperm genome (Figure 3D). These indicate that MpDN4MT1a cannot use 5mC as a substrate and suggest one reason why the CG dinucleotide specificity of this enzyme might be advantageous. In *Marchantia* somatic tissues, transposons and other repetitive elements are methylated in all sequence contexts; however, CG methylation is far more saturated than non-CG methylation (Figure 1C). Thus, restricting methylation to CG sites allows MpDN4MT1a to methylate exclusively non-repetitive regions. This way, through the two sequential waves of DNA methylation reprogramming, the *Marchantia* sperm genome is epigenetically distinguished at repetitive and genic regions, marked by 5mCG and 4mCG, respectively (Figure 6E).

Extensive 4mC over genic regions in the sperm (Figures 4A and 4C) may offer a mechanism for paternal genome recognition during post-fertilization development. Transcription in early *Marchantia* embryos was found to be maternally biased, while the paternal chromosomes are silenced by a repressive histone modification (histone H3 lysine 27 trimethylation [H3K27me3]) deposited by maternally expressed polycomb repressive complex 2 (PRC2) proteins (at ~3 days after fertilization [DAF]).[72] It is unknown how maternally provided PRC2 distinguishes paternal from maternal chromosomes and deposits H3K27me3 across all genic regions. As the sperm genome is densely marked by 4mC over genic regions, 4mC may serve as the primary imprint that allows paternal genome recognition by PRC2. Consistent with this idea, we observed significant developmental defects in embryos produced from wild-type female and Mp*dn4mt1* mutant male plants (Figure 6C). Among embryos produced from manual crosses using a mixture of Mp*dn4mt1* mutant and wild-type sperm cells, we also found that Mp*dn4mt1*/+ heterozygous embryos are 91% under-represented than wild-type embryos. As embryos were examined ~35 DAF, the reduced Mp*dn4mt1*/+ heterozygous embryos may include those aborted during early embryo development. Consistently, embryos produced from PRC2 female mutants, which show disrupted paternal genome silencing, were reported to have a ~75% reduced survival rate compared with wild-type when examined at 32 DAF.[72]

Plants and animals diverged more than a billion years ago from a unicellular, sexual eukaryote.[73] However, germ cells and multicellularity evolved independently in both lineages. Loss of 4mC in *Marchantia* compromises transcriptional regulation and impairs sperm function, reminiscent of defects in mouse *Dnmt3A* mutant, which is defective in 5mC reprogramming.[74] The convergent evolution of DNA methylation reprogramming during gamete production in plants and mammals underscores the crucial role of epigenetic reprogramming in sexual reproduction. A comprehensive understanding of DNA methylation reprogramming across different lineages is therefore essential for elucidating the fundamental principles of sexual reproduction. Conversely, the drastic changes during reproduction provide opportunities to uncover epigenetic mechanisms, such as the extensive 4mC deposition in *Marchantia* sperm.

Limitations of the study

For the detection of DNA modifications, each method has inherent limitations and potential biases, which have contributed to the ongoing debate regarding the presence of 4mC in eukaryotes.[48] Our study addresses these concerns by employing multiple independent techniques and rigorous controls, providing robust evidence that 4mC is a functionally relevant DNA modification in *Marchantia*. Further investigations using complementary and emerging technologies will help explore the broader occurrence and biological significance of 4mC in eukaryotic systems.

Our findings indicate that 4mC exerts a suppressive effect on chromatin accessibility and transcription over gene bodies (Figures 5A and 5D). The mechanism underlying this suppression remains unclear. Moreover, our data suggest that the loss of 4mC-mediated repression in the Mp*dn4mt1-1* mutant dilutes transcripts crucial for sperm function. Specifically, transcripts related to cilium- or flagellum-dependent cell motility are most reduced in the mutant sperm compared with wild-type (Table S4B). It is yet unclear whether this dilution or other impacts of 4mC loss on chromatin and transcription is responsible for impaired motility; however, it is plausible that such dilution leads to reduced functional proteins due to competition for translational machinery. Consistent with this idea, inhibition of protein translation in mammalian sperm is known to impair sperm motility, highlighting the importance of translational activity in mature sperm.[75]

Additionally, 4mC is lost in 12 DAF embryos (containing above 1,000 cells; Figure 1C), but the mechanisms underlying 4mC erasure are mysterious. As MpDN4MT1a is specifically expressed in the sperm, 4mC can simply be lost through passive dilution by embryonic cell divisions, and/or an active DNA demethylation mechanism might be involved. In mammalian preimplantation embryos, global 5mC methylation is actively removed by conversion to 5-hydroxymethylcytosine (5hmC) by TETs and

passively lost through cell divisions.[27] In plants, 5mC is actively removed via the excision of the methylated cytosine by DNA glycosylases such as REPRESSOR OF SILENCING1 (ROS1) and is subsequently replaced with a fresh unmethylated cytosine via DNA repair mechanisms.[16] Two orthologs of *ROS1* exist in *Marchantia*, one of which (Mp*ROS1x*) is situated on the female sex chromosome and is expressed in the embryo,[76] providing a potential candidate for the erasure of paternal DNA methylation.

RESOURCE AVAILABILITY

Lead contact

Further requests and information concerning this study should be addressed to the lead contact, Xiaoqi Feng (Xiaoqi.Feng@ista.ac.at).

Materials availability

All the reagents are available on request to the lead contact, Xiaoqi Feng (Xiaoqi.Feng@ista.ac.at).

Data and code availability

All sequencing data generated in this study have been deposited to GEO: GSE254154. Published datasets used for analysis are available through SRA: SRX4628137-42, SRA: SRX7548543-48, SRA: DRX162967-72, SRA: SRX8604154-55, and SRA: SRX8604169-70. Published software used in this study includes Bismark v0.22.288, TrimGalore v0.4.1, kallisto v0.43.0, and sleuth v0.30.0 and are available from https://github.com/FelixKrueger/TrimGalore, https://github.com/pachterlab/kallisto, and https://github.com/pachterlab/sleuth. SMRT Link v9.0, containing pbmm2 and ipdSummary, is available from https://www.pacb.com/support/software-downloads/.

ACKNOWLEDGMENTS

We thank Sir Richard Roberts (NEB) for the kind gift of anti-4mC antibodies. We are also grateful to the JIC Small Molecule Mass Spectrometry (Lionel Hill) and Chemistry (Martin Rejzek) platforms as well as the High Resolution Metabolomics Laboratory (Manfred Beckmann, Aberystwyth University) for their assistance with LC-MS. Additionally, we acknowledge the assistance of the JIC Bioimaging Facility and ISTA Imaging and Optics Facility for microscopy. Finally, we appreciate the High Performance Computing resources provided by the ISTA Scientific Computing Facility and Norwich BioScience Institute Partnership Computing Infrastructure. This work was funded by a Sainsbury Charitable Foundation studentship (J.W.), a UKRI-BBSRC Doctoral Training Partnerships studentship (BBT0087171 to J.T.), a European Research Council Starting Grant ("SexMeth" 804981 to J.W., S.X., and X.F.), two Biotechnology and Biological Sciences Research Council (BBSRC) grants (BBS0096201 and BBP0135111 to J.Z., M.V., and X.F.), an EMBO Long Term Fellowship (Y.L.), an ISTA Bridge Fellowship (S.X.), and ISTA core funding (Y.Y. and X.F.).

AUTHOR CONTRIBUTIONS

J.W., J.Z., and X.F. designed the study. J.W., J.Z., Y.L., S.X., W.O., and J.T. performed the experiments. J.W., J.Z., Y.Y., and M.V. analyzed the data. L.D., K.N., and X.F. oversaw the execution of the study. J.W. and X.F. wrote the manuscript. All co-authors edited and commented on the initial draft.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

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

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Plant material and growth conditions
- METHOD DETAILS
  - Plasmid construction and transformation
  - Sperm, antheridium and embryo extraction
  - Sectioning and microscopy of antheridium halves
  - Expression microscopy
  - DNA dot blot immunoassay
  - Liquid chromatography–mass spectrometry
  - Violin plots and heat maps
  - Protein sequence alignment and phylogenetic analysis
  - RNA and DNA extraction
  - Sequencing-library construction
  - Sequencing-library analysis
  - Sperm and post-fertilization phenotyping
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2025.03.014.

Received: February 9, 2022
Revised: June 30, 2024
Accepted: March 7, 2025
Published: April 9, 2025

REFERENCES

1. Beaulaurier, J., Schadt, E.E., and Fang, G. (2019). Deciphering bacterial epigenomes using modern sequencing technologies. Nat. Rev. Genet. *20*, 157–172. https://doi.org/10.1038/s41576-018-0081-3.
2. Sánchez-Romero, M.A., and Casadesús, J. (2020). The bacterial epigenome. Nat. Rev. Microbiol. *18*, 7–20. https://doi.org/10.1038/s41579-019-0286-2.
3. Jurkowska, R.Z., and Jeltsch, A. (2022). Mechanisms and Biological Roles of DNA Methyltransferases and DNA Methylation: From Past Achievements to Future Challenges. Adv. Exp. Med. Biol. *1389*, 1–19. https://doi.org/10.1007/978-3-031-11454-0_1.
4. Marinus, M.G., and Løbner-Olesen, A. (2014). DNA Methylation. EcoSal Plus *6*, 1–35. https://doi.org/10.1128/ecosalplus.ESP-0003-2013.
5. Kong, Y., Cao, L., Deikus, G., Fan, Y., Mead, E.A., Lai, W., Zhang, Y., Yong, R., Sebra, R., Wang, H., et al. (2022). Critical assessment of DNA adenine methylation in eukaryotes using quantitative deconvolution. Science *375*, 515–522. https://doi.org/10.1126/science.abe7489.
6. O'Brown, Z.K., Boulias, K., Wang, J., Wang, S.Y., O'Brown, N.M., Hao, Z., Shibuya, H., Fady, P.-E., Shi, Y., He, C., et al. (2019). Sources of artifact in measurements of 6mA and 4mC abundance in eukaryotic genomic DNA. BMC Genom. *20*, 445. https://doi.org/10.1186/s12864-019-5754-6.
7. Boulias, K., and Greer, E.L. (2022). Means, mechanisms and consequences of adenine methylation in DNA. Nat. Rev. Genet. *23*, 411–428. https://doi.org/10.1038/s41576-022-00456-x.
8. Ye, P., Luan, Y., Chen, K., Liu, Y., Xiao, C., and Xie, Z. (2017). MethSMRT: an integrative database for DNA N6-methyladenine and N4-methylcytosine generated by single-molecular real-time sequencing. Nucleic Acids Res. *45*, D85–D89. https://doi.org/10.1093/nar/gkw950.
9. Rodriguez, F., Yushenova, I.A., DiCorpo, D., and Arkhipova, I.R. (2022). Bacterial N4-methylcytosine as an epigenetic mark in eukaryotic DNA. Nat. Commun. *13*, 1072. https://doi.org/10.1038/s41467-022-28471-w.
10. Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916–919. https://doi.org/10.1126/science.1186366.

11. de Mendoza, A., Lister, R., and Bogdanovic, O. (2020). Evolution of DNA methylome diversity in eukaryotes. J. Mol. Biol. *432*, 1687–1705. https://doi.org/10.1016/j.jmb.2019.11.003.

12. 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.

13. Luo, C., Hajkova, P., and Ecker, J.R. (2018). Dynamic DNA methylation: in the right place at the right time. Science *361*, 1336–1340. https://doi.org/10.1126/science.aat6806.

14. Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell *99*, 247–257. https://doi.org/10.1016/S0092-8674(00)81656-6.

15. Sharif, J., Muto, M., Takebayashi, S.-i., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908–912. https://doi.org/10.1038/nature06397.

