



## OPEN ACCESS

## EDITED BY

Daoquan Xiang,  
National Research Council Canada (NRC),  
Canada

## REVIEWED BY

Jingpu Song,  
National Research Council Canada (NRC),  
Canada  
Wolfgang Lukowitz,  
University of Georgia, United States  
Zhaodong Hao,  
Nanjing Forestry University, China

## \*CORRESPONDENCE

Yeonhee Choi  
✉ yhc@snu.ac.kr

## †PRESENT ADDRESSES

Jennifer M. Frost,  
Genomics and Child Health, Queen Mary  
University of London, London,  
United Kingdom  
Ping-Hung Hsieh,  
DOE Joint Genome Institute, Lawrence  
Berkeley National Laboratory, Berkeley, CA,  
United States

†These authors have contributed equally to  
this work

RECEIVED 13 April 2023

ACCEPTED 23 May 2023

PUBLISHED 08 June 2023

## CITATION

Lee J, Lee S, Park K, Shin S-Y, Frost JM,  
Hsieh P-H, Shin C, Fischer RL, Hsieh T-F  
and Choi Y (2023) Distinct regulatory  
pathways contribute to dynamic  
CHH methylation patterns in  
transposable elements throughout  
*Arabidopsis* embryogenesis.  
*Front. Plant Sci.* 14:1204279.  
doi: 10.3389/fpls.2023.1204279

## COPYRIGHT

© 2023 Lee, Lee, Park, Shin, Frost, Hsieh,  
Shin, Fischer, Hsieh and Choi. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that  
the original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Distinct regulatory pathways contribute to dynamic CHH methylation patterns in transposable elements throughout *Arabidopsis* embryogenesis

Jaehoon Lee<sup>1,2†</sup>, Seunga Lee<sup>1,2†</sup>, Kyunghyuk Park<sup>1</sup>, Sang-Yoon Shin<sup>2,3</sup>, Jennifer M. Frost<sup>4†</sup>, Ping-Hung Hsieh<sup>4†</sup>, Chanseok Shin<sup>2,3</sup>, Robert L. Fischer<sup>4</sup>, Tzung-Fu Hsieh<sup>5,6</sup> and Yeonhee Choi<sup>1,2\*</sup>

<sup>1</sup>Department of Biological Sciences, Seoul National University, Seoul, Republic of Korea, <sup>2</sup>Research Center for Plant Plasticity, Seoul National University, Seoul, Republic of Korea, <sup>3</sup>Department of Agricultural Biotechnology, Seoul National University, Seoul, Republic of Korea, <sup>4</sup>Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, United States, <sup>5</sup>Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC, United States, <sup>6</sup>Plants for Human Health Institute, North Carolina State University, Kannapolis, NC, United States

CHH methylation (mCHH) increases gradually during embryogenesis across dicotyledonous plants, indicating conserved mechanisms of targeting and conferral. Although it is suggested that methylation increase during embryogenesis enhances transposable element silencing, the detailed epigenetic pathways underlying this process remain unclear. In *Arabidopsis*, mCHH is regulated by both small RNA-dependent DNA methylation (RdDM) and RNA-independent Chromomethylase 2 (CMT2) pathways. Here, we conducted DNA methylome profiling at five stages of *Arabidopsis* embryogenesis, and classified mCHH regions into groups based on their dependency on different methylation pathways. Our analysis revealed that the gradual increase in mCHH in embryos coincided with the expansion of small RNA expression and regional mCHH spreading to nearby sites at numerous loci. We identified distinct methylation dynamics in different groups of mCHH targets, which vary according to transposon length, location, and cytosine frequency. Finally, we highlight the characteristics of transposable element loci that are targeted by different mCHH machinery, showing that short, heterochromatic TEs with lower mCHG levels are enriched in loci that switch from CMT2 regulation in leaves, to RdDM regulation during embryogenesis. Our findings highlight the interplay between the length, location, and cytosine frequency of transposons and the mCHH machinery in modulating mCHH dynamics during embryogenesis.

## KEYWORDS

DNA methylation, CHH methylation, transposable elements, CMT2, RdDM, embryogenesis

## Introduction

DNA methylation is crucial for regulating gene transcription, transposon silencing, and genomic imprinting in eukaryotic genomes. It also plays a critical role in development, and aberrant DNA methylation results in abnormal gene expression and developmental defects in vertebrates and plants (Deniz et al., 2019; Greenberg and Bourchis, 2019). In addition to CG methylation, plant DNA also contains methylated cytosines in CHG and CHH contexts (Law and Jacobsen, 2010). Therefore, plants have evolved a comprehensive DNA methylation system involving multiple associated enzymes. Small RNA-directed DNA methylation (RdDM) is a representative feature of this system (Matzke and Mosher, 2014; Matzke et al., 2015). The RdDM pathway recruits DOMAINS REARRANGED METHYLTRANSFERASE 1/2 (DRM1/2) to their target sites, where *de novo* methylation occurs at all cytosine contexts (Cao and Jacobsen, 2002; Stroud et al., 2013). Symmetric CG and CHG methylation can be maintained by DNA METHYLTRANSFERASE 1 (MET1) and CHROMOMETHYLASE 3 (CMT3), respectively, on the opposite strand independently of small RNA (sRNA) once *de novo* methylation has occurred (Law and Jacobsen, 2010). For CHH methylation, in addition to DRM1/2, CMT2 also methylates CHH sites, mainly in the body of long transposable elements (TEs), independently of sRNA molecules (Zemach et al., 2013).

Transposable elements (TEs), which are mobile genetic entities, make up a significant portion of the genome across different species that vary from 12% of nematode *Caenorhabditis elegans* genome up to 85% of the maize genome, for example (Jurka et al., 2011; Chenais et al., 2012). TEs are considered as insertional mutagens and affect host genome size through amplifying TE copy number, thereby serving a major driving force of genome evolution (Lanciano and Cristofari, 2020). Although some TE insertions can lead to adaptive effects to the host genome by relocating *cis*-acting elements that influence gene expression, most TE insertions are neutral or mildly detrimental to the host (Deniz et al., 2019). TEs are classified into two large groups depending on their mechanism of replication; retrotransposons (class I) which transpose *via* an RNA intermediate, and DNA transposons (class II). Retrotransposons are subdivided depending on the presence or lacking of long terminal repeat (LTR). Gypsy and Copia superfamilies belong to LTR retrotransposons whereas LINEs and SINEs are non-LTR retrotransposons. Major *Arabidopsis* DNA transposons (Tn) are subdivided into two groups depending on the presence or lacking of terminal inverted repeats (TIR). TIR Tns harbor TIR at the borders such as MuDR and En-Spm superfamily, and the non-TIR Helitron Tns are known to replicate *via* a rolling circle mechanism (Quesneville, 2020).

While DNA methylation patterns are replicated faithfully during cell division, preserving genome stability and cell lineage, they also need to be dynamically reprogrammed during development to establish new transcriptional states and cell fates. The dynamic regulation of DNA methylation is essential for animal development and reproduction. In mammals, DNA methylation is reprogrammed in each generation to erase and re-establish parental imprints for the next generation (Seki et al., 2005; Sasaki and

Matsui, 2008; Seisenberger et al., 2013; Heard and Martienssen, 2014). In angiosperms, active demethylation occurs before fertilization in central and vegetative cells, adjacent companion cells of the egg and sperm, respectively (Schoft et al., 2011; Ibarra et al., 2012; Park et al., 2016; Kim et al., 2019). In *Arabidopsis*, this pre-fertilization epigenetic reconfiguration is performed by DEMETER (DME), a novel and essential DNA glycosylase that initiates the excision and removal of methylated cytosines *via* the base excision repair (BER) pathway. This process affects DNA methylation and gene transcription profiles in the endosperm (Gehring et al., 2006; Ibarra et al., 2012). Consequently, genome-wide DNA hypomethylation and gene imprinting are primarily observed in the endosperm upon fertilization in *Arabidopsis* (Gehring and Satyaki, 2017; Batista and Kohler, 2020).

While the presence of a bona fide DME ortholog in monocots remains debated, the ROS1a protein has been shown to be the functional counterpart of DME in the rice cultivar Nipponbare (Ono et al., 2012). The rice vegetative cell (VC) genome is also extensively hypomethylated compared to sperm, and this VC hypomethylation is dependent on ROS1a (Kim et al., 2019). Therefore, gamete companion cell epigenetic remodeling by DNA glycosylases is an evolutionarily conserved phenomenon in rice and *Arabidopsis*, despite their divergence over 150 million years ago. Multiple genome-wide DNA methylation studies have confirmed that the endosperm genome is hypomethylated relative to that of embryo in various plant species, including *Arabidopsis*, rice, maize, and Castro bean (Lauria et al., 2004; Gehring et al., 2009; Hsieh et al., 2009; Zemach et al., 2010; Xu et al., 2016; Gent et al., 2022; Xu et al., 2022).

Compared to the extensive epigenetic reprogramming observed in gamete companion cells and endosperm, the embryo methylome, at least in the context of CG methylation, appears to remain static during development and between generations, ensuring robust transgenerational epigenetic inheritance (Hsieh et al., 2016; Bewick and Schmitz, 2017; Picard and Gehring, 2017). In contrast, dynamic and extensive post-fertilization reconfiguration of CHH methylation has been reported during embryogenesis and germination processes (Kawakatsu et al., 2016; Bouyer et al., 2017; Lin et al., 2017). This gradual increase in CHH methylation during embryogenesis is associated with bursts of sRNA accumulation during embryo development and an increase in RdDM activity, particularly during embryo maturation (Papareddy et al., 2020). One function of this increase in CHH methylation is thought to help suppress the activity of TEs and prevent their transposition, maintaining genomic integrity (Lin et al., 2017).

