

Domains rearranged methyltransferases (DRMs)-mediated DNA methylation plays key roles in modulating gene expression and maintaining transposable element silencing in soybean

Hongwei Xun^{1†}, Lijie Lian^{2†}, Jing Yuan³, Jianhui Hong¹, Shanmeng Hao², Haonan Zhao², Shuhan Liu¹, Wanjie Feng², Huanran Yin², Bao Liu^{1*} and Xutong Wang^{2*}

1. Key Laboratory of Molecular Epigenetics of MOE, Northeast Normal University, Changchun 130024, China

2. National Key Laboratory of Crop Genetic Improvement, College of Plant Science and Technology, Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan 430070, China

3. National Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, China

[†]These authors contributed equally to this work.

*Correspondences: Bao Liu (baoliu@nenu.edu.cn); Xutong Wang (xtwang@mail.hzau.edu.cn). Dr. Wang is fully responsible for the distribution of all materials associated with this article in the manuscript)



Hongwei Xun



Xutong Wang

ABSTRACT

The domains rearranged methyltransferases (DRMs) play a critical role in the RNA-directed DNA methylation (RdDM) pathway in plants. However, the effects of inactivating the RdDM pathway on gene expression, transposable element (TE) activity, and phenotype in soybean remain unexplored. Here, we employed clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 gene editing to generate a quintuple mutant line in soybean (*Gmdrm2a*^{-/-}*2b*^{-/-}*2c*^{-/-}*3a*^{-/-}*3b*^{-/-}, designated *Gmdrm*). *Gmdrm* exhibited severe developmental abnormalities, including dwarfism and delayed growth, albeit remaining viable and fertile; however, the fully homozygous mutant could be maintained for a limited number of generations (T0–T3). Whole genome bisulfite sequencing revealed a significant reduction in DNA methylation across all cytosine sequence

contexts, with an average loss of 10%. The loss of ^mC was biased toward euchromatic regions, which is in contrast to the chromomethylase mutant. Transcriptome profiling identified 1,685 up-regulated genes, including photosynthesis-related genes, accompanied with altered chloroplast ultrastructure. Additionally, a cluster of resistance (*R*) genes on chromosome 16 was significantly up-regulated, coinciding with their reduced non-CG methylation. We also observed 3,164 differentially expressed TEs (DETs), of which, 2,655 were up-regulated and hypomethylated along their entire length. A substantial reduction in the abundance of 24-nt small interfering RNAs (siRNAs) in the *Gmdrm* mutant was detected by small RNA sequencing. Of note, the DRM-targeted TEs typically display higher levels of 24-nt siRNA abundance, shorter lengths, and are more AT-rich compared to chromomethylase-targeted TEs, highlighting 24-nt siRNAs as key determinants of DRM-dependent TE regulation. Together, this study documents a critical role of DRM-mediated DNA methylation in regulating gene expression, TE silencing, and normal development in soybean.

Keywords: 24-nt siRNAs, *de novo* DNA methylation, *Glycine max*, photosynthesis, *R* genes, RdDM, transposable element

Xun, H., Lian, L., Yuan, J., Hong, J., Hao, S., Zhao, H., Liu, S., Feng, W., Yin, H., Liu, B., et al. (2025). Domains rearranged methyltransferases (DRMs)-mediated DNA methylation plays key roles in modulating gene expression and maintaining transposable element silencing in soybean. *J. Integr. Plant Biol.* 00: 1–14.

INTRODUCTION

DNA methylation at the fifth position of cytosine is a crucial epigenetic modification in both animals and plants, playing a vital role in regulating stress responses and development. In plant genomes, this methylation occurs in three distinct cytosine sequence contexts: CG, CHG, and CHH (where H represents A, C, or T) (Henderson and Jacobsen, 2007; Stroud et al., 2014). Each context is established and maintained by specific families of DNA methyltransferases during DNA replication. Notably, DNA methyltransferase 1 (MET1) is responsible for maintaining methylation in the CG context. It recognizes hemimethylated CG dinucleotides, generated during DNA replication, and methylates the unmodified cytosine on the newly synthesized daughter strand. In contrast, CHG methylation is primarily maintained by the plant-specific chromomethylase 3 (CMT3), with additional contributions from CMT2. This process is facilitated by the histone modification H3K9me2 (Henderson and Jacobsen, 2007; Stroud et al., 2014). Collectively, these DNA methyltransferases ensure the proper establishment and maintenance of DNA methylation patterns across the plant genome, which are essential for normal growth and development.