16. 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.

17. Niederhuth, C.E., Bewick, A.J., Ji, L., Alabady, M.S., Kim, K.D., Li, Q., Rohr, N.A., Rambani, A., Burke, J.M., Udall, J.A., et al. (2016). Widespread natural variation of DNA methylation within angiosperms. Genome Biol. *17*, 194. https://doi.org/10.1186/s13059-016-1059-0.

18. Zhong, X., Du, J., Hale, C.J., Gallego-Bartolome, J., Feng, S., Vashisht, A.A., Chory, J., Wohlschlegel, J.A., Patel, D.J., and Jacobsen, S.E. (2014). Molecular mechanism of action of plant DRM de novo DNA methyltransferases. Cell *157*, 1050–1060. https://doi.org/10.1016/j.cell.2014.03.056.

19. 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.

20. Stroud, H., Greenberg, M.V.C., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. (2013). Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the *Arabidopsis* Methylome. Cell *152*, 352–364. https://doi.org/10.1016/j.cell.2012.10.054.

21. 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.

22. He, L., Huang, H., Bradai, M., Zhao, C., You, Y., Ma, J., Zhao, L., Lozano-Durán, R., and Zhu, J.-K. (2022). DNA methylation-free Arabidopsis reveals crucial roles of DNA methylation in regulating gene expression and development. Nat. Commun. *13*, 1335. https://doi.org/10.1038/s41467-022-28940-2.

23. Andrews, S., Krueger, C., Mellado-Lopez, M., Hemberger, M., Dean, W., Perez-Garcia, V., and Hanna, C.W. (2023). Mechanisms and function of de novo DNA methylation in placental development reveals an essential role for DNMT3B. Nat. Commun. *14*, 371. https://doi.org/10.1038/s41467-023-36019-9.

24. Mattei, A.L., Bailly, N., and Meissner, A. (2022). DNA methylation: a historical perspective. Trends Genet. *38*, 676–707. https://doi.org/10.1016/j.tig.2022.03.010.

25. Feinberg, A.P., Koldobskiy, M.A., and Göndör, A. (2016). Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. Nat. Rev. Genet. *17*, 284–299. https://doi.org/10.1038/nrg.2016.13.

26. 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.

27. Greenberg, M.V.C., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. Nat. Rev. Mol. Cell Biol. *20*, 590–607. https://doi.org/10.1038/s41580-019-0159-6.

28. 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.

29. Tang, W.W.C., Kobayashi, T., Irie, N., Dietmann, S., and Surani, M.A. (2016). Specification and epigenetic programming of the human germ line. Nat. Rev. Genet. *17*, 585–600. https://doi.org/10.1038/nrg.2016.88.

30. He, S., and Feng, X. (2022). DNA methylation dynamics during germline development. J. Integr. Plant Biol. *64*, 2240–2251. https://doi.org/10.1111/jipb.13422.

31. Walker, J., Gao, H., Zhang, J., Aldridge, B., Vickers, M., Higgins, J.D., and Feng, X. (2018). Sexual-lineage-specific DNA methylation regulates meiosis in Arabidopsis. Nat. Genet. *50*, 130–137. https://doi.org/10.1038/s41588-017-0008-5.

32. Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., et al. (2017). Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. Cell *171*, 287–304.e15. https://doi.org/10.1016/j.cell.2017.09.030.

33. Takuno, S., Ran, J.-H., and Gaut, B.S. (2016). Evolutionary patterns of genic DNA methylation vary across land plants. Nat. Plants *2*, 15222. https://doi.org/10.1038/nplants.2015.222. https://www.nature.com/articles/nplants2015222#supplementary-information.

34. Schmid, M.W., Giraldo-Fonseca, A., Rövekamp, M., Smetanin, D., Bowman, J.L., and Grossniklaus, U. (2018). Extensive epigenetic reprogramming during the life cycle of Marchantia polymorpha. Genome Biol. *19*, 9. https://doi.org/10.1186/s13059-017-1383-z.

35. Bewick, A.J., Niederhuth, C.E., Ji, L., Rohr, N.A., Griffin, P.T., Leebens-Mack, J., and Schmitz, R.J. (2017). The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. Genome Biol. *18*, 65. https://doi.org/10.1186/s13059-017-1195-1.

36. Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc. Natl. Acad. Sci. USA *89*, 1827–1831. https://doi.org/10.1073/pnas.89.5.1827.

37. Dai, Y., Yuan, B.F., and Feng, Y.Q. (2021). Quantification and mapping of DNA modifications. RSC Chem. Biol. *2*, 1096–1114. https://doi.org/10.1039/d1cb00022e.

38. Higo, A., Niwa, M., Yamato, K.T., Yamada, L., Sawada, H., Sakamoto, T., Kurata, T., Shirakawa, M., Endo, M., Shigenobu, S., et al. (2016). Transcriptional Framework of Male Gametogenesis in the Liverwort Marchantia polymorpha L. Plant Cell Physiol. *57*, 325–338. https://doi.org/10.1093/pcp/pcw005.

39. Montgomery, S.A., Tanizawa, Y., Galik, B., Wang, N., Ito, T., Mochizuki, T., Akimcheva, S., Bowman, J.L., Cognat, V., Maréchal-Drouard, L., et al. (2020). Chromatin organization in early land plants reveals an ancestral association between H3K27me3, transposons, and constitutive heterochromatin. Curr. Biol. *30*, 573–588.e7. https://doi.org/10.1016/j.cub.2019.12.015.

40. Yaari, R., Katz, A., Domb, K., Harris, K.D., Zemach, A., and Ohad, N. (2019). RdDM-independent de novo and heterochromatin DNA methylation by plant CMT and DNMT3 orthologs. Nat. Commun. *10*, 1613. https://doi.org/10.1038/s41467-019-09496-0.

41. Noy-Malka, C., Yaari, R., Itzhaki, R., Mosquna, A., Auerbach Gershovitz, N., Katz, A., and Ohad, N. (2014). A single CMT methyltransferase homolog is involved in CHG DNA methylation and development of Physcomitrella patens. Plant Mol. Biol. *84*, 719–735. https://doi.org/10.1007/s11103-013-0165-6.

42. Bujnicki, J.M., and Radlinska, M. (1999). Molecular evolution of DNA-(cytosine-N4) methyltransferases: evidence for their polyphyletic origin. Nucleic Acids Res. *27*, 4501–4509. https://doi.org/10.1093/nar/27.22.4501.

43. Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkienè, D., Butkus, V., and Janulaitis, A. (1989). Sequence motifs characteristic of DNA[cytosine-N4]methyltransferases: similarity to adenine and cytosine-C5 DNA-methylases. Nucleic Acids Res. *17*, 9823–9832. https://doi.org/10.1093/nar/17.23.9823.

44. Jurkowski, T.P., and Jeltsch, A. (2011). On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. PLoS One *6*, e28104. https://doi.org/10.1371/journal.pone.0028104.

45. Villarreal A, J.C., Crandall-Stotler, B.J., Hart, M.L., Long, D.G., and Forrest, L.L. (2016). Divergence times and the evolution of morphological complexity in an early land plant lineage (Marchantiopsida) with a slow molecular rate. New Phytol. *209*, 1734–1746. https://doi.org/10.1111/nph.13716.

46. Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., et al. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. *40*, D1178–D1186. https://doi.org/10.1093/nar/gkr944.

47. Vilkaitis, G., and Klimasauskas, S. (1999). Bisulfite sequencing protocol displays both 5-methylcytosine and N4-methylcytosine. Anal. Biochem. *271*, 116–119. https://doi.org/10.1006/abio.1999.4116.

48. Kong, Y., Mead, E.A., and Fang, G. (2023). Navigating the pitfalls of mapping DNA and RNA modifications. Nat. Rev. Genet. *24*, 363–381. https://doi.org/10.1038/s41576-022-00559-5.

49. Yu, M., Ji, L., Neumann, D.A., Chung, D.H., Groom, J., Westpheling, J., He, C., and Schmitz, R.J. (2015). Base-resolution detection of N4-methylcytosine in genomic DNA using 4mC-Tet-assisted-bisulfite-sequencing. Nucleic Acids Res. *43*, e148. https://doi.org/10.1093/nar/gkv738.

50. Davis, B.M., Chao, M.C., and Waldor, M.K. (2013). Entering the era of bacterial epigenomics with single molecule real time DNA sequencing. Curr. Opin. Microbiol. *16*, 192–198. https://doi.org/10.1016/j.mib.2013.01.011.

51. Clark, T.A., Murray, I.A., Morgan, R.D., Kislyuk, A.O., Spittle, K.E., Boitano, M., Fomenkov, A., Roberts, R.J., and Korlach, J. (2012). Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. Nucleic Acids Res. *40*, e29. https://doi.org/10.1093/nar/gkr1146.

52. Zhu, S., Beaulaurier, J., Deikus, G., Wu, T.P., Strahl, M., Hao, Z., Luo, G., Gregory, J.A., Chess, A., He, C., et al. (2018). Mapping and characterizing N6-methyladenine in eukaryotic genomes using single-molecule real-time sequencing. Genome Res. *28*, 1067–1078. https://doi.org/10.1101/gr.231068.117.

53. Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat. Methods *7*, 461–465. https://doi.org/10.1038/nmeth.1459.

54. Adam, S., Bräcker, J., Klingel, V., Osteresch, B., Radde, N.E., Brockmeyer, J., Bashtrykov, P., and Jeltsch, A. (2022). Flanking sequences influence the activity of TET1 and TET2 methylcytosine dioxygenases and affect genomic 5hmC patterns. Commun. Biol. *5*, 92. https://doi.org/10.1038/s42003-022-03033-4.

55. Ravichandran, M., Rafalski, D., Davies, C.I., Ortega-Recalde, O., Nan, X., Glanfield, C.R., Kotter, A., Misztal, K., Wang, A.H., Wojciechowski, M., et al. (2022). Pronounced sequence specificity of the TET enzyme catalytic domain guides its cellular function. Sci. Adv. *8*, eabm2427. https://doi.org/10.1126/sciadv.abm2427.

56. Xiong, J., Wang, P., Shao, W.-X., Li, G., Ding, J.-H., Xie, N.-B., Wang, M., Cheng, Q.-Y., Xie, C., Feng, Y.-Q., et al. (2022). Genome-wide mapping of N4-methylcytosine at single-base resolution by APOBEC3A-mediated deamination sequencing. Chem. Sci. *13*, 9960–9972. https://doi.org/10.1039/D2SC02446B.

57. Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. *20*, 207–220. https://doi.org/10.1038/s41576-018-0089-8.

58. Isbel, L., Grand, R.S., and Schübeler, D. (2022). Generating specificity in genome regulation through transcription factor sensitivity to chromatin. Nat. Rev. Genet. *23*, 728–740. https://doi.org/10.1038/s41576-022-00512-6.

59. Reynolds, W.F., and Wolfe, S.L. (1984). Protamines in plant sperm. Exp. Cell Res. *152*, 443–448. https://doi.org/10.1016/0014-4827(84)90645-1.

60. Reynolds, W.F., and Wolfe, S.L. (1978). Changes in basic proteins during sperm maturation in a plant, Marchantia polymorpha. Exp. Cell Res. *116*, 269–273. https://doi.org/10.1016/0014-4827(78)90448-2.

61. Santiago, J., Silva, J.V., Howl, J., Santos, M.A.S., and Fardilha, M. (2021). All you need to know about sperm RNAs. Hum. Reprod. Update *28*, 67–91. https://doi.org/10.1093/humupd/dmab034.

62. Lee, C.K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat. Genet. *36*, 900–905. https://doi.org/10.1038/ng1400.

63. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82. https://doi.org/10.1038/nature11232.

64. Ostermeier, G.C., Dix, D.J., Miller, D., Khatri, P., and Krawetz, S.A. (2002). Spermatozoal RNA profiles of normal fertile men. Lancet *360*, 772–777. https://doi.org/10.1016/S0140-6736(02)09899-9.

65. Jodar, M., Selvaraju, S., Sendler, E., Diamond, M.P., and Krawetz, S.A.; Reproductive; Medicine Network (2013). The presence, role and clinical use of spermatozoal RNAs. Hum. Reprod. Update *19*, 604–624. https://doi.org/10.1093/humupd/dmt031.

66. Pazour, G.J., and Witman, G.B. (2000). Forward and reverse genetic analysis of microtubule motors in Chlamydomonas. Methods *22*, 285–298. https://doi.org/10.1006/meth.2000.1081.

67. Klink, V.P., and Wolniak, S.M. (2001). Centrin Is Necessary for the Formation of the Motile Apparatus in Spermatids of Marsilea. Mol. Biol. Cell *12*, 761–776. https://doi.org/10.1091/mbc.12.3.761.

68. Koshimizu, S., Minamino, N., Nishiyama, T., Yoro, E., Sato, M., Wakazaki, M., Toyooka, K., Ebine, K., Sakakibara, K., Ueda, T., et al. (2022). Phylogenetic distribution and expression pattern analyses identified a divergent basal body assembly protein involved in land plant spermatogenesis. New Phytol. *236*, 1182–1196. https://doi.org/10.1111/nph.18385.

69. Gomes Pereira, S.G., Sousa, A.L., Nabais, C., Paixão, T., Holmes, A.J., Schorb, M., Goshima, G., Tranfield, E.M., Becker, J.D., and Bettencourt-Dias, M. (2021). The 3D architecture and molecular foundations of de novo centriole assembly via bicentrioles. Curr. Biol. *31*, 4340–4353.e7. https://doi.org/10.1016/j.cub.2021.07.063.

70. Shimamura, M. (2016). Marchantia polymorpha: Taxonomy, Phylogeny and Morphology of a Model System. Plant Cell Physiol. *57*, 230–256. https://doi.org/10.1093/pcp/pcv192.

71. Ma, J., Wang, S., Zhu, X., Sun, G., Chang, G., Li, L., Hu, X., Zhang, S., Zhou, Y., Song, C.-P., et al. (2022). Major episodes of horizontal gene transfer drove the evolution of land plants. Mol. Plant *15*, 857–871. https://doi.org/10.1016/j.molp.2022.02.001.

72. Montgomery, S.A., Hisanaga, T., Wang, N., Axelsson, E., Akimcheva, S., Sramek, M., Liu, C., and Berger, F. (2022). Polycomb-mediated repression of paternal chromosomes maintains haploid dosage in diploid embryos of Marchantia. eLife *11*, e79258. https://doi.org/10.7554/eLife.79258.

73. Embley, T.M., and Martin, W. (2006). Eukaryotic evolution, changes and challenges. Nature *440*, 623–630. https://doi.org/10.1038/nature04546.

74. Dura, M., Teissandier, A., Armand, M., Barau, J., Lapoujade, C., Fouchet, P., Bonneville, L., Schulz, M., Weber, M., Baudrin, L.G., et al. (2022). DNMT3A-dependent DNA methylation is required for spermatogonial stem cells to commit to spermatogenesis. Nat. Genet. *54*, 469–480. https://doi.org/10.1038/s41588-022-01040-z.

75. Gur, Y., and Breitbart, H. (2006). Mammalian sperm translate nuclear-encoded proteins by mitochondrial-type ribosomes. Genes Dev. *20*, 411–416. https://doi.org/10.1101/gad.367606.

76. Frank, M.H., and Scanlon, M.J. (2015). Transcriptomic Evidence for the Evolution of Shoot Meristem Function in Sporophyte-Dominant Land Plants through Concerted Selection of Ancestral Gametophytic and Sporophytic Genetic Programs. Mol. Biol. Evol. *32*, 355–367. https://doi.org/10.1093/molbev/msu303.

77. Kong, H., Lin, L.F., Porter, N., Stickel, S., Byrd, D., Posfai, J., and Roberts, R.J. (2000). Functional analysis of putative restriction–modification system genes in the Helicobacter pylori J99 genome. Nucleic Acids Res. *28*, 3216–3223. https://doi.org/10.1093/nar/28.17.3216.

78. Ikeda, Y., Nishihama, R., Yamaoka, S., Arteaga-Vazquez, M.A., Aguilar-Cruz, A., Grimanelli, D., Pogorelcnik, R., Martienssen, R.A., Yamato, K.T., Kohchi, T., et al. (2018). Loss of CG Methylation in Marchantia polymorpha Causes Disorganization of Cell Division and Reveals Unique DNA Methylation Regulatory Mechanisms of Non-CG Methylation. Plant Cell Physiol. *59*, 2421–2431. https://doi.org/10.1093/pcp/pcy161.

79. Arai, H., Yanagiura, K., Toyama, Y., and Morohashi, K. (2019). Genome-wide analysis of MpBHLH12, a IIIf basic helix-loop-helix transcription factor of Marchantia polymorpha. J. Plant Res. *132*, 197–209. https://doi.org/10.1007/s10265-019-01095-w.

80. Karaaslan, E.S., Wang, N., Faiß, N., Liang, Y., Montgomery, S.A., Laubinger, S., Berendzen, K.W., Berger, F., Breuninger, H., and Liu, C. (2020). Marchantia TCP transcription factor activity correlates with three-dimensional chromatin structure. Nat. Plants *6*, 1250–1261. https://doi.org/10.1038/s41477-020-00766-0.

81. Wu, T.-Y., Goh, H., Azodi, C.B., Krishnamoorthi, S., Liu, M.-J., and Urano, D. (2021). Evolutionarily conserved hierarchical gene regulatory networks for plant salt stress response. Nat. Plants *7*, 787–799. https://doi.org/10.1038/s41477-021-00929-7.

82. Ishizaki, K., Nishihama, R., Ueda, M., Inoue, K., Ishida, S., Nishimura, Y., Shikanai, T., and Kohchi, T. (2015). Development of Gateway Binary Vector Series with Four Different Selection Markers for the Liverwort Marchantia polymorpha. PLoS One *10*, e0138876. https://doi.org/10.1371/journal.pone.0138876.

83. Monte, I., Ishida, S., Zamarreño, A.M., Hamberg, M., Franco-Zorrilla, J.M., García-Casado, G., Gouhier-Darimont, C., Reymond, P., Takahashi, K., García-Mina, J.M., et al. (2018). Ligand-receptor co-evolution shaped the jasmonate pathway in land plants. Nat. Chem. Biol. *14*, 480–488. https://doi.org/10.1038/s41589-018-0033-4.

84. Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., et al. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380–1389. https://doi.org/10.1016/j.cell.2013.08.021.

85. Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., and Yamamoto, T. (2014). Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. Sci. Rep. *4*, 5400. https://doi.org/10.1038/srep05400.

86. Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., Huang, X., et al. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nat. Methods *11*, 399–402. https://doi.org/10.1038/nmeth.2857.

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. 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.

89. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

90. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

91. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/nar/gkw257.

92. Tian, F., Yang, D.-C., Meng, Y.-Q., Jin, J., and Gao, G. (2020). PlantRegMap: charting functional regulatory maps in plants. Nucleic Acids Res. *48*, D1104–D1113. https://doi.org/10.1093/nar/gkz1020.

93. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.

94. Katoh, K., Rozewicki, J., and Yamada, K.D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. *20*, 1160–1166. https://doi.org/10.1093/bib/bbx108.

95. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402. https://doi.org/10.1093/nar/25.17.3389.

96. Chiyoda, S., Ishizaki, K., Kataoka, H., Yamato, K.T., and Kohchi, T. (2008). Direct transformation of the liverwort Marchantia polymorpha L. by particle bombardment using immature thalli developing from spores. Plant Cell Rep. *27*, 1467–1473. https://doi.org/10.1007/s00299-008-0570-5.

97. Fujisawa, M., Hayashi, K., Nishio, T., Bando, T., Okada, S., Yamato, K.T., Fukuzawa, H., and Ohyama, K. (2001). Isolation of X and Y chromosome-specific DNA markers from a liverwort, Marchantia polymorpha, by representational difference analysis. Genetics *159*, 981–985. https://doi.org/10.1093/genetics/159.3.981.

98. Flores-Sandoval, E., Dierschke, T., Fisher, T.J., and Bowman, J.L. (2016). Efficient and Inducible Use of Artificial MicroRNAs in Marchantia polymorpha. Plant Cell Physiol. *57*, 281–290. https://doi.org/10.1093/pcp/pcv068.

99. Jones, V.A.S., and Dolan, L. (2017). MpWIP regulates air pore complex development in the liverwort Marchantia polymorpha. Development *144*, 1472–1476. https://doi.org/10.1242/dev.144287.

100. Tsuboyama, S., and Kodama, Y. (2014). AgarTrap: a simplified Agrobacterium-mediated transformation method for sporelings of the liverwort Marchantia polymorpha L. Plant Cell Physiol. *55*, 229–236. https://doi.org/10.1093/pcp/pct168.

101. Tsuboyama, S., Nonaka, S., Ezura, H., and Kodama, Y. (2018). Improved G-AgarTrap: A highly efficient transformation method for intact gemmalings of the liverwort Marchantia polymorpha. Sci. Rep. *8*, 10800. https://doi.org/10.1038/s41598-018-28947-0.

102. Feng, X., and Dickinson, H.G. (2010). Tapetal cell fate, lineage and proliferation in the Arabidopsis anther. Development *137*, 2409–2416. https://doi.org/10.1242/dev.049320.

103. Healey, A., Furtado, A., Cooper, T., and Henry, R.J. (2014). Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. Plant Methods *10*, 21. https://doi.org/10.1186/1746-4811-10-21.

104. Quinlivan, E.P., and Gregory, J.F. (2008). DNA digestion to deoxyribonucleoside: A simplified one-step procedure. Anal. Biochem. *373*, 383–385. https://doi.org/10.1016/j.ab.2007.09.031.

105. Linde, A.-M., Sawangproh, W., Cronberg, N., Szövényi, P., and Lagercrantz, U. (2020). Evolutionary history of the Marchantia polymorpha complex. Front. Plant Sci. *11*, 829. https://doi.org/10.3389/fpls.2020.00829.

106. Rich, M.K., Vigneron, N., Libourel, C., Keller, J., Xue, L., Hajheidari, M., Radhakrishnan, G.V., Le Ru, A., Diop, S.I., Potente, G., et al. (2021). Lipid

exchanges drove the evolution of mutualism during plant terrestrialization. Science *372*, 864–868. https://doi.org/10.1126/science.abg0929.

107. Marks, R.A., Smith, J.J., Cronk, Q., Grassa, C.J., and McLetchie, D.N. (2019). Genome of the tropical plant Marchantia inflexa: implications for sex chromosome evolution and dehydration tolerance. Sci. Rep. *9*, 8722. https://doi.org/10.1038/s41598-019-45039-9.

108. Linde, A.M., Singh, S., Bowman, J.L., Eklund, M., Cronberg, N., and Lagercrantz, U. (2023). Genome Evolution in Plants: Complex Thalloid Liverworts (Marchantiopsida). Genome Biol. Evol. *15*, evad014. https://doi.org/10.1093/gbe/evad014.

109. Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181. https://doi.org/10.1038/nprot.2014.006.

110. Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat. Methods *11*, 817–820. https://doi.org/10.1038/nmeth.3035.