Although DNA methylation changes during plant embryo development have been reported in multiple plant species (Xing et al., 2015; Bouyer et al., 2017; Kawakatsu et al., 2017; Lin et al., 2017; Grover et al., 2020; Rajkumar et al., 2020), a comprehensive understanding of methylation dynamics for different TE groups or families during embryo development has not been reported. Recently, we described methods for isolating high-purity stage-specific embryos for low-input methylome library construction and data analysis (Yoo et al., 2021). In this study, we used these methods to investigate comprehensive methylome changes in developing *Arabidopsis* embryos at five stages. Our findings revealed distinct

methylation dynamics for different TE classes according to their length, superfamily, genome location, and cytosine ratio. Shorter, heterochromatic TEs in embryos are prone to be newly targeted by the RdDM pathway upon fertilization. Furthermore, we discovered that many CHH-methylated regions expand into neighboring unmethylated regions during embryo maturation, accompanied by the accumulation and expansion of 24nt sRNA clusters. Overall, our study provides a detailed analysis of the dynamic changes in TE methylomes during *Arabidopsis* embryo development, shedding light on the underlying mechanisms of TE regulation and their potential roles in embryo development.

## Materials and methods

### Plant materials and growth conditions

The *Arabidopsis thaliana*, Col-*gl* (Columbia-glabrous) ecotype was used for embryo isolation. Plants were cultivated in a controlled environment room with a long photoperiod of 16 hours of light and 8 hours of darkness, at a temperature of 22°C with cool white fluorescent light (100 mmole/m<sup>2</sup>/s).

### Embryo isolation and construction of whole-genome bisulfite sequencing

After 24 hours of emasculating, fully matured WT ovules were pollinated with WT pollen. The pollinated plants were then incubated in the growth room until they reached a suitable stage for sampling. *Arabidopsis* embryos at DAP4, 5, 7, 9, and 12, corresponding to the globular, heart, torpedo, bending torpedo, and mature green stage embryos, respectively, were dissected from seeds using various methods specific to each stage. Globular embryos were obtained by gentle grinding the seeds with a pestle, while heart embryos were excised by making incision of the seeds with tweezers. For later-stage embryos, from torpedo to mature green, embryos were carefully extracted by puncturing the seed coat and endosperm with tweezers, followed by pushing embryos out. Each released embryo was manually collected using microcapillaries, one by one. All procedures used in this study, including pollination for embryo preparation, sampling of the embryos, genomic DNA extraction, bisulfite library construction, and the processing of whole-genome bisulfite sequencing (WGBS) data, were previously described in detail (Yoo et al., 2021).

### Global DNA methylation analysis

Embryo DNA methylation data was divided into 50 bp fractional windows that contained at least three cytosines with at least five aligned reads in all stages for each cytosine context. For the

circular global map, we used 5kb windows that contained more than ten cytosines with at least five reads for each methylation context.

### Small RNA-Seq analysis

The small RNA-Seq dataset was obtained from the NCBI GEO (GSE98553, GSE132066, GSE152971) (Lutzmayer et al., 2017; Plotnikova et al., 2019; Papareddy et al., 2020). We performed adapter trimming, quality control, and 18-to-26 nt-length read selection using Trimmomatic (0.39) and Cutadapt (v3.4). The pre-processed reads were mapped to rRNA/tRNA/snRNA sequences (RNACentral, v17) using bowtie (v1.3.0, with parameters -v 1 -m 0 -a) to filter out structural non-coding RNA reads. The remaining reads were first aligned using bowtie (1.3.0, -v 0 -m 0 -a), and then the aligned reads were re-aligned with ShortStack (v3.8.5, with parameters -align\_only -mismatches 0 -mmap u -bowtie\_m 1000 -ranmax 3) using the TAIR10 genome sequence as a reference. Reads aligned to known microRNA precursors were removed using Bedtools intersect (v2.30.0). The microRNA-free reads were then used as input for downstream analysis. Small RNA read clustering was conducted in a sample-by-sample manner using ShortStack (with parameters -dicermis 20 -dicermis 24 -foldsize 300 -pad 1(or 75), -mincov 1.0rpm -nohp) (Johnson et al., 2016).

### Average length of merged windows for the methylated block

To calculate the average length of the regions where consecutive methylated probes were present at each stage of embryo development, we used 10 bp windows with at least one methylation site and at least five aligned reads. We merged windows with methylation levels above 5%, 10%, or 15% in all stages for each cytosine context if they were closer than 40 bp apart. The resulting merged windows were used to determine the average length of the methylated regions at each stage.

### Relationship between sRNA clustering and spreading of DNA methylation

To identify regions with consecutive CHH methylation and sRNA clustering simultaneously in all stages of embryo development, we selected regions with consecutive CHH methylated probes in at least three out of five stages and a methylation level cutoff of 15%, along with sRNA clusters common in at least five out of eight stages. We divided the consecutive CHH methylated regions into two groups based on whether they overlapped with sRNA clustering or not. The average length of the regions with or without sRNA clustering was then calculated.

## Classification of TE by CMT2 or RdDM dependency

To determine the dependency of TE methylation on CMT2 and RdDM pathways, we utilized previously identified lists of differentially methylated regions (DMRs) (Stroud et al., 2013), and recently derived CHH methylation pathway DMR classifications (Hsieh et al., 2023): DMRs that are regulated by the CMT2 or RdDM pathways in both leaf and embryos are classified as cCMT2 DMRs or cRdDM DMRs, respectively. DMRs that are regulated by CMT2 or unmethylated in the leaf and become newly targeted by the RdDM pathway in embryo are classified as eRdDM DMRs. Here, TEs were allocated into the three DMR classes using bedtools intersect -wa -a TE list -b DMR list. TEs allocated to more than one DMR class were removed, and only TEs associated with a single type of DMR class chosen for further analyses.

## Results

### DNA methylation dynamics during *Arabidopsis* embryo development

We utilized our recently reported optimized method to profile the genome-wide *Arabidopsis* embryo methylome at five developmental stages: DAP4, 5, 7, 9, and 12, which correspond to the globular, heart, torpedo, bending torpedo, and mature green stage embryos, respectively (Figure 1A) (Yoo et al., 2021). All libraries were made up of at least two biological replicates with robust CT conversion (>99%) and genome coverage (>40X), complying with the recommended standard for methylation data analysis (Ziller et al., 2015) (Figure 1B).

As previously reported (Cokus et al., 2008; Lister et al., 2008), *Arabidopsis* chromosomes exhibited high levels of pericentromeric methylation in all cytosine contexts, due to the accumulation of repetitive sequences and heterochromatic TEs (Figure 1B). Consistent with previous reports (Bouyer et al., 2017; Kawakatsu et al., 2017; Lin et al., 2017; Papareddy et al., 2020), CG methylation levels remained relatively constant throughout embryo development, whereas non-CG methylation was more dynamic (Figure 1C). In particular, as others have reported, there was a gradual increase in global CHH methylation levels in pericentromeric regions during embryo development (Figure 1B), which culminated as the seeds mature (Figure 1C). The increase in CHH methylation levels primarily resulted from increases at TEs, whereas gene body methylation levels in all cytosine contexts remained relatively stable (Figure 1C). CHH methylation levels in gene bodies increased only slightly (Figure 1C). CHG methylation levels in TEs showed a slight decrease, particularly from DAP7 to DAP9 (Figure 1C), but otherwise remained relatively stable as the seeds mature.

We next analyzed the methylation dynamics of *Arabidopsis* embryo development within genomic features, focusing on genes and TEs (Figure 1). As observed genome-wide, we found that the methylation change from the beginning to end of embryogenesis was small in genic regions in all contexts (<0.7%; Figures 1D–F).

However, we observed distinct gene body methylation levels for both CHG and CHH contexts at each developmental stage (Figures 1E, F), indicating a coordinated change in methylome status during development. For CHH methylation at the 5' and 3' ends of genes, there was a slightly larger increase (1%) that was distinguishable from the globular-torpedo to the bending torpedo-mature green stages, likely due to the presence of overlapping TE sequences in the flanking regions of genes (Figure 1F). As noted above, the torpedo embryo stage at DAP7 showed the lowest CHG methylation levels (Figures 1C–E).

At TEs, CG methylation increased marginally throughout embryogenesis (Figures 1G–I). CHH methylation levels at TEs, as previously noted, demonstrated a profound and continuous increase during development (Figure 1I). Interestingly, a striking increase was shown in CHH methylation at TEs from DAP5 to DAP7, which coincided with a decrease in CHG methylation levels (Figures 1C–H). This observation may indicate that CHG and CHH methylation work together in a coordinated manner to repress TEs during heart (H) to torpedo (T) transition phase, which marks the beginning of maturation after pattern formation.

### CHH methylation spreading is accompanied by sRNA cluster expansion during embryo development

In mammalian embryogenesis, a unique form of *de novo* CG methylation occurs through the spreading of methylation from cis-regulatory regions to distal CpG sites (Turker, 2002). To investigate whether this phenomenon of methylation spreading also occurs in plants during embryo development, we calculated the average length of the region where CHH methylated probes exist consecutively at each embryo developmental stage. We merged windows with methylation levels above 5%, 10%, or 15% in all stages for each cytosine context if they were closer than 40 bp apart. As the embryos developed, the average length of consecutively methylated regions gradually increased, especially when we used higher methylation cutoffs (15%; Figure 2A). This result suggests that genome-wide CHH methylation spreads as the embryo matures.