Asymmetric CHH methylation is established and maintained by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), a homolog of DNMT3, primarily through RNA-directed DNA methylation (RdDM) to regulate *de novo* cytosine methylation in all sequence contexts, with additional contributions from CMT2 and the chromatin remodeling factor DECREASE IN DNA METHYLATION 1 (DDM1). DRM2 preferentially localizes to short, evolutionarily young transposable elements (TEs), as well as other repetitive sequences within euchromatic chromosome arms and at the edges of long TEs. In *Arabidopsis thaliana*, three *DRM* genes—*DRM1*, *DRM2*, and *DRM3*—have been identified. The *drm1 drm2* double mutant exhibits reduced CHH methylation and influences the silencing of specific genes and TEs (Cao and Jacobsen, 2002a; Cao et al., 2003; Tran et al., 2005; Stroud et al., 2013). Interestingly, while neither the *drm1 drm2* double mutant nor the *cmt3* single mutant displays overt morphological changes, the *drm1 drm2 cmt3* triple mutant shows pleiotropic developmental defects, such as dwarfism, distorted leaves, reduced fertility, and misregulation of numerous genes (Lindroth et al., 2001; Cao and Jacobsen, 2002a, 2002b). DRM3 is known to stimulate the catalytic activities of DRM1 and DRM2 (Henderson et al., 2010; Costa-Nunes et al., 2014; Zhong et al., 2015). In maize, *DRM2* RNA interference (RNAi) lines (*dmt103*) exhibit severe defects in seed morphology during reproductive development but remain unaffected morphologically during vegetative growth (Garcia-Aguilar et al., 2010). Similarly, in *Setaria viridis*, the *Drm1a Drm1b* double mutant plants exhibit delayed growth and flowering, although they remain viable, with a genome-wide reduction in CG and CHG methylation by 2%–6% and CHH methylation by ~33%, particularly in regions with high CHH methylation in wild-type (WT) plants (Read et al., 2022). However, the loss of CHH methylation surrounding the transcription

start site or in nearby promoter regions does not markedly alter gene expression, suggesting a limited regulatory role on gene expression by RdDM (Read et al., 2022). In rice, the majority of CHH methylation is abolished in the *Osdrm2* mutant (Moritoh et al., 2012; Hu et al., 2021). However, while single mutants of *OsDRM1a*, *OsDRM1b*, or *OsDRM3* do not significantly affect overall CHG or CHH methylation levels, *OsDRM2*, *OscMT2*, and *OscMT3a* work together to regulate TE and gene transcription (Hu et al., 2021). Notably, *Oscmt3a* has a more pronounced effect on TE regulation than *Osdrm2* or *Oscmt2*, suggesting that CHG methylation plays a larger role in TE silencing than CHH methylation in rice (Hu et al., 2021). Collectively, these studies highlight the critical role of DRM2 in CHH methylation and its functional redundancy with CMT3 in maintaining non-CG methylation across *Arabidopsis*, rice, and *Setaria viridis*.

As a leading oilseed crop worldwide, soybean (*Glycine max* L. Merr.) has a complex genomic structure in which approximately 75% of its genes occur in multiple copies, and around 59% of the genome consists of TEs (Schmutz et al., 2010; Liu et al., 2020). The RdDM pathway plays a critical role in regulating both gene expression and TE silencing (Erdmann and Picard, 2020). However, until now, no studies have explored the effects of RdDM pathway inactivation on gene expression, TE activity, or phenotypic changes in soybean. In this study, we successfully generated a quintuple mutant line targeting five putative functional DRM orthologs in soybean. We employed whole genome bisulfite sequencing (WGBS), transcriptome analyses and small RNA (sRNA) sequencing to investigate the role of DRM-mediated DNA methylation in regulating gene expression and TE silencing. Our findings provide new insights into the complex interplay between DNA methylation and transcriptional regulation, offering valuable knowledge about the epigenetic mechanisms that control plant growth, development, and safeguard genome stability.

RESULTS

Developmental abnormalities in *Gmdrm* quintuple mutant affect both vegetative and reproductive stages

Soybean, an ancient polyploid species, contains five copies of *DRM* genes in its genome, as revealed through sequence analysis and phylogenetic studies. These five *DRM* genes, designated as *GmDRM2a* (Glyma.05G005600), *GmDRM2b* (Glyma.19G006100), *GmDRM2c* (Glyma.02G035700), *GmDRM3a* (Glyma.07G233200), and *GmDRM3b* (Glyma.17G038300) (Figure 1A), contrast with *Arabidopsis*, which possesses only three *DRM* genes. To determine their possible tissue-specific functions, we performed quantitative reverse transcription polymerase chain reaction (qRT-PCR) to examine the expression levels of these *GmDRM* genes across various soybean tissues, including roots, leaves, flowers, seeds, and pods. Our analysis revealed that *GmDRM2a* and *GmDRM2b* are predominantly expressed