111. Kingan, S.B., Heaton, H., Cudini, J., Lambert, C.C., Baybayan, P., Galvin, B.D., Durbin, R., Korlach, J., and Lawniczak, M.K.N. (2019). A High-Quality de novo Genome Assembly from a Single Mosquito Using PacBio Sequencing. Genes (Basel) *10*, 62. https://doi.org/10.3390/genes10010062.

112. Ding, P., Sakai, T., Krishna Shrestha, R., Manosalva Perez, N., Guo, W., Ngou, B.P.M., He, S., Liu, C., Feng, X., Zhang, R., et al. (2021). Chromatin accessibility landscapes activated by cell-surface and intracellular immune receptors. J. Exp. Bot. *72*, 7927–7941. https://doi.org/10.1093/jxb/erab373.

113. 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.

114. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

115. Higo, A., Kawashima, T., Borg, M., Zhao, M., López-Vidriero, I., Sakayama, H., Montgomery, S.A., Sekimoto, H., Hackenberg, D., Shimamura, M., et al. (2018). Transcription factor DUO1 generated by neofunctionalization is associated with evolution of sperm differentiation in plants. Nat. Commun. *9*, 5283. https://doi.org/10.1038/s41467-018-07728-3.

116. Tinevez, J.-Y., Perry, N., Schindelin, J., Hoopes, G.M., Reynolds, G.D., Laplantine, E., Bednarek, S.Y., Shorte, S.L., and Eliceiri, K.W. (2017). TrackMate: an open and extensible platform for single-particle tracking. Methods *115*, 80–90. https://doi.org/10.1016/j.ymeth.2016.09.016.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-N4-methyl-C	Kong et al.[77]	N/A
Anti-rabbit IgG, HRP	GE Healthcare	Cat#NA934V
Bacterial and virus strains		
Stellar Competent Cells	Takara	Cat#636763
Agrobacterium tumefaciens GV6620	GoldBio	Cat#CC-270-5x50
Chemicals, peptides, and recombinant proteins		
In-Fusion HD Cloning Plus	Takara	Cat#638909
N4-methyl-2'-dCTP	Trilink	Cat#O-0329
Deoxycytidine	Sigma	Cat#D3897-100MG
5-Me-2'-deoxycytidine	Cayman chemicals	Cat#16166
2'-Deoxy-N4-methylcytidine	CarboSynth	Cat#ND153424
N4,5-Dimethyldeoxycytidine	Toronto Research Chemicals	Cat#D468100
Unmethylated cl857 *Sam7* Lambda DNA	Promega	Cat#D152A
CG methyltransferase M. *SssI*	NEB	Cat#M0226M
Ligation high ver2	TOYOBO	Cat#LGK-201
Quick Taq ® HS DyeMix	2BSCIENTIFIC LTD	Cat#TYB-DTM-101
Critical commercial assays		
EpiTect Fast Bisulfite Conversion	Qiagen	Cat#59802
Ovation Ultralow Methyl-Seq	NuGEN	Cat#0336
Dynabeads® mRNA DIRECT™ Micro Kit	Invitrogen	Cat#00577121
Imprint DNA Modification Kit	Sigma	Cat#SLCF7072
Technovit 7100 kit	KULZER	Cat#T218
Gateway LR clonase II Enzyme Mix	Invitrogen	Cat#11791-020
Gateway BP clonase II Enzyme Mix	Invitrogen	Cat#11780-020
SMRTbell Express Prep kit	Pacific Biosciences	Cat#101-685-400
SMRTbell gDNA Sample Amplification Kit	Pacific Biosciences	Cat#101-980-000
Deposited data		
Raw and analyzed data	This paper	GEO: GSE254154
Differential gene expression analysis	Ikeda et al.,[78] Arai et al.,[79] Karaaslan et al.,[80] Wu et al.[81]	SRA: SRX4628137-42; SRA: SRX7548543-48; SRA: DRX162967-72; SRA: SRX8604154-55; SRA: SRX8604169-70
Experimental models: Organisms/strains		
Takaragaike-1 (Tak-1)	Bowman et al.[32]	N/A
Takaragaike-2 (Tak-2)	Bowman et al.[32]	N/A
Mp*dnmt3b* knockdown mutant	This paper	N/A
Mp*cmta* knockdown mutant	This paper	N/A
Mp*dn4mt1-1*	This paper	N/A
Mp*dn4mt1-2*	This paper	N/A
*pro*Mp*DN4MT1a*:Mp*DN4MT1a-Citrine*	This paper	N/A
*pro*Mp*DN4MT1b*:Mp*DN4MT1b-Citrine*	This paper	N/A
Mp*dn4mt1-1* (*pro*Mp*DN4MT1a*:Mp*DN4MT1a*)	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mp*dn4mt1-1* (*pro*Mp*DN4MT1b*:Mp*DN4MT1b*)	This paper	N/A
Mp*dn4mt1-1* (*pro*Mp*DN4MT1a*: Mp*DN4MT1a* - catalytic mutant)	This paper	N/A
Recombinant DNA		
pMpGWB103	Ishizaki et al.[82], Addgene	Cat#68557
pMpGWB207	Ishizaki et al.[82], Addgene	Cat#68598
pMpGE017	Monte et al.[83]	N/A
pMpGE_En04	Ran et al.,[84] Sakuma et al.,[85] Shen et al.[86]	N/A
pBC-GE12	Ran et al.,[84] Sakuma et al.,[85] Shen et al.[86]	N/A
pBC-GE23	Ran et al.,[84] Sakuma et al.,[85] Shen et al.[86]	N/A
pBC-GE34	Ran et al.,[84] Sakuma et al.,[85] Shen et al.[86]	N/A
pUC19	NEB	Cat#N3041S
*pro*Mp*DN4MT1a*:Mp*DN4MT1a-Citrine*	This paper	N/A
*pro*Mp*DN4MT1b*:Mp*DN4MT1b-Citrine*	This paper	N/A
*pro*Mp*DN4MT1a*:Mp*DN4MT1a*	This paper	N/A
*pro*Mp*DN4MT1b*:Mp*DN4MT1b*	This paper	N/A
*pro*Mp*DN4MT1a*:Mp*DN4MT1a* - catalytic mutant	This paper	N/A
Software and algorithms		
Bismark v0.22.288	Krueger and Andrews[87]	https://github.com/FelixKrueger/Bismark
Kallisto v0.43.0	Bray et al.[88]	https://github.com/pachterlab/kallisto
sleuth package v0.30.0	Pachter Lab	https://github.com/pachterlab/sleuth
TrimGalore v0.4.1	Felix Krueger	https://github.com/FelixKrueger/TrimGalore
DESeq 2 v3.16	Love et al.[89]	https://github.com/mikelove/DESeq2
ggplot2 v3.3.6	CRAN	https://cran.r-project.org/web/packages/ggplot2/index.html
pbmm2 v1.1.0	Pacific Biosciences	https://www.pacb.com/support/software-downloads/
ipdSummary v2.0	Pacific Biosciences	https://www.pacb.com/support/software-downloads/
Primrose v1.3	Pacific Biosciences	https://github.com/PacificBiosciences/primrose
pb-CpG-tools v1.1.0	Pacific Biosciences	https://github.com/PacificBiosciences/pb-CpG-tools/
ccs v6.4	Pacific Biosciences	https://github.com/PacificBiosciences/ccs
Bowtie 2 v2.3.4.1	Langmead and Salzberg[90]	https://github.com/BenLangmead/bowtie2
deepTools v3.1.1	Ramírez et al.[91]	https://github.com/deeptools/deepTools
PlantRegMap	Tian et al.[92]	http://plantregmap.gao-lab.org/index.php
ImageJ v1.53c	Schindelin et al.[93]	https://imagej.net/software/fiji/downloads
MAFFT v7	Katoh et al.[94]	https://mafft.cbrc.jp/alignment/server/
BoxShade v3.2; NCBI Protein BLAST	Altschul et al.[95]	https://sourceforge.net/projects/boxshade/; https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Plant material and growth conditions

Male and female *Marchantia polymorpha, L*. subsp. *ruderalis*, accessions Takaragaike-1 (Tak-1, male) and Takaragaike-2 (Tak-2, female), were used. The study also included *Mpdn4mt1-1*, *Mpdn4mt1-2*, *Mpdnmt3b* knockdown, and *Mpcmta* knockdown mutants,

as well as transgenic lines expressing *proMpDN4MT1a:MpDN4MT1a-Citrine* and *proMpDN4MT1b:MpDN4MT1b-Citrine* that were generated in this study. Plants were grown on plates containing ½ × strength Gamborg's B5 medium with 1% agar (Sigma) under constant light at 21 °C, and Tak-1 thallus tissue was collected from 2-week-old plants. Plants were transferred to Jiffy 7 peat compost pellets (Greens Hydroponics UK) after 2 weeks on plates and placed in a growth chamber at 21 °C with 70% humidity under constant light with far-red irradiation for induction of sexual reproduction, as described previously.[34,96] Spores were sterilized with Milton's sterilizing solution (Milton Pharmaceutical UK) as described.[32] Untransformed spores were grown to sexual maturity as above and used as wild-type (WT) controls in phenotypic studies. Male and female plants were determined with sex chromosome-specific markers, rbm27 and rhf72[97] (Table S5).

METHOD DETAILS

Plasmid construction and transformation

The Mp*DNMT3b* (Mp5g09290) knockdown construct was generated by synthesizing an artificial microRNA (amiRNA) construct (JW004; Table S5, ENSA), based on MpMIR160[98] and designed as described[99] to target an exon of Mp*DNMT3b*, and thereafter inserted into MpGWB103[82] using the Gateway system (Invitrogen). For the Mp*CMTa* (Mp6g08650) knockdown construct, the amiRNA sequence was amplified and modified with primers JW490, JW495, and JW539-542 (Table S5) to target an exon of Mp*CMTa* and inserted into MpGWB103.

To generate *pro*Mp*DN4MT1a:*Mp*DN4MT1a-Citrine* and *pro*Mp*DN4MT1b:*Mp*DN4MT1b-Citrine* constructs, Tak-1 genomic DNA for each gene (Mp6g18330 and Mp6g18340) plus 2-kb upstream sequences were amplified using primers YL176, YL257, YL178, and YL258 (Table S5), and inserted into the destination vector pMpGWB207[82] using the Gateway system (Invitrogen).

For CRISPR/Cas9 deletion of Mp*DN4MT1a* and Mp*DN4MT1b*, double nicking was carried out as previously described[84,86] using pMpGE017.[83] Briefly, four different gRNAs (Table S5) were cloned into the BsaI site of pMpGE_En04 vector or into the multiplex vectors pBC-GE12, pBC-GE23 or pBC-GE34.[84–86] The four gRNAs cassettes were then cloned into pMpGE017 binary vector by LR reaction (Invitrogen).