The RdDM pathway plays a critical role in *de novo* methylation in *Arabidopsis* (Matzke and Mosher, 2014; Chow et al., 2020; Erdmann and Picard, 2020), however, it is possible that sRNAs are only required for the initial CHH methylation setting during early embryogenesis. This initial CHH methylation could then act as a signal for its spreading to nearby sites at later stages of embryogenesis, similarly to CG methylation in mammals. In this scenario, CHH methylation spreading would occur even in the absence of matching sRNAs to nearby sites. Alternatively, sRNA clusters corresponding to neighboring regions may be continuously required for the spreading of CHH methylation. We measured the average length of 24nt sRNA clusters that are merged when they exist within 75 bp at each stage using publicly available sRNA expression data during embryogenesis (Papareddy et al., 2020). As the embryos developed, the average length of merged sRNA clusters for 24nt sRNAs gradually increased (Figure 2B). This suggests that



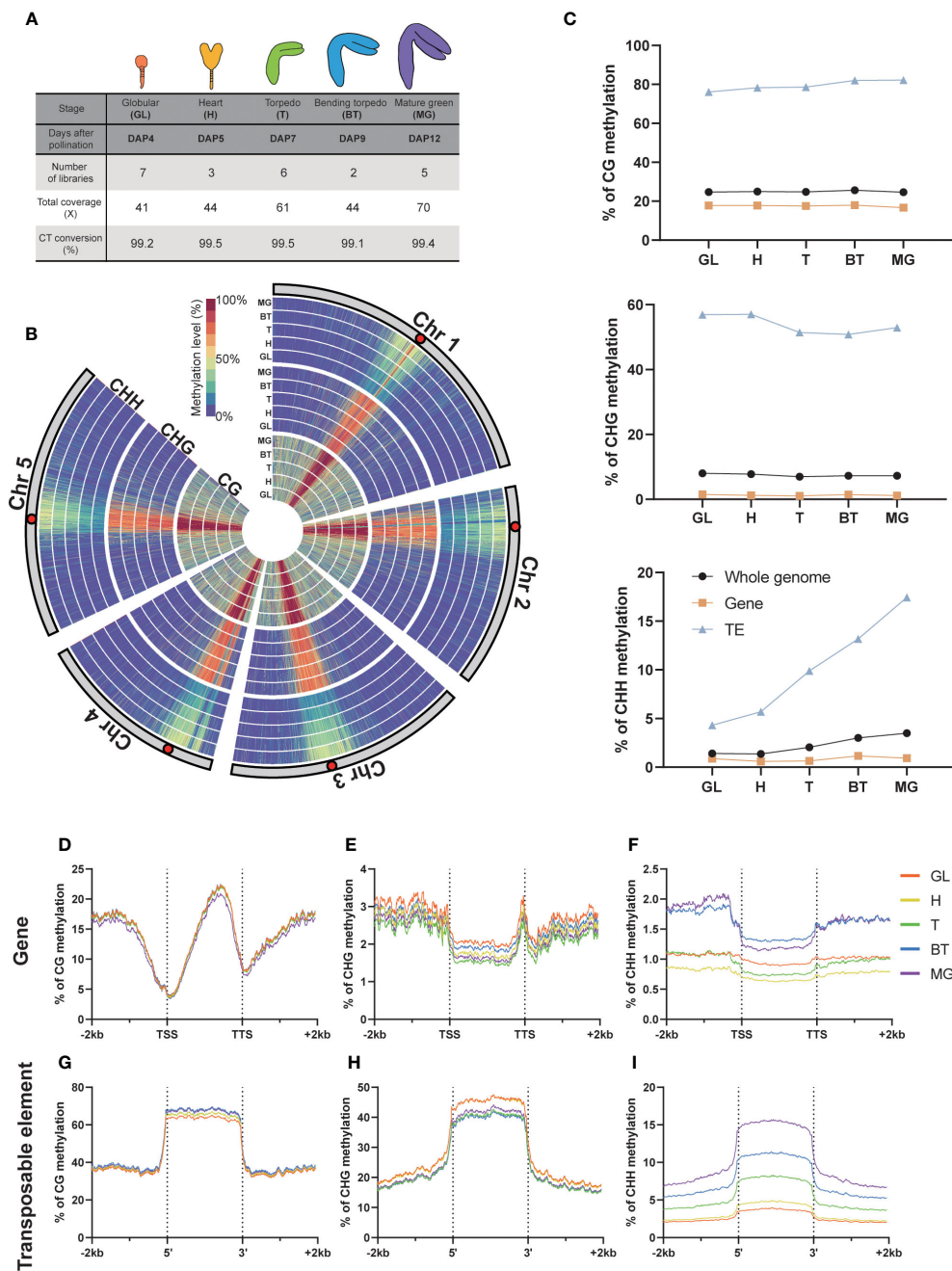


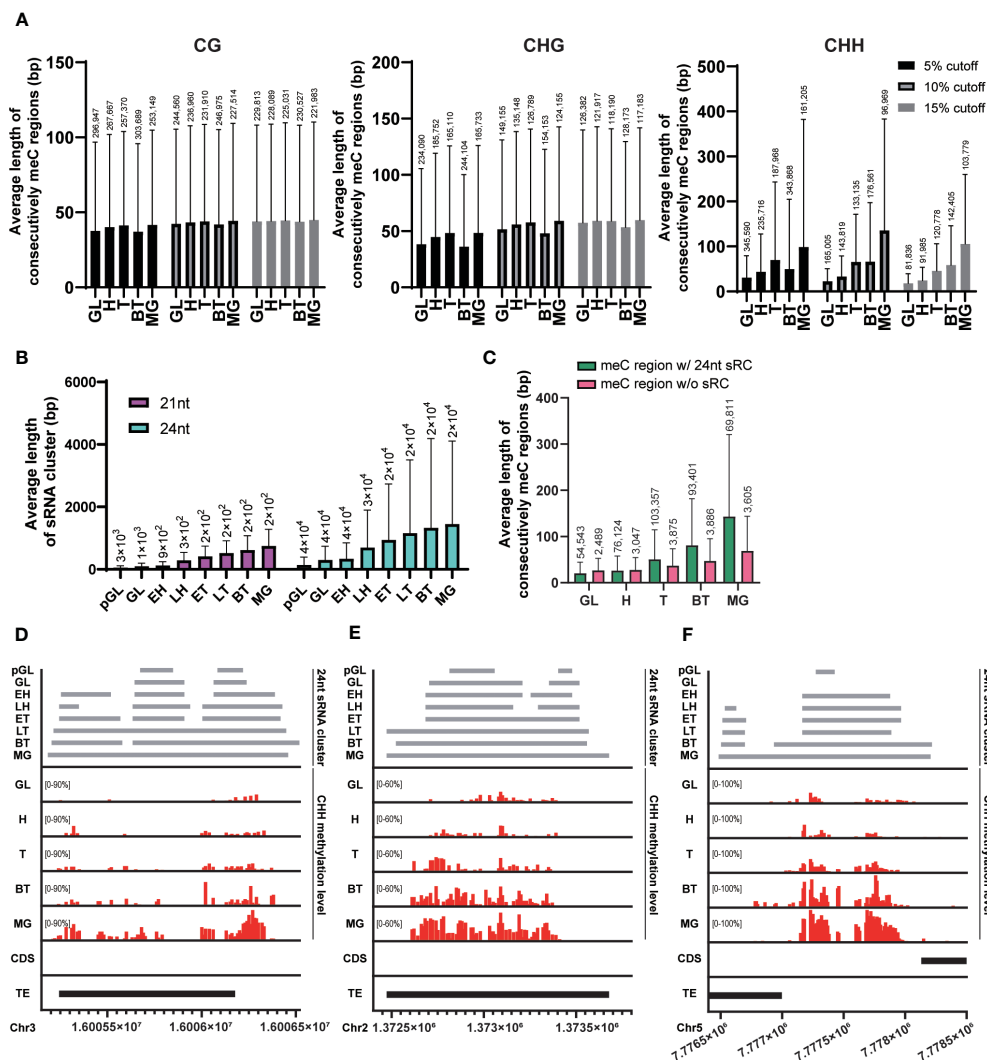
FIGURE 1

DNA methylation in global and genomic features. (A) Whole Genome Bisulfite Sequencing (WGBS) libraries used in this study. The percentage of C-to-T conversion was calculated by measuring 100% minus the methylation level of plastids that are known to be unmethylated. DAP, days after pollination. (B) Circular heatmap displaying the global DNA methylation levels during embryo development, with the inner circle representing the earlier developmental stages. GL, globular; H, heart; T, torpedo; BT, bending torpedo; MG, mature green. (C) The average DNA methylation levels by the different genomic features. (D-I) The DNA methylation levels of genomic features and their surrounding regions are shown.

expression of sRNAs corresponding to nearby sites is still required for the methylation of neighboring unmethylated CHH sites during embryo development.

To investigate the relationship between the expansion of sRNA expression and the spreading of CHH methylation during embryo development, we compared the average length of regions where consecutive CHH-methylated windows overlapped with merged 24nt sRNA clusters. Our results revealed a linear increase in the

length of the sRNA clusters during development in CHH overlapping regions, demonstrating that there is a positive correlation between the spreading of CHH methylation and expansion of sRNA-producing regions (Figure 2C, green bars). CHH methylated regions were also expanded in the absence of sRNA clusters as embryos matured (Figure 2C, pink bars). However, these regions were formed of far fewer loci and reflected less extensive CHH spreading. In addition, when we



**FIGURE 2** Spreading of CHH methylation and expansion of sRNA expression. (A) The average length of regions where the methylated probes exist consecutively in each cytosine context during embryo developmental stage. (B) The average length of merged sRNA clusters in each embryo developmental stage. (C) The average length of regions where consecutively CHH-methylated overlapped or did not overlap with merged 21nt or 24nt sRNA clusters (sRC) during each embryo developmental stage. (A–C) The height of box and whisker indicate mean and standard deviation, respectively. The number of loci (for the methylated regions and sRNA clusters) are indicated on each whisker. (D–F) Example regions illustrating the spread of CHH methylation. GL, globular; H, heart; T, torped; BT, bending torpedo; MG, mature green.

mapped the position of these non-sRNA overlapping CHH methylated regions, we found that they were tightly centromeric, and as such likely to be conferred by CMT2 (Supplementary Figure S1). To further investigate the positive correlation of CHH methylation spreading with sRNA clusters, we visualized individual overlapping loci, and observed that expanding sRNA expression was coincident with CHH methylation spreading and accumulation (Figures 2D–F), providing support for previous observations that increased embryo CHH methylation is conferred by the RdDM pathway (Hsieh et al., 2023).