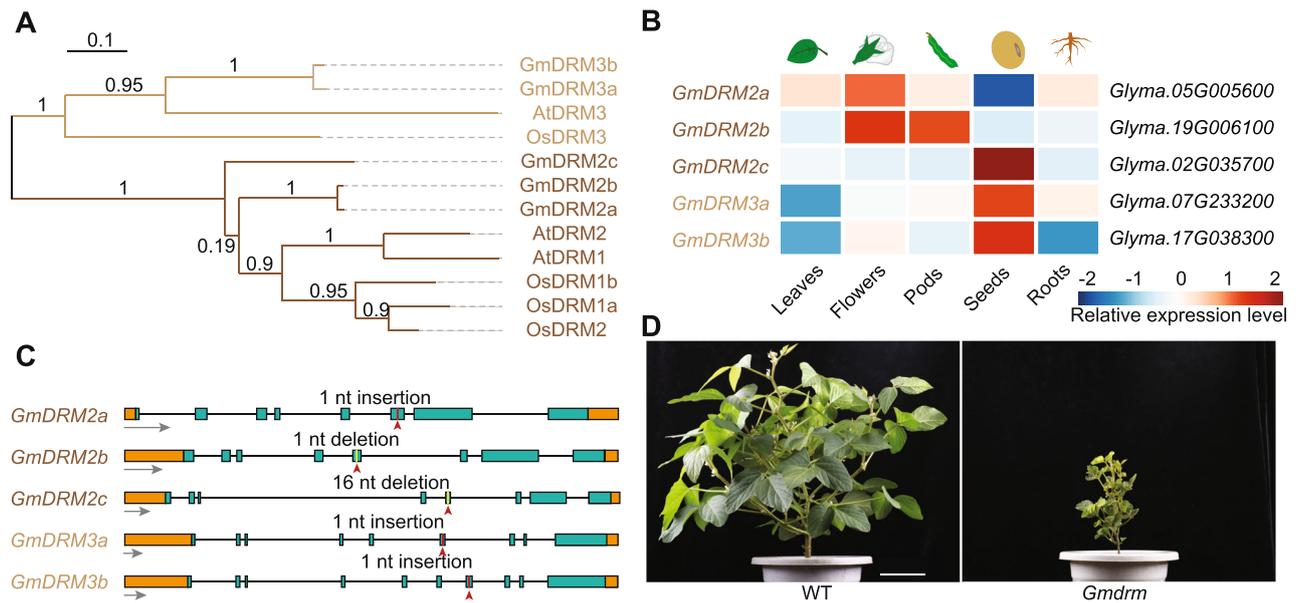


Figure 1. Generation and phenotypic characterization of the *Gmdrm* quintuple mutant in soybean

(A) A neighbor-joining tree was constructed to illustrate the evolutionary relationships among *GmDRM* family members. Branch support values are indicated above the branches, and the scale bar represents branch length, providing a measure of evolutionary distance. (B) Expression levels of *GmDRMs* were quantified in various tissues, including leaves (collected from 28-d-old plants), flowers (collected at the R2 stage), pods (collected at the R5 stage), and seeds (collected at the R8 stage), roots (collected from 28-d-old plants). The expression of *GmActin11* was used as an internal control to normalize the data. (C) The successful generation of the *Gmdrm* quintuple mutant was confirmed through polymerase chain reaction (PCR) amplification, followed by Sanger sequencing, verifying the editing status of all five *DRM* genes. (D) Comparison of the phenotypes of the *Gmdrm* quintuple mutant and wild-type (WT) plants grown under greenhouse conditions. Photographs were taken of 65-d-old plants. Scale bar = 12 cm.

in flowers and pods, while *GmDRM2c*, *GmDRM3a*, and *GmDRM3b* exhibit higher expression in seeds (Figure 1B), suggesting their differential roles in tissue-specific development.