Gemmae (Mp*dnmt3b* and Mp*cmta* knockdown mutants) or F1 spores (Mp*dn4mt1-1* and Mp*dn4mt1-2* CRISPR/Cas9 deletion mutants and *pro*Mp*DN4MT1a/b:*Mp*DN4MT1a/b-Citrine* lines) were transformed with the above constructs using agrobacteria transformation (strain GV6620) and positive transformants were selected on hygromycin (Mp*dnmt3b* and Mp*cmta* knockdown mutants, and Mp*dn4mt1-1* and Mp*dn4mt1-2* knockout mutant) or gentamycin (*pro*Mp*DN4MT1a/b:*Mp*DN4MT1a/b-Citrine* lines), as described.[100,101] The knockdown lines were checked by qPCR using antheridiophores as previously described,[31] with primers JW778 and JW779 for Mp*cmta* knockdown, primers JW448 and JW449 for Mp*dnmt3b* knockdown, and primers JW438 and JW439 (Table S5) to compare expression to Mp*EF1α* (Mp3g23400). Mp*dn4mt1-1 and* Mp*dn4mt1-2* deletion was detected via PCR with primers YL1 and YL2 and sequenced with YL1 (Table S5). For Mp*dn4mt1-1* an 8807-bp deletion (Chr6: 20,611,704 - 20,620,510) including the tandem repeats of Mp*DN4MT1a* and *1b* (Chr6: 20,613,540 – 20,619,372) was obtained except for the inclusion of a 9-bp sequence (AGGTTTCAA; Chr6: 20,612,442 – 20,612,450) outside Mp*DN4MT1s*. For Mp*dn4mt1-2* a similar deletion was obtained (8,785-bp; Chr6: 20,611,726 – 20,620,510) except for the inclusion of a 26-bp sequence (unplaced).

To generate complementation lines, *pro*Mp*DN4MT1a:*Mp*DN4MT1a* and *pro*Mp*DN4MT1b:*Mp*DN4MT1b* were inserted into the destination vector pMpGWB201[82] by the Gateway system (Invitrogen) using primers YL176, YL257, YL178, and YL258 (Table S5). Mp*DN4MT1a* was also mutated at the catalytic site (TSPPY to AAAAA) and inserted into pMpGWB201 using primers YL176, YL629, YL630, and YL257. F1 spores from a cross between Mp*dn4mt1-1* and Tak-2 females were transformed with the complementation constructs individually (*pro*Mp*DN4MT1a:*Mp*DN4MT1a*, *pro*Mp*DN4MT1b:*Mp*DN4MT1b*, and *pro*Mp*DN4MT1a:*Mp*DN4MT1a*-catalytic mutant) using agrobacteria transformation (strain GV6620) and positive transformants were selected on gentamycin, as described.[100,101] Successful transformants within the Mp*dn4mt1-1* background were then confirmed with primers YL1, YL2, YL107, YL108, YL115, and YL116.

Sperm, antheridium and embryo extraction

For sperm, drops of sterile water (approximately 15 μl) were applied to the upper surface of at least 50 antheridiophore heads at stage 5.[38] Water was collected after waiting at least 2 minutes to allow sufficient sperm release. The collected solution (approximately 750 μl was used for each RNA- or bisulfite-seq library construction) was centrifuged at 10,000 rpm for 10 minutes to pellet sperm, which was subsequently resuspended in 20 μl of sterile water. For each antheridium, antheridiophore receptacles of Stage 5[38] were manually cut longitudinally with razor blades to expose antheridia. Individual antheridia were thereafter picked up with tweezers and each was transversely cut in half using fine needles (0.5 mm x 25 mm). One half was placed into a DNase-/RNase-free PCR tube and immediately frozen in liquid nitrogen, while the other half was deposited into a separate PCR tube containing fixation buffer (4% paraformaldehyde, PBS, pH 7.5, 0.1% Triton X-100, 0.1% Tween-20). Embryos were dissected from fertilized gametophores 12 days following sperm application and the surrounding pseudoperianths were removed.

Sectioning and microscopy of antheridium halves

With the antheridium halves in fixation buffer, a series of ethanol dehydration steps were carried out followed by resin embedding using a Technovit 7100 kit (Electron Microscopy Sciences). The fixed antheridium halves were then cut into 5 μm sections with a

microtome RM2250 (Leica) and stained with Toluidine blue.[102] Imaging was carried out with an Axio Imager Z2 (Zeiss) with Hamamatsu camera. Based on the morphology of containing spermatocytes/spermatids, antheridia are categorized into early (containing spermatocytes), middle (early spermiogenesis, containing mostly round spermatids) or late (late spermiogenesis when DNA compaction and sperm tail formation occur, containing mostly elongating spermatids) stages.

Expression microscopy

Antheridiophores of *pro*Mp*DN4MT1a/b*:Mp*DN4MT1a/b-Citrine* lines corresponding to stage 5,[38] with obvious anthocyanin accumulation in the receptacle that suggest mature sperm formation, were manually cut longitudinally. Antheridiophore slices were subsequently prepared in VECTASHIELD mounting medium (H-1000, Vector Laboratories) for observation under a fluorescence stereo microscope (Leica M205FA), and fluorescence was captured with a Leica DFC310FX color camera illuminated by LED 470nm-GFP. Antheridia with Citrine signal were manually taken out and observed under a confocal microscope (Leica SP8X) with a 40x/1.30 oil objective.

DNA dot blot immunoassay

dCTP and 4mdCTP PCR products were generated from pUC19 (NEB) synthesized with either dCTP or N4-methyl-dCTP (4mdCTP; Trilink) using primers JZ294/JZ295 (Table S5). Thallus and sperm DNA was extracted as described.[103] 60 ng DNA for each sample was spotted onto a nitrocellulose membrane (Millipore), air-dried and fixed by UV cross-linking (1.2 mJ/cm^2 for 1 min). The cross-linked membrane was then blocked in a 5% non-fat milk in PBST (containing 0.05% v/v Tween) for 1 h. The membrane was then incubated with the primary antibody (1: 2000 anti-N4-methyl-C antibody) at room temperature for 2 h. The membrane was washed three times with PBST (10 min each), incubated with 1:10000 secondary donkey anti-rabbit antibody conjugated with horseradish peroxidase (HRP; GE Healthcare) at room temperature for 1 h, and finally washed three times with PBST (10 min each). Positive signals were detected using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific). DNA dots were visualized by developing the X-ray film which captured the light emitted by destabilized LumiGLO substrates in a dark room.

Liquid chromatography–mass spectrometry

Thallus and sperm DNA was digested into deoxyribonucleosides as previously described.[104] The digested samples were centrifuged through 0.2 μm filters and 5 μl was analyzed by LC-MS/MS (Waters Aquity UPLC and XevoTQS MS). Separation was carried out on a 100 × 2.1 mm 1.7 μ Kinetex C18 column (Phenomenex) using the following gradient of methanol (solvent B) versus 0.1% formic acid buffered to pH 8.3 with ammonium hydroxide (solvent A), run at 40 °C: 0 min, 0% B; 2 min, 0% B; 4 min, 10% B; 4.5 min, 90% B; 5.5 min, 90% B; 5.6 min, 0% B; 9.6 min, 0% B. The flow rate was 0.5 ml·min^{-1} except during the periods 4.0 – 4.5 min and 5.5 – 6.5 min to avoid excessive pressure. Analytes were quantified against external standard curves by multiple reaction monitoring (MRM) detecting the following transitions from positive-mode electrospray: 228→112 (unmethylated C), 242→126 (4mC and 5mC), and 256→140 (4,5mC). 4mC and 5mC eluted at different retention times. Collision energies and cone voltages were optimized using Waters' Intellistart software. Spray chamber conditions were 3.5 kV spray voltage, 500 °C desolvation temperature, 900 l·hr^{-1} desolvation gas, 150 l·hr^{-1} cone gas, and 7.0 bar nebulizer pressure. Deoxycytidine (Sigma), 5-Me-2'-deoxycytidine (Cayman chemicals), 2'-Deoxy-N4-methylcytidine (CarboSynth), and N4,5-Dimethyldeoxycytidine (Toronto Research Chemicals) were used as standards.

Violin plots and heat maps

For all the violin plots in this study, the plot shows the distribution of the data and its probability density, with the box enclosing the middle 50% of the distribution (the horizontal line marks the median).

Heat maps (otherwise referred to as circus plots; www.circos.ca) were generated with fractional CG, CHG, and CHH methylation of 10-kb genomic windows.

Protein sequence alignment and phylogenetic analysis

Protein sequences representing 5mC and β-class 4mC methyltransferases were obtained[42,44] and closest methyltransferase homologs were found using Standard Protein BLAST of N4CMT_A and MpDN4MT1s against the NCBI database with default parameters.[95] Protein sequences up to motif III were taken from 5mC methyltransferases and relocated onto the C-terminus. Sequences were aligned with MAFFT (v7) using default settings.[94] A phylogenetic tree was constructed with the neighbor joining method using 100 bootstraps. A multiple sequence alignment of MpDN4MT1a, MpDN4MT1b, and prokaryotic 4mC methyltransferases was obtained and shaded with BoxShade (v3.2) using default parameters (http://sourceforge.net/projects/boxshade/). Motifs were labelled according to published data.[42] Amino acid sequences of MpDN4MT1a/b homologs from *Marchantia polymorpha L.* subsp. *Polymorpha* (MppBR5_0228s0040 and MppBR5_0228s0050), *Marchantia polymorpha L.* subsp. *Montivagans* (MpmSA2_0387s0050.1 and MpmSA2_0387s0050.2), *Marchantia paleacea* (Marpal_scaff4134_Fgenesh3.1), *Marchantia inflexa* (unannotated), and *Lunularia cruciata* (Lc_composite_17444) were obtained via BLAST of the respective genomes[105–108] and sequences were refined using published annotations in combination with the *Marchantia polymorpha L.* subsp. *Ruderalis* sequence structure. Due to the incomplete genome sequence of *Marchantia inflexa*, the N-terminus of MpDN4MT1a/b homologous gene is likely truncated. A second copy of MpDN4MT1a/b was also detected in *Marchantia paleacea* (Marpal_scaff4134_Fgenesh4.1) with the key TSPPY motif but the

complete protein sequence could not be resolved. A multiple sequence alignment of MpDN4MT1a, MpDN4MT1b and these homologs was created and annotated as above. Alternative plant genomes (259 published genomes in total) were explored via BLAST of the protein sequences of MpDN4MT1a and MpDN4MT1b with Phytozome (v13)[46] using default parameters, and no homologs were identified.

RNA and DNA extraction

RNA extraction was performed as previously described.[109] For antheridium halves, genomic DNA extraction was also performed from the same sample using this protocol.[110] Briefly, samples were ground in sodium lysis buffer (100 mM Tris pH 7.5, 500 mM NaCl, 10 mM EDTA, 5 mM DTT, 1% SDS) in a PCR tube by a plastic pestle. RNA was sequestered with Dynabeads® Oligo $(dT)_{25}$ from the mRNA DIRECT™ Micro Kit (Invitrogen), while genomic DNA in the supernatant was pipetted out into a new collection tube and treated with pronase for protein digestion. Thallus, sperm and embryo DNA were extracted based on a CTAB-modified protocol as described.[103]

Sequencing-library construction

4mC and 5mC spike-in controls were generated from pUC19 (NEB) PCR product synthesized with N4-methyl-dCTP (4mdCTP; Trilink) using primers JZ294/JZ295 (Table S5) and genomic lambda cl857 *Sam7* DNA (Promega) methylated with 5mC CG methyltransferase M. *SssI* (NEB), respectively, and added to DNA samples prior to library construction where appropriate for testing bisulfite conversion, as previously described.[49]

RNA-seq library preparation was carried out as previously described.[109] Briefly, the Smart-seq2 method was employed to generate full-length cDNA and to allow sequencing via standard reagents. RNA-seq libraries were sequenced on NextSeq 500 (Illumina) to produce 75-bp single-end data. RNA-seq data for thallus and embryo was obtained from published sources.[38,76]

For antheridium halves (two biological replicates each for early, middle, and late stage used in Figures 1 and S1), sperm (one biological replicate used in Figures 1, 2, and S1), Mp*DNMT3b* and Mp*CMTa* knockdown sperm (one biological replicate each used in Figures 2 and S1), a single-cell-based method for BS-seq library preparation was carried out as previously described.[110] For one biological replicate each of thallus, embryo, WT sperm and Mp*dn4mt1-1* sperm, single-end BS-seq libraries were also constructed with Ovation Ultralow Methyl-Seq Library Systems (Nugen, 0336) and EpiTect Bisulfite Kit (Qiagen, 59104) according to manufacturer's instructions, with the incorporation of either one (WT sperm, Mp*dn4mt1-1* mutant sperm, and Mp*dn4mt1-2* mutant sperm used in Figures 4 and S5A) or two rounds (WT sperm used in Figure S4, and thallus and embryo used in Figures 1 and S1) of bisulfite conversion.