Taken together, our data suggest that the RdDM pathway plays an important role in both initiating and spreading CHH methylation to nearby regions, as well as in the accumulation of CHH methylation levels. This can be achieved, at least in part, by the continuous expression of sRNA clusters followed by expression

of sRNA clusters in neighboring regions, resulting in wider CHH methylation regions and higher CHH methylation levels at later stages of embryogenesis.

### Chromosomal location shapes distinctive methylation dynamics in CMT2 and RdDM DMRs

CHH methylation is known to be regulated by two pathways; the RNA-independent CMT2 pathway and the RdDM pathway (Zemach et al., 2013; Matzke and Mosher, 2014; Kenchanmane Raju et al., 2019). RdDM targets are methylated by *de novo* methyltransferases, DRM1 and 2. Recent papers described the delineation of embryo CHH loci into classes based on their

mechanism of catalysis using *drm1/2* and *cmt2* mutants (Kawakatsu et al., 2017; Hsieh et al., 2023). Briefly, if an embryonic CHH DMR is targeted by CMT2 both in the vegetative (leaves) and reproductive embryo stages, that DMR is denoted as a conserved CMT2 (cCMT2) DMR. If a DMR is targeted by DRM1/2 both in leaves and embryos, that DMR is denoted as a conserved RdDM (cRdDM) DMR. If a DMR is targeted by CMT2 or unmethylated in leaves but becomes exclusively targeted by DRM1/2 in embryos, that DMR is denoted an embryonic RdDM (eRdDM) DMR (Hsieh et al., 2023).

In order to understand the methylation dynamics of these three DMR classes during development, we divided our data according to DMR class and examined their average methylation level during embryogenesis (Figure 3A, also see Methods). Average CG methylation levels of cCMT2 DMRs increased from DAP4 to DAP12 ( $p < 0.0001$ ), whereas those of cRdDM DMRs decreased during the same period ( $p < 0.0001$ ) (Figure 3A). Both the levels and pattern of CG methylation in eRdDM DMRs laid between those of cCMT2 DMRs and cRdDM DMRs (Figure 3A). Average CHG methylation level change was unique, whereby cCMT2 and eRdDM DMRs showed decreases from DAP5 to DAP7 ( $p < 0.0001$  for cCMT2 DMRs and  $p < 0.0001$  for eRdDM DMRs) whereas cRdDM DMRs showed large increases in CHG methylation from DAP7 to DAP12 ( $p < 0.0001$ ) (Figure 3A). CHH methylation levels were continuously increased during embryogenesis in all classes (Figure 3A). The most dynamic epigenetic change observed during embryogenesis remained the constant increase in CHH methylation, and among them, cRdDM-targets displayed the largest change. This was particularly striking from DAP9 to DAP12, reaching 60% CHH methylation levels at DAP12 (Figure 3A).

To explore the overall location of CMT2/RdDM-dependent DMRs, we mapped cCMT2-, cRdDM-, and eRdDM DMRs to chromosomes. All three classes were enriched in pericentric chromatin (Figure 3B). However, consistent with previous reports, cCMT2 DMRs were almost entirely located in pericentromeric regions (He et al., 2021) (Figure 3B, green trace), whereas cRdDM DMRs were much more evenly spread throughout chromosome arms (Figure 3B, pink trace). Interestingly, eRdDM DMRs were most similar in profile to cCMT2 DMRs, with clear

pericentromeric enrichment, although they demonstrated comparative increases in chromosome arm enrichment (Figure 3B, black trace). These data indicate that chromosomal location, which is highly associated with local chromatin state, may be key to determining CHH enzyme dependency, consistent with previous reports (Sigman and Slotkin, 2016).

Recently, it was reported that some regions regulated by CMT2 in wild-type *Arabidopsis* seedlings are accessed by the RdDM pathway in the absence of the DDM1 nucleosome remodeling factor, which in wild-type plants mediates heterochromatic DNA methylation in all cytosine contexts (Zemach et al., 2013; He et al., 2021). These regions were referred to as “CMT2-to-RdDM loci” by the authors. Some eRdDM DMRs exhibited similar traits to CMT2-to-RdDM loci, as they were initially methylated by CMT2 in somatic tissues and subsequently switch to the RdDM pathway after fertilization during embryonic development (Hsieh et al., 2023). This prompted us to investigate the extent of overlap between eRdDM DMRs and CMT2-to-RdDM loci by He et al. We conducted a parallel comparison of these regions and found that most of the DMRs were not shared between two datasets (Figure 3C), indicating that eRdDM DMRs are distinct from the CMT2-to-RdDM loci.

## Methylation levels of TE- flanking regions differ according to CHH regulatory mechanisms

The purpose of CHH methylation in *Arabidopsis* is widely accepted to be TE silencing, to prevent insertional mutagenesis in key tissues, such as the developing embryo. To explore TE dynamics during embryo development, we categorized TEs based on their dependency on CHH methylation pathway. As for our whole-genome analyses, we annotated TEs according to the three DMR classes outlined above (cCMT2, cRdDM and eRdDM) (Hsieh et al., 2023), and calculated the proportion of each (Figure 4A). Then, only those TEs annotated exclusively with a single category of DMR were used for further analyses; 3,790 for cCMT2 TEs, 5,402 for cRdDM TEs and 2,646 for eRdDM TEs (Figure 4A).

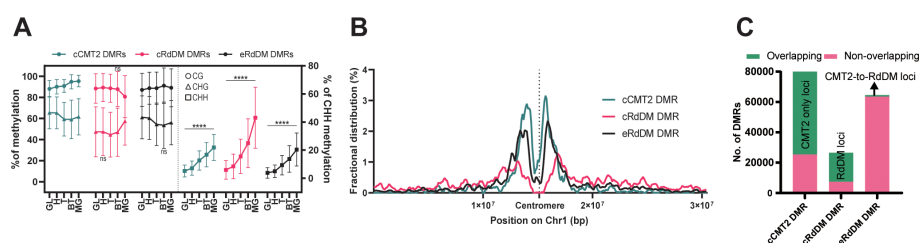
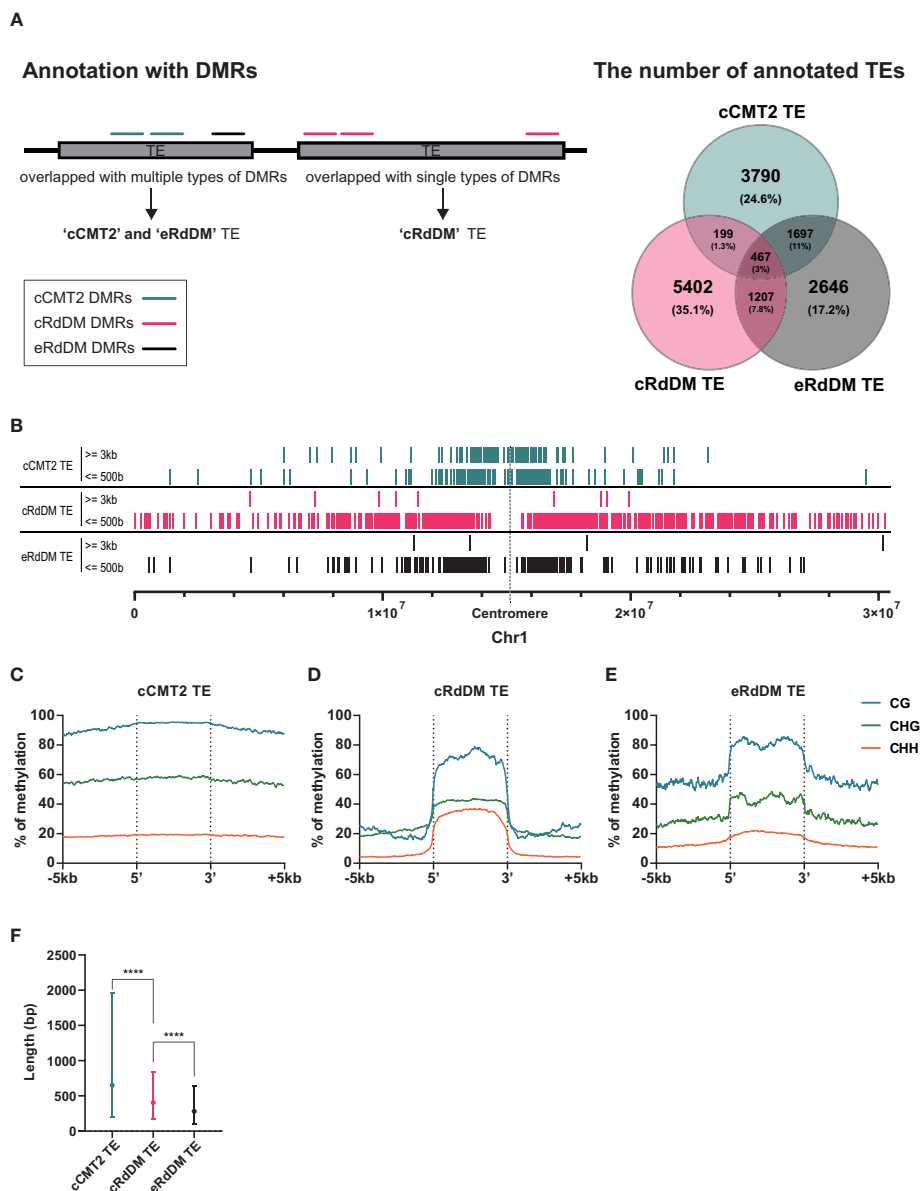


FIGURE 3

Characteristics of three types of DMRs. (A) The average DNA methylation levels of the three types of DMRs were measured at each stage. The levels of mCG and mCHG are shown on the left side, while the levels of mCHH were shown on the right side. The error bar represents the standard deviation. The  $p$ -value for each developmental stage was determined by comparing it to the prior stage. CHH methylation levels during embryo development were significantly changed in all stages ( $p < 0.0001$ ). Similarly, the  $p$ -values for mCG and mCHG also showed significant changes, except for the stages denoted as ‘ns’ (not significant). (B) Fractional distributions of the position of the three types of DMRs on chromosome1 were shown. (C) The three DMR classes used in this study were compared with the previously identified DMRs using a similar approach for the regions found in the seedling stage (He et al., 2021). The counts of DMRs, defined at the loci from (He et al., 2021), were colored in green, and those not found were colored in pink.