To investigate the functional importance of these genes in plant development, we used the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) system to generate a quintuple *Gmdrm* mutant. We successfully identified a homozygous knockout line with targeted mutations in all five *DRM* genes: *GmDRM2a* harbored a 1 bp insertion in the sixth exon, *GmDRM2b* had a 1 bp deletion in the sixth exon, *GmDRM2c* carried a 16 bp deletion and one single nucleotide polymorphism in the fifth exon, *GmDRM3a* exhibited a 1 bp insertion in the sixth exon, and *GmDRM3b* contained a 1 bp insertion in the seventh exon (Figures 1C, S1). Under greenhouse conditions, the *Gmdrm* mutant of T2 generation exhibited a range of developmental abnormalities during the vegetative stage, including growth retardation, dwarfism, and leaf shrinkage. Additionally, at the reproductive stage, it showed delayed heading and defects in seed development. Although the mutant remained fertile, it could only be propagated as a homozygous mutant for three generations due to poor seed quality (Figures 1D, S2). These observations highlight the significant phenotypic alterations associated with the *Gmdrm* mutant, underscoring its essential role in soybean growth and development.

To investigate functional redundancies between *DRM2* and *DRM3* genes, we screened additional mutant lines, including several newly generated partial mutants:

Gmdrm2c^{-/-}*3a*^{-/-}*3b*^{-/-}, *Gmdrm2a*^{-/-}*2b*^{-/-}*2c*^{-/-}*3a*^{-/-}*3b*^{+/-}, and *Gmdrm2b*^{-/-}*2c*^{-/-}*3a*^{-/-}*3b*^{-/-}*2a*^{+/-} (Figure S3). Methylome analysis of these mutants revealed that the *Gmdrm2c*^{-/-}*3a*^{-/-}*3b*^{-/-} mutant showed no reduction in DNA methylation compared to the WT. In contrast, the *Gmdrm2a*^{-/-}*2b*^{-/-}*2c*^{-/-}*3a*^{-/-}*3b*^{+/-} and *Gmdrm2b*^{-/-}*2c*^{-/-}*3a*^{-/-}*3b*^{-/-}*2a*^{+/-} mutants exhibited reductions of 3.6% in mCG, 5.7% in mCHG, and 16% in mCHH methylation. Notably, these hypomethylation levels were substantially lower than the 7.2% and 8.6% reductions observed for mCG and mCHG methylation, respectively, in the *Gmdrm* quintuple mutant (Figure S3). These findings suggest that *DRM2* and *DRM3* exhibit functional redundancy in soybean, as null mutations in either gene result in only limited changes in DNA methylation. Our results indicate that the simultaneous inactivation of both *DRM2* and *DRM3* genes is necessary to induce significant methylation changes, underscoring their overlapping roles in maintaining DNA methylation in the soybean genome.

Widespread DNA hypomethylation in *Gmdrm* mutant, particularly in euchromatic regions

To explore the role of *GmDRMs* in soybean genome methylation, we performed WGBS on leaf tissues from both the *Gmdrm* mutant and WT plants. When dividing the genome into euchromatin and heterochromatin regions, we observed higher methylation levels in heterochromatin than in euchromatin regions for both the WT and mutants in all three sequence contexts (Table S1), as expected. The loss-of-function of all five *DRM* genes resulted in significant

reductions in DNA methylation: 9.6% in CG, 12.1% in CHG, and 10.5% in CHH contexts compared to the WT (Figure 2A, B). Further analyses identified a greater number of hypomethylated differentially methylated regions (DMRs) than hypermethylated ones in the *Gmdrm* mutant compared to the WT (Figure 2C). When comparing these DMRs to those in the *Gmcmt* mutant that we characterized recently (Xun et al., 2024), we found that the hypo-DMRs in *Gmcmt* were predominantly enriched in heterochromatin, while in *Gmdrm*, hypomethylation was mainly concentrated in euchromatin (Figure 2D) (Xun et al., 2024).

Specifically, methylation in the *Gmdrm* mutant was reduced by 13%, 19%, and 20% in CG, CHG, and CHH contexts, respectively, in euchromatin, and only by 4%, 3%, and 2% in heterochromatin for the same contexts (Figure 2A; Table S2).

Notably, 57.2% of hypomethylated DMRs in the CHH context were shared between *Gmcmt* and *Gmdrm*, while only 31.7% in CHG and 7.6% in CG were common between the two mutants (Figure 2C; Tables S3, S4). This suggests that the genomic regions regulated by DRM- and CMT-dependent methylation are highly disparate. These findings