Native WT and Mp*dn4mt1-1* mutant sperm libraries for SMRT-seq were prepared using the SMRTbell Express Prep kit (v2.0; Pacific Biosciences, Menlo Park, CA, USA) following the low DNA input workflow as described.[111] Whole genome amplified libraries were also constructed, utilizing the SMRTbell gDNA Sample Amplification Kit (Pacific Biosciences, Menlo Park, CA, USA). Libraries were sequenced under CCS running mode (one Sequel II SMRT Cell for each native library and one Sequel II SMRT Cell for the amplified libraries) via DNA Link (Seoul, Korea).

WT and Mp*dn4mt1-1* mutant 4mC-TAB-seq sperm libraries were prepared with Ovation Ultralow Methyl-Seq Library Systems (Nugen, 0336) and EpiTect Bisulfite Kit (Qiagen, 59104) according to manufacturer's instructions, with the incorporation of one round of bisulfite conversion. Prior to adaptor ligation, sheared DNA was treated with TET2 (NEB #E7120S) following manufacturer's instructions. WT and Mp*dn4mt1-1* mutant 4mC-AMD-seq sperm libraries were prepared with the ACCEL-NGS methyl seq DNA library kit (Qiagen, 59104) after treatment of sheared DNA with APOBEC (NEB #E7120S), following manufacturer's instructions.

For ATAC-seq libraries, *Marchantia* sperm nuclei were collected and sorted by FACS through SYBR Green staining. Approximately 10,000 sperm nuclei were used for each replicate. ATAC-seq libraries were constructed using Illumina Tagment DNA TDE1 Enzyme and Buffer Kits, catalogue number: 20034197, as described.[112] Briefly, sorted nuclei were kept on ice and centrifuged at 4 °C, 800 g for 10 min before use. Most of the supernatant was removed and 5 μl of the tagmentation mix (2.5 μl tagmentation buffer, 0.5 μl enzyme, and nuclease free water) was added. Samples were incubated at 37 °C for 30 min, then the samples were amplified for 5 cycles (72 °C 5 min, 98 °C 30 s, (98 °C 10 s, 63 °C 30 s, 72 °C 1 min) x 5 cycles, 72 °C 10 min final extension). Amplified DNA was quantified by qPCR as previously described[31] before any remaining additional PCR cycles required was performed. Libraries were sequenced on NextSeq 500 (Illumina) to produce 75-bp paired-end data.

Sequencing-library analysis

For RNA-seq, low-quality reads and potential adaptor sequences were first trimmed using TrimGalore (v0.4.2) with default parameters. RNA-seq reads were processed with kallisto (v0.43.0) and statistical analysis was carried out with the sleuth package (v0.30.0) using the *Marchantia* (v5.1) gene annotation. Transcripts > 10 TPM in either WT or Mp*dn4mt1-1* were retained and those with a $-\log_{10}$ (*Q*-value) < 0.1 between the two genotypes were discarded, leaving a total of 5,970 genes. Gene Ontology analysis was performed using the Plant Transcriptional Regulatory Map (PlantRegMap; Center for Bioinformatics Peking University).[92]

For BS-seq, low-quality reads and potential adaptor sequences were first trimmed using TrimGalore (v0.4.2) with default parameters. Biological replicates for antheridia were combined and DNA methylation analysis was performed using Bismark.[87] Methylation data was mapped to the *Marchantia* chromosome assembly (v5.1)[39] as well as to the pUC19 vector sequence and lambda cl857 *Sam7* genome where appropriate.

SMRT libraries were mapped to the *Marchantia* (v5.1) genome using pbmm2 (v1.1.0). Base modifications were detected using ipdSummary (v2.0) with the following flags set (–identify 4mC –methylFraction). For 5mCG analysis, circular consensus reads were generated with ccs (v6.4) using the "—hifi-kinetics" option to generate consensus kinetics tags, and primrose (v1.3) was used to predict 5mC modification of each CG motif and generate base modification ("MM") and base modification probability ("ML") BAM tags. Pileup-based consensus methylation sites and probabilities were generated by the script "aligned_bam_to_cpg_scores.py" from pb-CpG-tools (v1.1.0) with the "-q 1 -m reference -p model -c 10" options.

Transposon and gene meta-analysis as well as 5mC methylation analysis was performed as previously described.[113] The proportion of the *Marchantia* genome considered to be methylated in thallus and during sperm development was taken as the percentage of 100-bp genomic windows (w100s) with fractional methylation > 0.2 for CG, > 0.1 for CHG, > 0.05 for CHH, > 0.1 for non-CG, or >0.1 for all cytosines (Cmethyl) out of the total number of sequenced w100s in the respective context. For comparisons between WT and knockdown sperm, a fractional methylation level > 0.3 in each of the contexts was used. w100s were considered to be associated with repeats if they overlapped or were within 500-bp of the *Marchantia* (v5.1) transposon annotation. For comparisons to LC-MS, methylation levels from BS-seq data were calculated using the average fractional methylation of single cytosines from Cmethyl. For SMRT-seq data, the proportion of N4 methylated cytosines for each w100 was taken as the average fractional methylation for 4mC cytosines detected over the total number of cytosines within the w100 sequence for each cytosine context. The same was performed for 5mCG analysis of SMRT-seq data. Violin plots were generated using w100s with at least ten informative sequenced cytosines in the respective context for BS-seq data while all w100s were considered for SMRT-seq data. For both violin plots and calculations regarding the fractional 5mC and 4mC methylation of repeats, only w100s overlapping the transposon annotation with CG fractional methylation > 0.5 in thallus were considered.

Profiles of ATAC-seq reads were generated using bigwig files with a bin size of 50. First, ATAC and input reads were mapped to the *Marchantia* genome (v5.1) with Bowtie 2(v2.3.4.1).[90] The read counts were calculated using featureCounts (v2.0.6),[114] and the scale factors were estimated using DESeq2 (v3.16).[89] bigWig files were then created from these mapped reads using the command bamCompare with a bin size of 50 and the calculated scale factors. Spearman correlations between ATAC-seq replicates and antheridia RNA-seq replicates were calculated using multiBamSummary and plotCorrelation.

Sperm and post-fertilization phenotyping

Swimming dynamics of *Marchantia* sperm were examined as previously described,[115] with modifications. Sperm discharged in water were observed under an Axio Imager Z2 (Zeiss) with Hamamatsu camera and 40x/0.75 air objective in dark field and video-recorded at the rate of 10 frames per second (fps) for 10s. To track sperm, TrackMate (v6.0.1),[116] a plugin bundled in ImageJ (v1.53c; the Fiji distribution)[92] was used. Swimming velocity was calculated by track displacement divided by track duration time. Directionality was calculated by track displacement divided by sum of links displacement.

To examine the fertility of Mp*dn4mt1-1* mutant sperm, freshly collected sperm cells from WT and Mp*dn4mt1-1* mutant plants were quantified by measuring the optical density at a wavelength of 600 nm (OD600). Subsequently, the sperm samples were equally mixed according to the OD600 values, and the mixed sample was applied to archegoniophores of sexually mature WT Tak-2 plants. At 35 DAF, embryos were collected and genotyped using primers YL1, YL2, and SX382 (Table S5).

For post-fertilization phenotyping, freshly collected sperm was crossed to sexually mature WT Tak-2 plants. The archegoniophores were dissected every two days from 35 DAF to record days required for maturation and the numbers of normal and defective embryos. Normal mature embryos were identified as yellow ones that produce yellow spores; defective embryos were identified as dry, brown embryos that did not produce spores (Figure S6H).

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed using R (3.6.3). The statistical tests used for each experiment are listed in figure legends. Before applying parametric tests, we assessed whether the data met the required assumptions. For datasets that did not meet normality assumptions, non-parametric alternatives were used. For sperm motility assays (Figures 6A and 6B), motile sperm were analyzed per condition (WT: n = 30, Mp*dn4mt1-1*: n = 28, Mp*dn4mt1-2*: n = 31, MpDN4MT1a reintroduced into Mp*dn4mt1-1*: n = 31, Mp*dnmt3b* kd: n = 31, Mp*cmta* kd: n = 33). Velocity and directionality differences were assessed using a Kolmogorov-Smirnov test. For embryo viability assays (Figure 6C), sample sizes ranged from n = 193 to 416 embryos per condition, representing the number of embryos analyzed for viability post-fertilization. Statistical significance was determined using a Fisher's exact test. For embryo maturation timing (Figure 6D), the number of embryos assessed per condition was recorded, with development tracked at 35 days after fertilization (DAF) and analyzed using a Kolmogorov-Smirnov test to compare maturation timing distributions. For gene expression analysis by qRT-PCR (Figure S1E), n = 3 biological replicates per condition were analyzed to assess transcript levels, with statistical significance determined using a two-tailed t-test.

CellPress
OPEN ACCESS

Supplemental figures

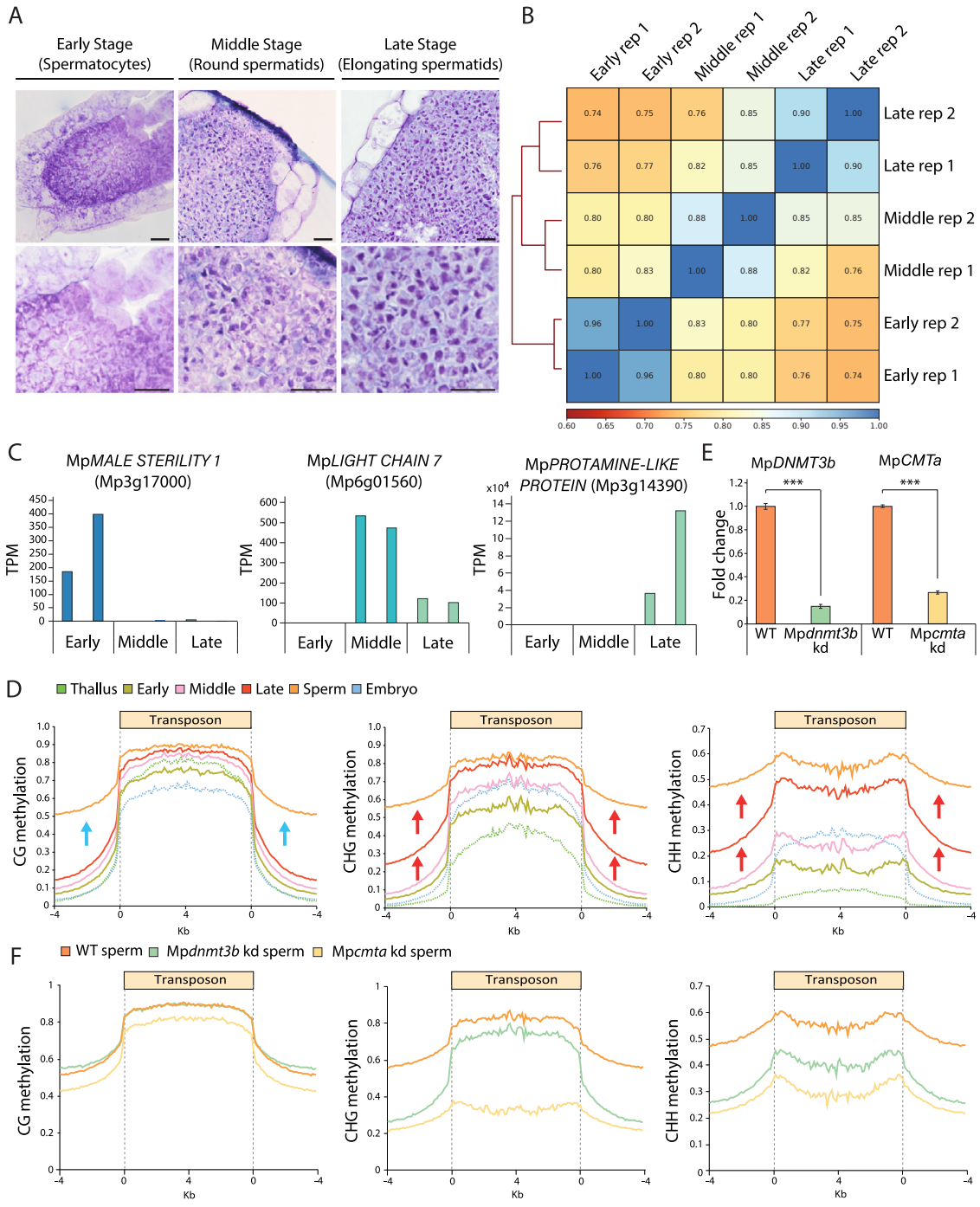


Figure S1. Gene expression and DNA methylation during *Marchantia* spermatogenesis, related to Figures 1 and 2