**FIGURE 4** Distinctive characteristics of TEs categorized into three types. **(A)** A diagram explaining how to categorize TEs is shown on the left side, and the classified TEs are compared on the right side. Only the exclusive parts were used for the following analyses. **(B)** The positions of classified TEs divided by length on chromosome 1 were analyzed. **(C-E)** DNA methylation levels of the three classified TEs and their surrounding regions were analyzed. **(F)** The length of the three classified TEs was compared. Dots and error bars indicate the average and standard deviation, respectively. \*\*\*\* indicates a significant difference ( $p < 0.0001$ ).

To investigate the overall location of TEs depending on their CHH mechanistic class, we mapped cCMT2, cRdDM, and eRdDM TEs to the chromosomes. Similar to the chromosomal location of each CHH DMR (Figure 3B), we found cCMT2 TEs were more highly concentrated in pericentromeric regions than other classes of TEs. Furthermore, we observed that longer cCMT2 TEs ( $\geq 3$ kb) were more highly concentrated in pericentromeric regions compared to shorter cCMT2 TEs ( $\leq 500$  bp) (Figure 4B). In contrast, cRdDM TEs were located throughout the chromosomes

and most of them were shorter than 500bp (Figure 4B). Interestingly, the majority of eRdDM TEs were also shorter than 500bp, similar to cRdDM TEs (Figure 4B), indicating that accessible length is a fundamental attribute of loci targeted by RdDM. Additionally, we observed that eRdDM TEs were predominantly located similarly to the shorter cCMT2 TEs, with a slight shift towards chromosomal arms (Figure 4B).

We compared the overall cytosine methylation levels for the three TE classes, mainly revealing differences in flanking regions.



Consecutive high levels of methylation in all contexts surrounding TE bodies was observed in cCMT2 TEs (Figure 4C), but not in cRdDM TEs (Figure 4D). This is consistent with previous observations of comprehensive heterochromatin formation and DNA methylation in pericentromeric regions, since cCMT2 TEs were almost all pericentromeric, and cRdDM TEs were much more spread along the chromosome (Figure 4B) (Zemach et al., 2013; Yelagandula et al., 2014; Zhao et al., 2019; Bourguet et al., 2021). eRdDM TE methylation levels were intermediate between the other two classes, displaying the same clear delineation of CG and CHG methylation levels between TE bodies and flanking regions as cRdDM sites, but much less so for CHH methylation (Figure 4E).

We next examined the average length of TEs classified into the three DMR classes. cCMT2 TEs tended to be longer than cRdDM TEs, although with large variation (Figure 4F,  $p < 0.0001$ ). eRdDM TEs were even shorter in size than cRdDM TEs (Figure 4F,  $p < 0.0001$ ). Considering that 29% of eRdDM TEs were cCMT2 TEs in leaves (Figure 4A, Supplementary Figure S2), this is quite intriguing. Based on their size and location, this result raises the idea that eRdDM TEs that originated as CMT2 TEs in leaves were shorter, less-heterochromatic CMT2 sites. We speculate that these sites transitioned to RdDM targets due to changes in local chromatin organization and the presence of highly robust RdDM activity during the developmental transition upon fertilization.

## TE size is a critical factor for CHH dependency even within the same TE superfamily

To test whether the composition of TE superfamilies within each TE class is correlated to CHH-conferring enzyme dependency, we investigated the proportion of representative TE superfamilies within each TE class. In *Arabidopsis*, retroelements represent the most common TE type, comprising ~10Mb of the genome, closely followed by DNA transposon orders Helitrons (8 Mb) and TIR elements (7 Mb) (Ahmed et al., 2011). TIR elements are so called due to their Terminal Inverted Repeats, and two of the most frequent TIR superfamilies in *Arabidopsis* are Mutator-like (MuDR) which is also

one of the most mobile elements (Quadrana et al., 2016), and En-Spm. Retroelements can be further delineated into LTR (long terminal repeat) TEs such as Gypsy and Copia [another highly mobile elements, (Quadrana et al., 2016)], and non-LTRs such as LINES and SINEs (Quesneville, 2020).

We examined whether any TE superfamilies were enriched in a particular TE class. As shown in Figures 5A, B, cCMT2 TEs were highly enriched for Gypsy LTR retrotransposons, which is known to be the longest TE in mean length and concentrated in pericentromeric regions (Quesneville, 2020). So most Gypsy elements acquire CHH methylation *via* the CMT2 pathway. In contrast, both cRdDM and eRdDM TEs were depleted for LTR targets, instead enriched for the Helitron and MuDR superfamilies of TEs, of which the mean lengths were significantly shorter than LTR retrotransposons (Figures 5A, B). All TE superfamilies, regardless of their average length, exhibited the same trend: longer TEs were targeted by the CMT2 pathway, while shorter TEs were targeted by the RdDM pathway, within each family (Figure 5B). For instance, SINEs were generally the shortest TE among the seven different superfamilies we examined. However, there were SINE TEs targeted by CMT2 which were, although long relative to other SINEs, were still shorter than RdDM-targeted TEs in the Helitron superfamily (Figure 5B). In addition, the mean length of eRdDM TEs were shorter than cRdDM TEs even within each superfamily (Figure 5B). This suggests a sequence based mechanism influences CHH methylation targeting, whereby once an element reaches a certain size, the sequence composition is such that CHH targeting switches from RdDM to CMT2 based mechanisms. Shorter TEs that are unmethylated, or even CMT2-regulated in leaves, may become accessible to the RdDM pathway after the transition to reproductive development in the embryo. Taken together, and consistent with a previous report (Zemach et al., 2013), our data highlight that TE size, relative to the superfamily, is one of the most important factors for CHH methylation targeting mechanisms.

All three DMR classes of CHH-methylated loci in embryo were found across all TE superfamilies (Figures 5A, B). Next we examined methylation dynamics of each TE family during embryo development. In general, cCMT2 TEs showed uniform

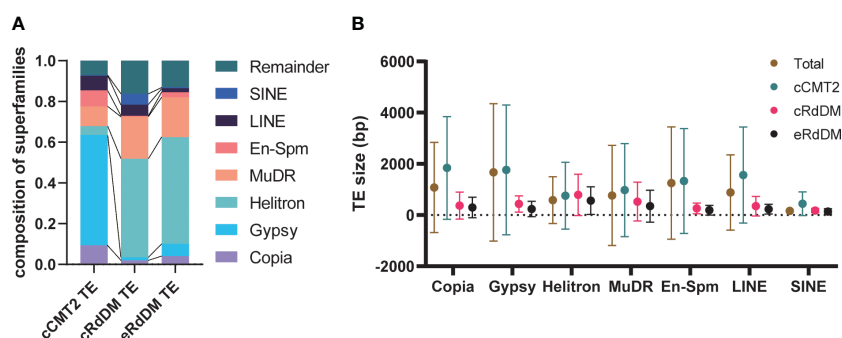


FIGURE 5

Characteristics of the three TE classes. (A) The composition of the seven superfamilies in three TE classes were analyzed. Remainder indicates the TEs that are not included in the seven superfamilies. (B) The average TE size according to superfamilies was analyzed. Dots and error bars indicate the average and standard deviation, respectively.

methyloome dynamics throughout the superfamilies (Supplementary Figure S3A), whereas both cRdDM- and eRdDM TEs showed more variable methyloome dynamics in each superfamily (Supplementary Figures S3B, C). Consistent with the result shown in Figure 3A, while all the TEs showed gradual increase of CHH methylation, cRdDM TEs showed the most striking increase in CHH methylation for all superfamilies (Supplementary Figure S3B, cRdDM\_CHH). CG methylation levels were decreased specifically at DAP12 (Figure 3A), an all seven superfamilies of cRdDM TEs followed a similar pattern (Supplementary Figure S3B, cRdDM\_CG). It is tempting to speculate that the decrease in CG methylation in cRdDM TEs located in chromosomal arms might induce an active demethylation mechanism, resulting in an increase in sRNA generation and CHH *de novo* methylation. The large increase in CHH methylation of TEs during late embryo development is supported by previous reports that genes involved in RdDM such as AGO4, DRM2 and DMS3, and concomitant sRNAs, showed highest expression during late embryogenesis (Belmonte et al., 2013; Bouyer et al., 2017; Papareddy et al., 2020). Interestingly, Helitron transposons, which were the most enriched in terms of copy number and genome coverage among the superfamilies we analyzed (Figure 5A), showed the lowest CG and CHG methylation levels in RdDM at all stages (Supplementary Figures S3B, C). Taken together, our data suggest that the CHH regulatory pathways mediating TE methylation dynamics are independent of TE superfamily.

## The length of each TE class can impact methylation levels during embryogenesis

We next investigated the impact of TE length on embryo methylation levels and dynamics in different CHH regulatory pathways at DAP4 (Figures 6A–C) and DAP12 (Figures 6D–F). Almost all longer (>3 kb) TEs belonged to the cCMT2 TE category

in all cytosine contexts (green in Figures 6A–F), which was consistent with the findings in Figures 4F, 5B. Moreover, methylation levels for cCMT2 TEs tended to converge at a specific level as TE length increased in all cytosine contexts. Methylation levels increased from DAP4 to DAP12, especially in CG and CHH contexts (Figures 6A–F). This suggests that cCMT2 TEs maintain specific ranges of methylation levels regardless of their length, consistent with their collective methylation pattern around the centromere, observed above (Figure 4C; Supplementary Figure S3A). The CG methylation range for cCMT2 TEs was broader at DAP4, becoming narrower at DAP12 (Figures 6A–D). Interestingly, the opposite trend was observed for CHH methylation, as DAP12 exhibited a broader range of CHH methylation. These results indicate that as the embryo matures, cCMT2 TEs gain a higher and more uniform level of CG methylation (around 95%) whereas CHH increases are more variable, resulting in a wider range of levels DAP12. Intriguingly, although many cCMT2 TEs displayed low CHH methylation at DAP4 (around 5%), some short cCMT2 TEs exhibited very high CHH methylation (Figure 6C).

Unlike cCMT2 TEs, the majority of both RdDM TEs were found to reside in shorter TEs, with eRdDM TEs being slightly shorter than cRdDM TEs (Figures 6A–F). The overall methylation patterns, particularly for CG and CHG methylation, were similar between cRdDM TEs and eRdDM TEs (Figures 6A–E). The eRdDM TEs exhibiting high CG methylation levels showed almost the same levels as those in cCMT2 TEs, suggesting that these eRdDM TEs in the reproductive stage likely originated from cCMT2 TEs in the vegetative stage. CHH methylation levels of eRdDM TEs were found to be lower than those in cRdDM TEs at DAP4, and the increasing rate of CHH methylation of during embryogenesis was slower. Therefore, the highest CHH methylation level observed in eRdDM TEs was around 45%, compared with cRdDM TEs, which exhibited levels around 95% at DAP12 (Figures 6C–F).

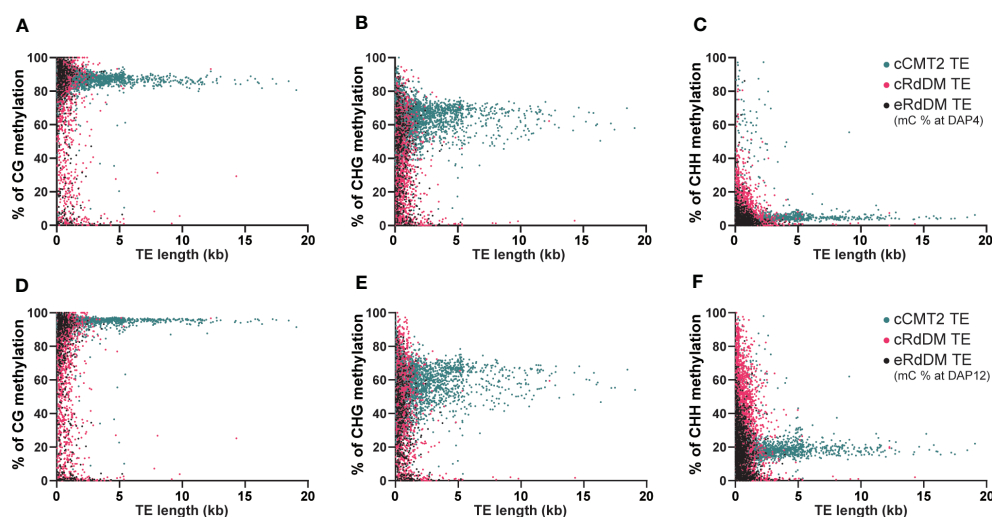


FIGURE 6

Analysis of DNA methylation levels according to the length of TEs. The DNA methylation levels at globular embryo (DAP4) (A–C) and those of mature green embryo (DAP12) (D–F) were used for this analysis. The DNA methylation levels on TEs were measured by calculating the mean percentage of mC at every cytosine site.

## Cytosine frequency and methylation levels in different TE classes

While it has been known that heterochromatic TEs have higher GC content than euchromatic TEs (Zemach et al., 2013), examining the frequency of all cytosine contexts may provide insights into the relationship between CHH regulatory dependency of TEs and their sequence composition. Our analysis revealed that cCMT2 TEs exhibited significantly higher cytosine frequency across all contexts compared to cRdDM TEs and eRdDM TEs (Figure 7A,  $p < 0.0001$ ). eRdDM TEs demonstrated the lowest CG and CHG frequency, but slightly higher CHH frequency compared to cRdDM TEs (Figure 7A). We measured the frequencies of CG, CHG and CHH in the three TE classes as well as those of all *Arabidopsis* TEs as a reference based on each TE superfamily. We also calculated the expected frequencies using arithmetic calculations (Table 1). As a result, we observed that the cytosine frequencies of most cCMT2 TEs were higher than two RdDM classes, as well as all TEs across all cytosine contexts. However, the patterns of cytosine frequencies differed between cRdDM TEs and eRdDM TEs, depending on the context. Both cRdDM TEs and eRdDM TEs exhibited lower CHG frequency compared to all TEs except for the SINE superfamily in cRdDM TEs (0.052) (Table 1; Figure 7A). Between cRdDM TEs and eRdDM TEs, eRdDM TEs showed lower CHG frequency but higher CHH frequency, except for the En-Spm superfamily (Table 1, Figure 7A). Notably, the CHH frequency of cRdDM TEs was the lowest among the three TE classes, suggesting that TEs with low CHH frequency are preferentially targeted by the RdDM pathway during vegetative growth. Since the region with the highest CHG frequency in cCMT2 TEs is likely to be more heterochromatic than the CHG region in RdDM TEs, our analysis suggests that the difference between cCMT2 TEs and eRdDM TEs may be attributed, at least in part, to the higher CHG frequency observed in cCMT2 TEs. This further implies that shorter CMT2 TEs with lower CHG frequency may contribute to epigenetic reprogramming during the transition from the vegetative to the reproductive lifecycle, and they

may be newly targeted by the RdDM pathway, becoming eRdDM TEs.

Recent research has shown that although euchromatic TEs are usually regulated by RdDM, young and LTR transposons in euchromatin are typically regulated by CMTs. Furthermore, longer and younger LTR transposons tend to have higher CHG methylation levels (Wang and Baulcombe, 2020). We analyzed the cytosine frequencies of predicted *de novo* LTRs using publicly available data (Wang and Baulcombe, 2020) and found that intact LTRs with higher LTR sequence similarity, which are thought to be more recently introduced in the genome, displayed diverse cytosine frequencies. By contrast, older LTRs with low similarity due to mutations had similar cytosine frequencies, suggesting they have decayed over time during evolution (Figure 7B). Taken together, our findings highlight the correlation between length, location, cytosine ratio, DNA methylation, and regulatory pathways of TEs.

## Discussion

In embryos, a programmed and gradual increase in CHH methylation throughout seed development has been reported in several dicot plants, including *Arabidopsis* (Bouyer et al., 2017; Kawakatsu et al., 2017; Lin et al., 2017), *Brassica rapa* (Grover et al., 2020), soybean (Lin et al., 2017), and chickpea (Rajkumar et al., 2020). However, the epigenetic pathways involved in this methylation reconfiguration have not been fully elucidated. In *Arabidopsis*, this gradual increase in CHH methylation is accompanied by at least two waves of massive sRNAs accumulation produced by thousands of TEs (Papareddy et al., 2020). Here, we provide results consistent with previous studies that during *Arabidopsis* embryo development, the level of CHH methylation gradually increases (Bouyer et al., 2017; Kawakatsu et al., 2017; Lin et al., 2017), whereas CG and CHG methylation remain high and constant. Our study reveals that the gradual increase in CHH methylation during embryo development is

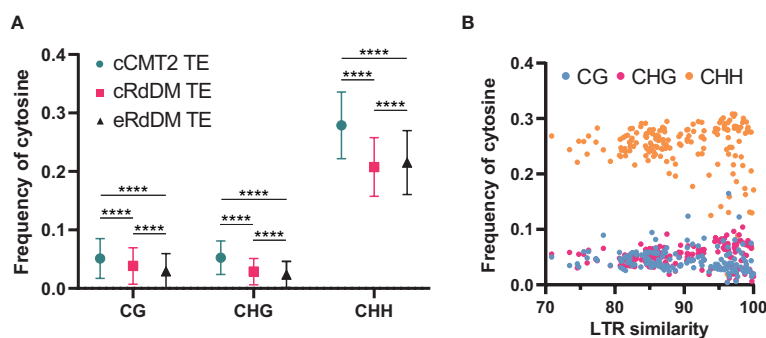


FIGURE 7

Calculation of methylation context frequencies on TEs. (A) The frequencies of each methylation context on TEs were calculated by dividing the number of the same methylation context with the length of TE. \*\*\*\* indicates a significant difference ( $p < 0.0001$ ). (B) The frequencies of each cytosine context on *de novo* intact LTRs were analyzed. Intact LTRs and their similarity were defined using a previous report (Wang and Baulcombe, 2020).

TABLE 1 The frequencies of cytosine methylation contexts on TE superfamily.