(A) Transverse sections for each second biological replicate of developing antheridia stained with toluidine blue, staged as in Figure 1B. Scale bars, 20 μm.

(B) Heatmap showing the Spearman correlation coefficient of pairwise comparison between the transcriptomes from each of the six individual antheridia.

(C) Transcript abundances detected in early, middle, or late stage for the six individual antheridia for three genes previously known to be developmentally regulated as shown by RNA *in situ* hybridization.[38] TPM, transcripts per million.

(legend continued on next page)

(D) DNA methylation over TEs in the antheridia (of early, middle, or late stages) and sperm (illustrated by solid lines), in comparison to the thallus and embryo (dashed green and blue lines, respectively). TEs were aligned at the 5′ and 3′ ends (dashed gray lines), and average methylation levels for each 100-bp interval are plotted. Blue arrows indicate expansion of CG methylation to non-repetitive regions at the last stage of sperm maturation (after the late stage, the red line). Red arrows illustrate expansion of non-CG methylation into non-repetitive regions in antheridia, starting from the middle stage (pink line).
(E) RT-qPCR showing the expression of Mp*DNMT3b* and Mp*CMTa* in Mp*dnmt3b* KD or Mp*cmta* KD mutant antheridiophores. ***$p < 0.001$, two-tailed t test; $n = 3$.
(F) Methylation over TEs in WT, Mp*dnmt3b* KD, and Mp*cmta* KD mutant sperm, as in (D).

Figure S2. MpDN4MT1a and MpDN4MT1b share homology with prokaryotic 4mC Dnmts, related to Figure 3

Protein sequence alignment of MpDN4MT1a and b with β-class 4mC methyltransferases.[42] Amino acids with shared identity between sequences are shaded green while similar amino acids are shaded blue. The TSPPY motif characteristic of 4mC methyltransferases is highlighted in orange.

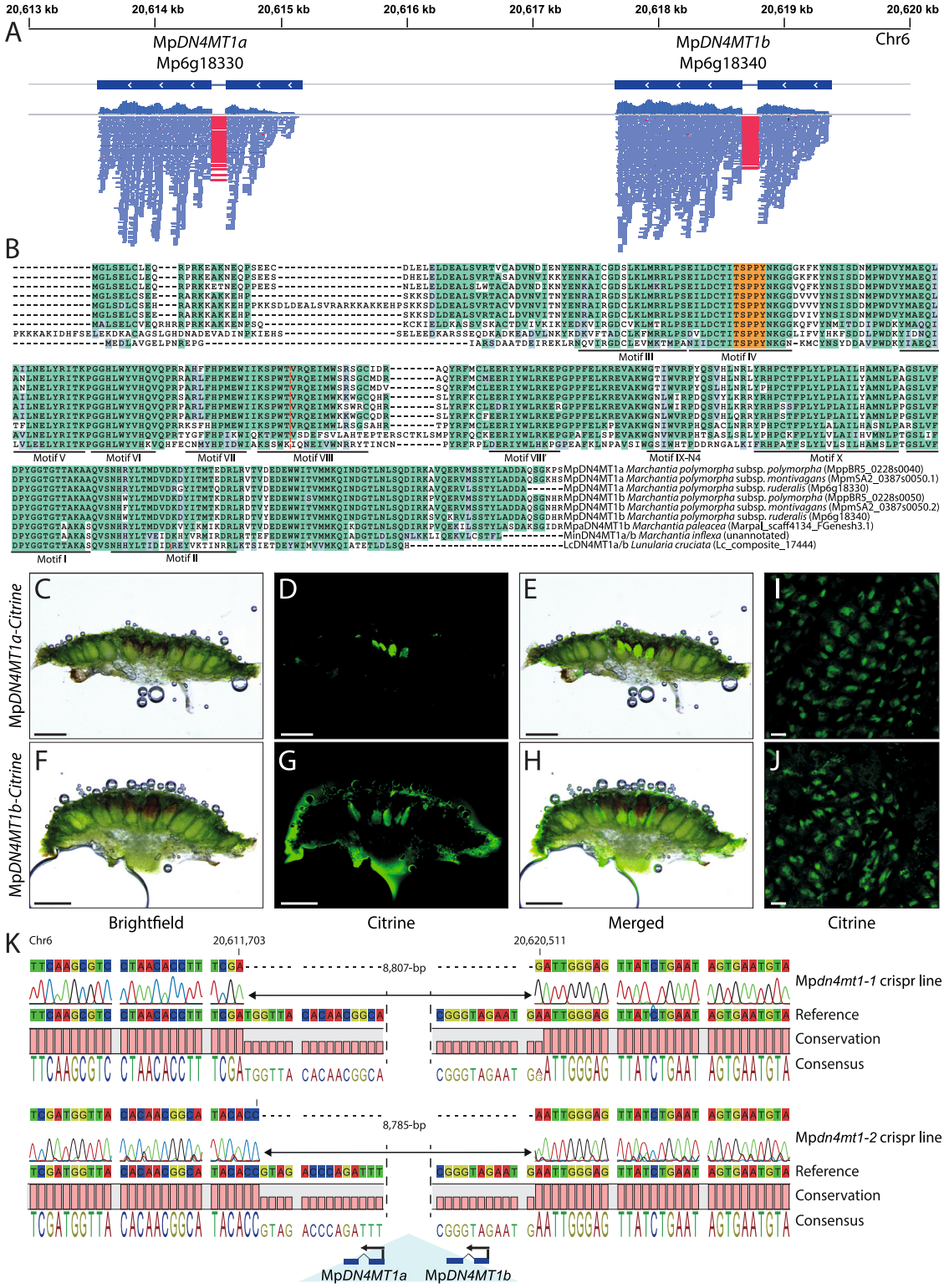


Figure S3. Genomic, expression, and mutant characterization of MpDN4MT1a and MpDN4MT1b, related to Figure 3

(A) Alignment of WT sperm RNA-seq reads to the *Marchantia polymorpha* v5.1 genomic region containing the genes that encode MpDN4MT1a and MpDN4MT1b. Red lines indicate split reads that map across the introns of each gene.

(legend continued on next page)

(B) Protein sequence alignment of MpDN4MT1a and b from the three *Marchantia polymorpha* subspecies *ruderalis* (used in this study), *polymorpha*, and *montivagans*, a MpDN4MT1a/b homolog from *Marchantia paleacea*, as well as a MpDN4MT1a/b homolog each from *Marchantia inflexa* and *Lunularia cruciata*. Amino acids with shared identity between sequences are shaded green while similar amino acids are shaded blue. The TSPPY motif characteristic of 4mC methyltransferases is highlighted in orange. Dashed red line indicates the corresponding splice site of each gene.
(C–H) Longitudinal sections of Mp*DN4MT1a-citrine* (C–E) and Mp*DN4MT1b-citrine* (F–H) antheridiophores, showing mature antheridia with positive signals. Antheridiophores correspond to stage 5,[38] with obvious anthocyanin accumulation in the receptacle, indicating the presence of antheridia containing mature sperm. Scale bars, 500 μm.
(I and J) Zoomed-in view of elongating spermatids, in which MpDN4MT1a-citrine (I) or MpDN4MT1b-citrine (J) signals are specifically observed. Scale bars, 5 μm.
(K) DNA sequence alignment illustrating the 5′ and 3′ ends for the CRISPR-Cas9 deletion of Mp*DN4MT1a* and Mp*DN4MT1b* in the Mp*dn4mt1-1* and Mp*dn4mt1-2* mutants. Exons are indicated by dark blue bars.

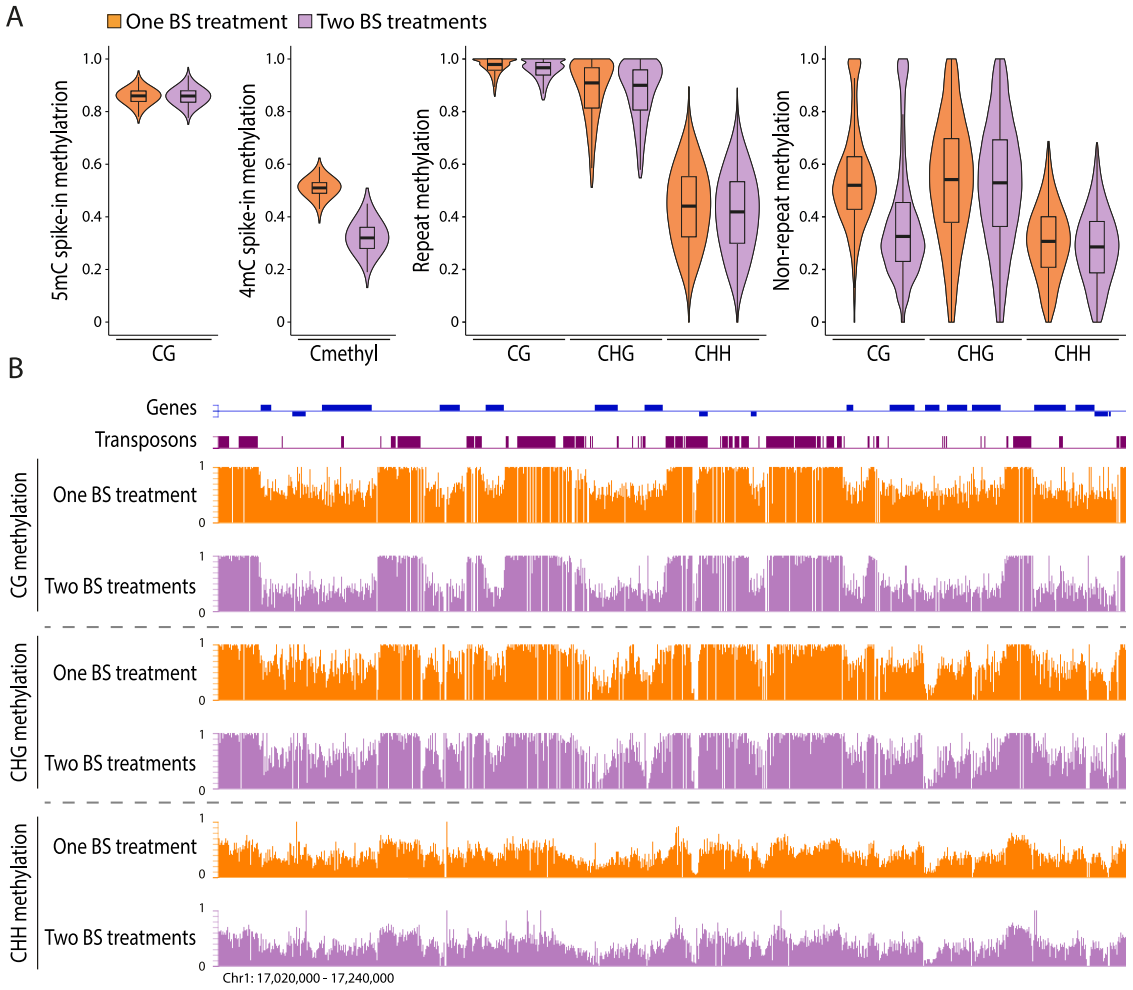


Figure S4. CG methylation at non-repetitive regions is sensitive to bisulfite treatment, related to Figure 4