Class		All TEs			cCMT2 TEs			cRdDM TEs			eRdDM TEs		
mC context		CG	CHG	CHH	CG	CHG	CHH	CG	CHG	CHH	CG	CHG	CHH
Expectation		0.063	0.047	0.28	0.063	0.047	0.28	0.063	0.047	0.28	0.063	0.047	0.28
Observation	Copia	0.039	0.042	0.261	0.042	0.052	0.274	0.038	0.031	0.250	0.032	0.027	0.259
	Gypsy	0.055	0.053	0.280	0.057	0.055	0.290	0.049	0.037	0.257	0.035	0.032	0.270
	Helitron	0.027	0.023	0.201	0.034	0.040	0.240	0.030	0.022	0.195	0.024	0.020	0.205
	MuDR	0.036	0.030	0.220	0.049	0.050	0.265	0.036	0.026	0.213	0.032	0.025	0.313
	En-Spm	0.050	0.045	0.252	0.049	0.052	0.260	0.063	0.025	0.236	0.053	0.031	0.222
	LINE	0.054	0.048	0.258	0.050	0.060	0.291	0.061	0.042	0.231	0.049	0.040	0.259
	SINE	0.059	0.048	0.245	0.071	0.055	0.289	0.070	0.052	0.251	0.035	0.034	0.269

Note: Observed frequency was calculated by dividing each the counts of cytosine context by the length of TE. The background color was determined by higher (green) or lower (orange) than the frequency of all TEs (blue). Expected frequencies of CG, CHG and CHH were calculated as following formula; CG (0.25x0.25), CHG (0.25x0.75x0.25), CHH (0.25x0.75x0.75x2, double strands).

accompanied by the spreading of methylation in many loci. We showed that this process coincided with the expansion of sRNA expression, which was highly correlated with the expansion of CHH-methylated regions, as well as the increase in CHH methylation levels. This reconfiguration of CHH methylation was associated with sRNA-dependent features and sRNA-independent enzymes such as CMT2. Indeed, the expression of sRNAs and associated enzymes were positively correlated with the resetting of CHH methylation (Papareddy et al., 2020).

Our study showed that the dependency of individual loci on previously defined CHH methylation pathways (Hsieh et al., 2023) is likely dictated by chromosomal location, with cCMT2 DMRs predominantly concentrated at pericentromeric regions, cRdDM DMRs distributed throughout the chromosomes, and eRdDM DMRs located in-between. eRdDM DMRs are unique from He et al.'s CMT2-to-RdDM loci, which were identified in *ddm1* mutants. Firstly, most of the DMRs are not shared between the eRdDM class and He et al.'s CMT2-to-RdDM loci. Secondly, eRdDM TEs have an average length of less than 500 bp, while the average length of TEs overlapped with CMT2-to-RdDM DMRs are over 4kb (He et al., 2021). These results suggested that eRdDM TEs are prone to undergo a natural conversion from heterochromatic to euchromatic state during developmental transition, likely due to the reprogramming of CHH methylation during the transition from the vegetative to reproductive life cycle. Thus, while eRdDM DMRs are targeted by CMT2 in the leaf, they become accessible to the RdDM pathway following CHH reprogramming upon fertilization. This may be due to the fact that eRdDM TEs tend to be shorter and located towards the chromosomal arms compared to cCMT2 targets. In contrast, CMT2-to-RdDM loci are heterochromatic regions maintained by DDM1 and are thought to be stably regulated by CMT2 in the wild-type state.

Our analyses of TEs regulated by the three different CHH pathways revealed significant differences in terms of their size, methylation levels, and chromosomal location. Longer TEs tend to be targeted by CMT2 whereas shorter TEs are targeted by the

RdDM pathway. Interestingly, this trend remains consistent even within the same superfamily. Our results demonstrated that the size of TEs even within each TE superfamily is the crucial factor for their CHH enzyme dependency, regardless of their classification into different superfamilies. Therefore, the size of the TEs appears to be a more important factor than their family or origin. Compared with cRdDM TEs or eRdDM TEs, cCMT2 TEs tend to maintain specific ranges of methylation levels, especially for longer TEs. Intriguingly, shorter cCMT2 TEs show more variations in non-CG methylation levels at DAP12 compared to DAP4, which is the opposite of what is observed in CG methylation levels. The majority of both cRdDM TEs and eRdDM TEs are short in length and exhibit similar overall methylation patterns. When analyzing the frequency of cytosines in relation to TE length, we found cCMT2 TEs had a higher cytosine frequency at all contexts than both cRdDM TEs and eRdDM TEs. Specifically, a significant difference in CHG frequency depending on the regulatory pathway was observed in that CHG frequency in cCMT2 TE superfamilies tended to be higher than expected (green in CHG of Table 1) which was not observed in cRdDM TEs and eRdDM TEs superfamilies (Table 1). This raises the possibility that shorter CMT2 TEs with lower CHG frequency may be prone to undergoing epigenetic reprogramming during the transition and becoming accessible to RdDM machinery.

Taken together, our findings highlight the correlation between length, location, cytosine ratio, DNA methylation, and regulatory pathways of TEs. These factors all contribute to the evolutionary trajectory of TEs and by examining these characteristics, we can gain insights into the underlying mechanisms by which TEs evolved in host genomes.

## Data availability statement

The data presented in the study are deposited in the NCBI, accession number PRJNA944400 (<https://www.ncbi.nlm.nih.gov/project/PRJNA944400>).



## Author contributions

JL, KP, and YC designed the research. JL, KP, and SL performed the experiments. JL, KP, SL, JF, P-HH, RF, T-FH, and YC analyzed the methylome data. SS and CS analyzed small RNA clusters. JL, SL, T-FH, and YC wrote the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This work was funded by the National Research Foundation of Korea (NRF) Grant 2020R1A2C2009382 and 2021R1A5A1032428 to YC. This work was also supported by the Stadelmann-Lee Scholarship Fund, Seoul National University, to JL and SL and by the NSF Grant MCB-1715115 to T-FH.

## Acknowledgments

We thank Hyunjin Yoo for helping with embryo collection.

## References

- Ahmed, I., Sarazin, A., Bowler, C., Colot, V., and Quesneville, H. (2011). Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in arabidopsis. *Nucleic Acids Res.* 39, 6919–6931. doi: 10.1093/nar/gkr324
- Batista, R. A., and Kohler, C. (2020). Genomic imprinting in plants-revisiting existing models. *Genes Dev.* 34, 24–36. doi: 10.1101/gad.332924.119
- Belmonte, M. F., Kirkbride, R. C., Stone, S. L., Pelletier, J. M., Bui, A. Q., Yeung, E. C., et al. (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc. Natl. Acad. Sci.* 110 (5), E435–E444. doi: 10.1073/pnas.1222061110
- Bewick, A. J., and Schmitz, R. J. (2017). Gene body DNA methylation in plants. *Curr. Opin. Plant Biol.* 36, 103–110. doi: 10.1016/j.pbi.2016.12.007
- Bourguet, P., Picard, C. L., Yelagandula, R., Pelissier, T., Lorkovic, Z. J., Feng, S., et al. (2021). The histone variant H2A.W and linker histone H1 co-regulate heterochromatin accessibility and DNA methylation. *Nat. Commun.* 12, 2683. doi: 10.1038/s41467-021-22993-5
- Bouyer, D., Kramdi, A., Kassam, M., Heese, M., Schnittger, A., Roudier, F., et al. (2017). DNA Methylation dynamics during early plant life. *Genome Biol.* 18, 179. doi: 10.1186/s13059-017-1313-0
- Cao, X., and Jacobsen, S. E. (2002). Role of the arabidopsis DRM methyltransferases in *De novo* DNA methylation and gene silencing. *Curr. Biol.* 12, 1138–1144. doi: 10.1016/S0960-9822(02)00925-9
- Chenais, B., Caruso, A., Hiard, S., and Casse, N. (2012). The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene* 509, 7–15. doi: 10.1016/j.gene.2012.07.042
- Chow, H. T., Chakraborty, T., and Mosher, R. A. (2020). RNA-Directed DNA methylation and sexual reproduction: expanding beyond the seed. *Curr. Opin. Plant Biol.* 54, 11–17. doi: 10.1016/j.pbi.2019.11.006
- Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., et al. (2008). Shotgun bisulphite sequencing of the arabidopsis genome reveals DNA methylation patterning. *Nature* 452, 215–219. doi: 10.1038/nature06745
- Deniz, O., Frost, J. M., and Branco, M. R. (2019). Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* 20, 417–431. doi: 10.1038/s41576-019-0106-6
- Erdmann, R. M., and Picard, C. L. (2020). RNA-Directed DNA methylation. *PLoS Genet.* 16, e1009034. doi: 10.1371/journal.pgen.1009034
- Gehring, M., Bubb, K. L., and Henikoff, S. (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* 324, 1447–1451. doi: 10.1126/science.1171609
- Gehring, M., Huh, J. H., Hsieh, T. F., Penterman, J., Choi, Y., Harada, J. J., et al. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* 124, 495–506. doi: 10.1016/j.cell.2005.12.034
- Gehring, M., and Satyaki, P. R. (2017). Endosperm and imprinting, inextricably linked. *Plant Physiol.* 173, 143–154. doi: 10.1104/pp.16.01353
- Gent, J. I., Higgins, K. M., Swentowsky, K. W., Fu, F. F., Zeng, Y. B., Kim, D. W., et al. (2022). The maize gene maternal derepression of r1 encodes a DNA glycosylase that demethylates DNA and reduces siRNA expression in the endosperm. *Plant Cell* 34, 3685–3701. doi: 10.1093/plcell/koac199
- 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. doi: 10.1038/s41580-019-0159-6
- Grover, J. W., Burgess, D., Kendall, T., Baten, A., Pokhrel, S., King, G. J., et al. (2020). Abundant expression of maternal siRNAs is a conserved feature of seed development. *Proc. Natl. Acad. Sci. U.S.A.* 117, 15305–15315. doi: 10.1073/pnas.2001332117
- He, L., Zhao, C., Zhang, Q., Zinta, G., Wang, D., Lozano-Duran, R., et al. (2021). Pathway conversion enables a double-lock mechanism to maintain DNA methylation and genome stability. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2107320118. doi: 10.1073/pnas.2107320118
- Heard, E., and Martienssen, R. A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* 157, 95–109. doi: 10.1016/j.cell.2014.02.045
- Hsieh, P.-H., Frost, J. M., Choi, Y., Hsieh, T.-F., Zilberman, D., and Fischer, R. L. (2023). Embryo-specific epigenetic mechanisms reconstitute the CHH methylation landscape during arabidopsis embryogenesis. *bioRxiv* 2023-04. doi: 10.1101/2023.04.06.535361
- Hsieh, P. H., He, S. B., Buttress, T., Gao, H. B., Couchman, M., Fischer, R. L., et al. (2016). Arabidopsis male sexual lineage exhibits more robust maintenance of CG methylation than somatic tissues. *Proc. Natl. Acad. Sci. United States America* 113, 15132–15137. doi: 10.1073/pnas.1619074114
- Hsieh, T. F., Ibarra, C. A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R. L., et al. (2009). Genome-wide demethylation of arabidopsis endosperm. *Science* 324, 1451–1454. doi: 10.1126/science.1172417
- Ibarra, C. A., Feng, X., Schoft, V. K., Hsieh, T. F., Uzawa, R., Rodrigues, J. A., et al. (2012). Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* 337, 1360–1364. doi: 10.1126/science.1224839
- Johnson, N. R., Yeoh, J. M., Coruh, C., and Axtell, M. J. (2016). Improved placement of multi-mapping small RNAs. *G3 (Bethesda)* 6, 2103–2111. doi: 10.1534/g3.116.030452