(A) Violin plots showing DNA methylation detected from BS-seq with one or two rounds of BS treatment of *Marchantia* sperm, with a 5mC spike-in control (CG methylated Lambda genomic DNA) and a 4mC spike-in control (pUC19 PCR product synthesized with 4mdCTP). All cytosine contexts (Cmethyl) are examined for the 4mC spike-in control. Methylation in *Marchantia* sperm is separated into different sequence contexts and genomic regions (repeats and non-repeats).
(B) Snapshots of methylation levels detected by one or two BS treatments in different cytosine contexts over repeats and genic regions.

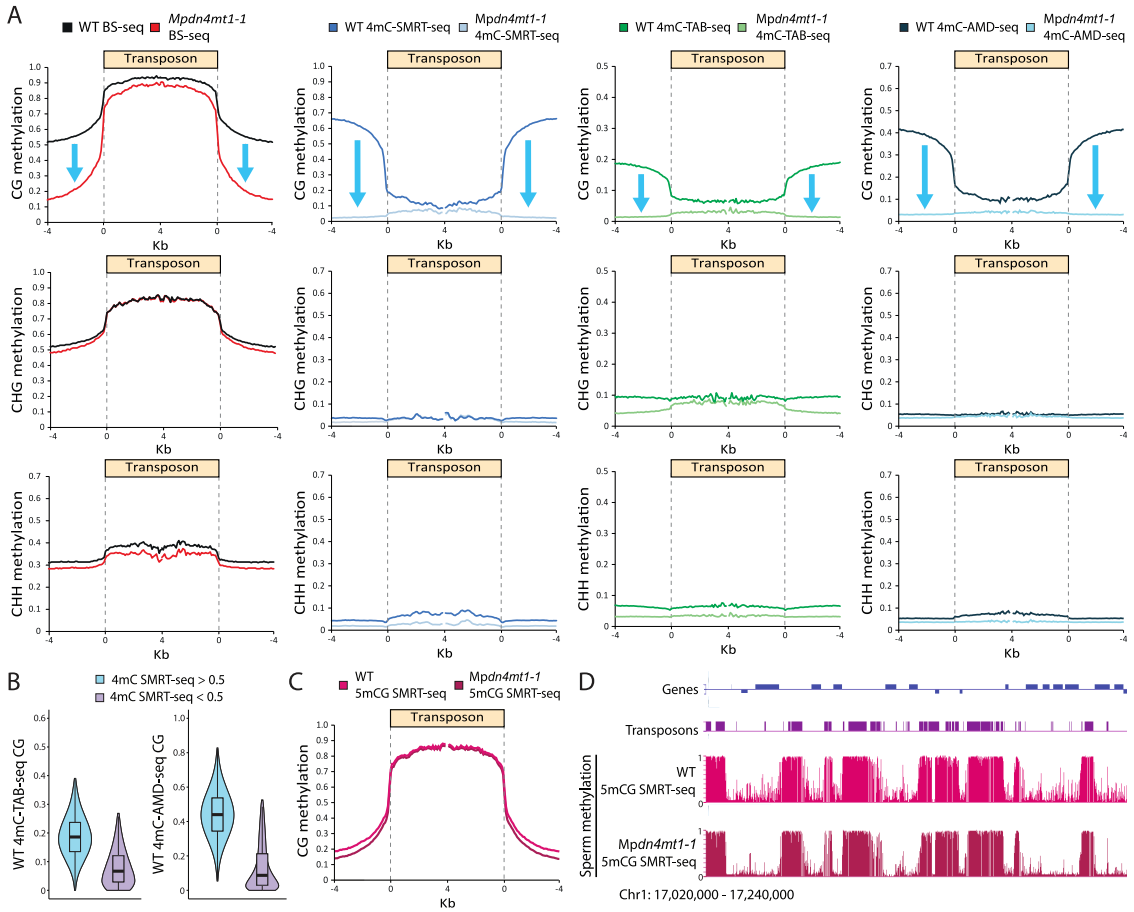


Figure S5. Cytosine methylation in WT and Mp*dn4mt1-1* mutant sperm measured by BS-seq, SMRT-seq, 4mC-TAB-seq, and 4mC-AMD-seq, related to Figure 4

(A) Transposons were aligned at the 5′ and 3′ ends (dashed lines), and average methylation levels for each 100-bp interval are plotted. The dashed line at zero represents the point of alignment. Blue arrows indicate loss of CG methylation outside of transposons in Mp*dn4mt1-1* mutant sperm.

(B) Cytosine methylation in WT sperm measured by 4mC-TAB-seq and 4mC-AMD-seq in comparison to SMRT-seq 4mC calls. Cytosines were separated by SMRT-seq 4mC fractional methylation greater than or less than 0.5 (the latter including absent 4mC calls).

(C) 5mCG methylation in WT and Mp*dn4mt1-1* mutant sperm measured by SMRT-seq, as in (A).

(D) Snapshot as in Figure 4C showing 5mCG methylation called by SMRT-seq.

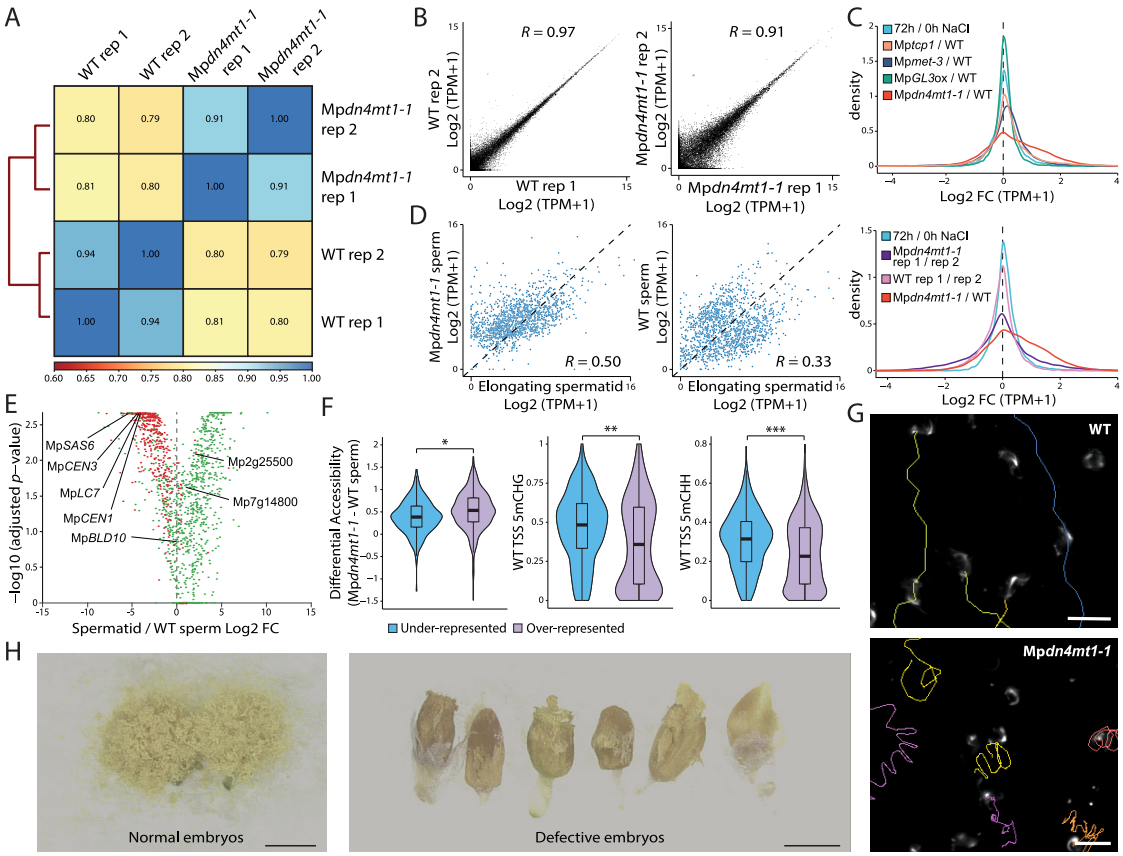


Figure S6. Functional and transcriptomic characterization of Mp*dn4mt1-1* mutant sperm, related to Figures 5 and 6

(A) Heatmap showing the Spearman correlation coefficient of pairwise comparison across the entire genome from each of the sperm ATAC-seq biological replicates (rep).

(B) Scatterplots showing high correlations between RNA-seq biological replicates of WT or Mp*dn4mt1-1* mutant sperm. *R*, Pearson's correlation coefficient.

(C) Kernel density plots displaying distributions of differential gene expression for genes expressed (TPM > 0) in at least one out of the two datasets compared. Unlike thallus datasets from stress-response, mutants, or gene-overexpression (ox) compared with their respective controls,[78–81] Mp*dn4mt1-1* mutant sperm displayed broad transcriptional shifts (top), indicative of non-uniform global changes in gene expression. As a control for noise, differential expression between biological replicates of Mp*dn4mt1-1* mutant and WT sperm is shown (bottom), both displaying central peaks at zero. This confirms that the broader distribution in the mutant versus WT sperm reflects genuine global transcriptional changes, not experimental noise.

(D) Transcript abundance in indicated samples for genes with significantly different transcript abundance between Mp*dn4mt1-1* sperm and WT sperm. Transcript abundance is shown in $\log_2$ (TPM + 1). Pearson's *R* values are displayed, as in (B).

(E) Volcano plot showing the difference in expression between elongating spermatids and WT sperm for transcripts that are over-represented (green) or under-represented (red) in Mp*dn4mt1-1* sperm compared with WT sperm (highlighted in Figure 5C). Vertical dashed line marks 0-fold change (FC).

(F) Violin plots displaying differential chromatin accessibility (measured by ATAC-seq) between Mp*dn4mt1-1* sperm and WT sperm (left) and DNA methylation in WT sperm at 200 bp downstream of TSS (5mCHG, middle; 5mCHH, right) at genes that are under- or over-represented in Mp*dn4mt1-1* sperm compared with WT sperm. *p* values (Kolmogorov-Smirnov test) are shown: $*p < 2.16e{-}10$, $**p < 7.33e{-}15$, $***p < 1.27e{-}12$.

(G) Snapshots of Data S1A and S1B showing the swimming paths taken by example WT or Mp*dn4mt1-1* sperm (indicated by colored lines). Scale bars, 20 μm.

(H) Examples of two normally burst embryos (the sporophytes) and examples of defective embryos that failed to release spores. Scale bars, 1 mm.