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1204279/full#supplementary-material>



- Jurka, J., Bao, W., and Kojima, K. K. (2011). Families of transposable elements, population structure and the origin of species. *Biol. Direct* 6, 44. doi: 10.1186/1745-6150-6-44
- Kawakatsu, T., Huang, S. S., Jupe, F., Sasaki, E., Schmitz, R. J., Urlich, M. A., et al. (2016). Epigenomic diversity in a global collection of arabidopsis thaliana accessions. *Cell* 166, 492–505. doi: 10.1016/j.cell.2016.06.044
- Kawakatsu, T., Nery, J. R., Castanon, R., and Ecker, J. R. (2017). Dynamic DNA methylation reconfiguration during seed development and germination. *Genome Biol.* 18, 171. doi: 10.1186/s13059-017-1251-x
- Kenchanmane Raju, S. K., Ritter, E. J., and Niederhuth, C. E. (2019). Establishment, maintenance, and biological roles of non-CG methylation in plants. *Essays Biochem.* 63, 743–755. doi: 10.1042/EBC20190032
- Kim, M. Y., Ono, A., Scholten, S., Kinoshita, T., Zilberman, D., Okamoto, T., et al. (2019). DNA Demethylation by ROS1a in rice vegetative cells promotes methylation in sperm. *Proc. Natl. Acad. Sci. U.S.A.* 116, 9652–9657. doi: 10.1073/pnas.1821435116
- Lanciano, S., and Cristofari, G. (2020). Measuring and interpreting transposable element expression. *Nat. Rev. Genet.* 21, 721–736. doi: 10.1038/s41576-020-0251-y
- Lauria, M., Rupe, M., Guo, M., Kranz, E., Pirona, R., Viotti, A., et al. (2004). Extensive maternal DNA hypomethylation in the endosperm of zea mays. *Plant Cell* 16, 510–522. doi: 10.1105/tpc.017780
- 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. doi: 10.1038/nrg2719
- Lin, J. Y., Le, B. H., Chen, M., Henry, K. F., Hur, J., Hsieh, T. F., et al. (2017). Similarity between soybean and arabidopsis seed methylomes and loss of non-CG methylation does not affect seed development. *Proc. Natl. Acad. Sci. United States America* 114, E9730–E9739. doi: 10.1073/pnas.1716758114
- Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., et al. (2008). Highly integrated SINGLe-base resolution maps of the epigenome in arabidopsis. *Cell* 133, 523–536. doi: 10.1016/j.cell.2008.03.029
- Lutzmayer, S., Enugutti, B., and Nodine, M. D. (2017). Novel small RNA spike-in oligonucleotides enable absolute normalization of small RNA-seq data. *Sci. Rep.* 7, 1–6. doi: 10.1038/s41598-017-06174-3
- Matzke, M. A., Kanno, T., and Matzke, A. J. (2015). RNA-Directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants. *Annu. Rev. Plant Biol.* 66, 243–267. doi: 10.1146/annurev-arplant-043014-114633
- Matzke, M. A., and Mosher, R. A. (2014). RNA-Directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* 15, 394–408. doi: 10.1038/nrg3683
- Ono, A., Yamaguchi, K., Fukada-Tanaka, S., Terada, R., Mitsui, T., and Iida, S. (2012). A null mutation of ROS1a for DNA demethylation in rice is not transmittable to progeny. *Plant J.* 71, 564–574. doi: 10.1111/j.1365-313X.2012.05009.x
- Papareddy, R. K., Paldi, K., Paulraj, S., Kao, P., Lutzmayer, S., and Nodine, M. D. (2020). Chromatin regulates expression of small RNAs to help maintain transposon methylome homeostasis in arabidopsis. *Genome Biol.* 21, 251. doi: 10.1186/s13059-020-02163-4
- Park, K., Kim, M. Y., Vickers, M., Park, J. S., Hyun, Y., Okamoto, T., et al. (2016). DNA Demethylation is initiated in the central cells of arabidopsis and rice. *Proc. Natl. Acad. Sci. U.S.A.* 113, 15138–15143. doi: 10.1073/pnas.1619047114
- Picard, C. L., and Gehring, M. (2017). Proximal methylation features associated with nonrandom changes in gene body methylation. *Genome Biol.* 18, 73. doi: 10.1186/s13059-017-1206-2
- Plotnikova, A., Kellner, M. J., Schon, M. A., Mosiolek, M., and Nodine, M. D. (2019). MicroRNA dynamics and functions during arabidopsis embryogenesis. *Plant Cell* 31, 2929–2946. doi: 10.1105/tpc.19.00395
- Quadrana, L., Bortolini Silveira, A., Mayhew, G. F., Leblanc, C., Martienssen, R. A., Jeddloh, J. A., et al. (2016). The arabidopsis thaliana mobilome and its impact at the species level. *Elife* 5, e15716. doi: 10.7554/eLife.15716.046
- Quesneville, H. (2020). Twenty years of transposable element analysis in the arabidopsis thaliana genome. *Mob DNA* 11, 28. doi: 10.1186/s13100-020-00223-x
- Rajkumar, M. S., Gupta, K., Khemka, N. K., Garg, R., and Jain, M. (2020). DNA Methylation reprogramming during seed development and its functional relevance in seed size/weight determination in chickpea. *Commun. Biol.* 3. doi: 10.1038/s42003-020-1059-1
- Sasaki, H., and Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat. Rev. Genet.* 9, 129–140. doi: 10.1038/nrg2295
- Schoft, V. K., Chumak, N., Choi, Y., Hannon, M., Garcia-Aguilar, M., Machlicova, A., et al. (2011). Function of the DEMETER DNA glycosylase in the arabidopsis thaliana male gametophyte. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8042–8047. doi: 10.1073/pnas.1105117108
- 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. doi: 10.1098/rstb.2011.0330
- Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., and Matsui, Y. (2005). Extensive and orderly reprogramming of genomic-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev. Biol.* 278, 440–458. doi: 10.1016/j.ydbio.2004.11.025
- Sigman, M. J., and Slotkin, R. K. (2016). The first rule of plant transposable element silencing: location, location, location. *Plant Cell* 28, 304–313. doi: 10.1105/tpc.15.00869
- Stroud, H., Greenberg, M. V., 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. doi: 10.1016/j.cell.2012.10.054
- Turker, M. S. (2002). Gene silencing in mammalian cells and the spread of DNA methylation. *Oncogene* 21, 5388–5393. doi: 10.1038/sj.onc.1205599
- Wang, Z., and Baulcombe, D. C. (2020). Transposon age and non-CG methylation. *Nat. Commun.* 11, 1221. doi: 10.1038/s41467-020-14995-6
- Xing, M. Q., Zhang, Y. J., Zhou, S. R., Hu, W. Y., Wu, X. T., Ye, Y. J., et al. (2015). Global analysis reveals the crucial roles of DNA methylation during rice seed development. *Plant Physiol.* 168, 1417–U1554. doi: 10.1104/pp.15.00414
- Xu, Q., Wu, L., Luo, Z., Zhang, M., Lai, J., Li, L., et al. (2022). DNA Demethylation affects imprinted gene expression in maize endosperm. *Genome Biol.* 23, 77. doi: 10.1186/s13059-022-02641-x
- Xu, W., Yang, T., Dong, X., Li, D. Z., and Liu, A. (2016). Genomic DNA methylation analyses reveal the distinct profiles in castor bean seeds with persistent endosperms. *Plant Physiol.* 171, 1242–1258. doi: 10.1104/pp.16.00056
- Yelagandula, R., Stroud, H., Holec, S., Zhou, K., Feng, S., Zhong, X., et al. (2014). The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in arabidopsis. *Cell* 158, 98–109. doi: 10.1016/j.cell.2014.06.006
- Yoo, H., Park, K., Lee, J., Lee, S., and Choi, Y. (2021). An optimized method for the construction of a DNA methylome from small quantities of tissue or purified DNA from arabidopsis embryo. *Mol. Cells* 44, 602–612. doi: 10.14348/molcells.2021.0084
- Zemach, A., Kim, M. Y., Hsieh, P. H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., et al. (2013). The arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 153, 193–205. doi: 10.1016/j.cell.2013.02.033
- Zemach, A., Kim, M. Y., Silva, P., Rodrigues, J. A., Dotson, B., Brooks, M. D., et al. (2010). Local DNA hypomethylation activates genes in rice endosperm. *Proc. Natl. Acad. Sci. United States America* 107, 18729–18734. doi: 10.1073/pnas.1009695107
- Zhao, S., Cheng, L., Gao, Y., Zhang, B., Zheng, X., Wang, L., et al. (2019). Plant HP1 protein ADCP1 links multivalent H3K9 methylation readout to heterochromatin formation. *Cell Res.* 29, 54–66. doi: 10.1038/s41422-018-0104-9
- Ziller, M. J., Hansen, K. D., Meissner, A., and Aryee, M. J. (2015). Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nat. Methods* 12, 230–232. doi: 10.1038/nmeth.3152