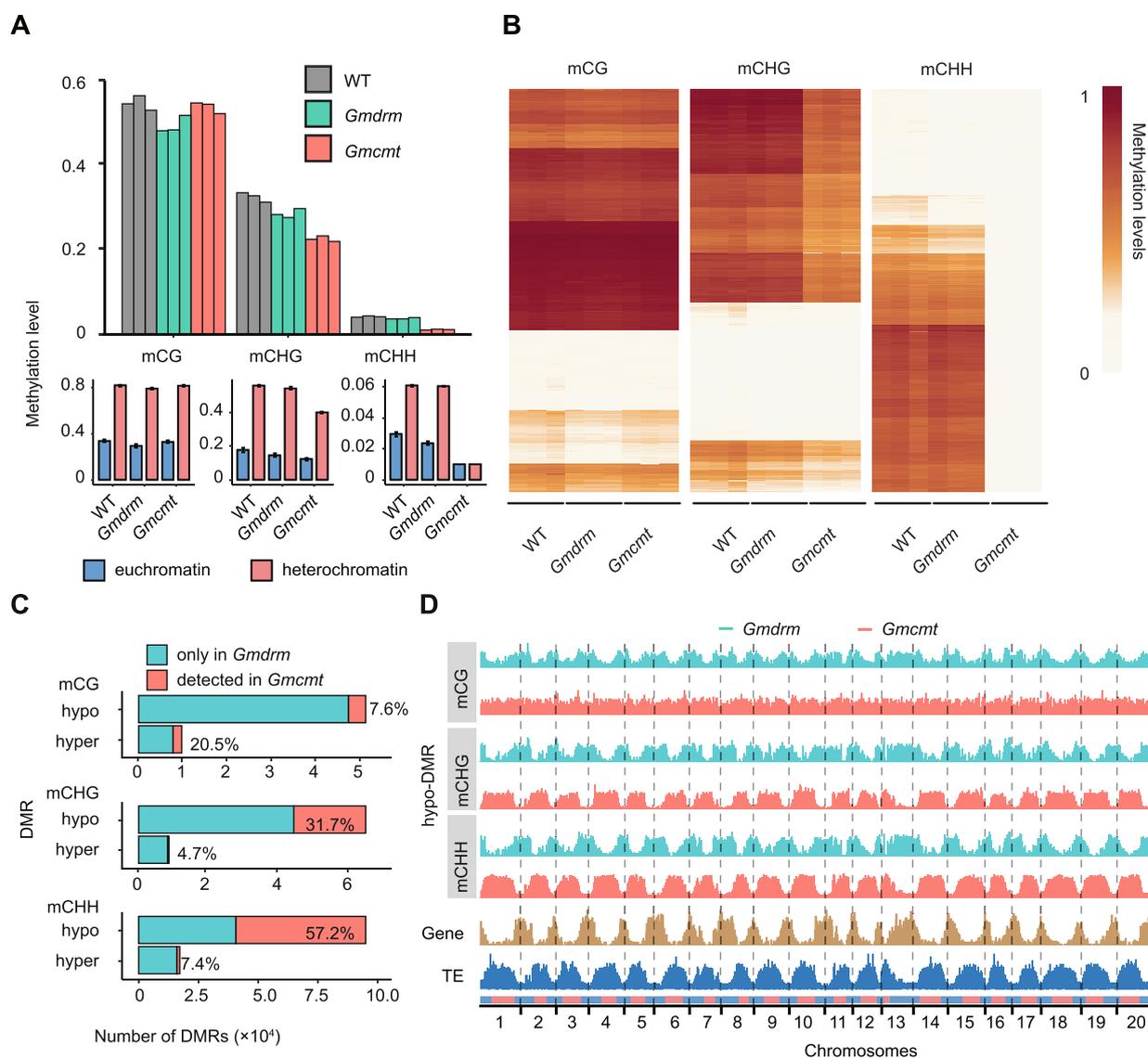


Figure 2. The effects on CG, CHG, and CHH DNA methylation in the *Gmdrm* and *Gmcmt* mutants

(A) Comparative analysis of mCG, mCHG, and mCHH methylation patterns between *Gmdrm*, *Gmcmt*, and wild-type (WT) plants across the whole genome (upper panel) and specifically within euchromatic and heterochromatic regions (lower panel). (B) Heatmap comparison of mCG, mCHG, and mCHH methylation patterns between *Gmdrm*, *Gmcmt*, and WT. Columns represent the different genotypes, while rows display methylation levels at cytosine sites, categorized by their mC context. (C) Number of differentially methylated regions (DMRs) detected in *Gmdrm* across the three cytosine contexts (CG, CHG, CHH). The green box represents DMRs unique to *Gmdrm*, while the red box indicates DMRs shared between *Gmdrm* and *Gmcmt*. (D) Genome-wide distribution of hypo-DMRs in *Gmdrm* and *Gmcmt*. Green shading represents the distribution of hypo-DMRs in *Gmdrm*, while red shading represents those in *Gmcmt*. Brown and blue lines indicate gene and transposable element (TE) densities, respectively, across the 20 soybean chromosomes. The pink and blue lines on the chromosomes represent the arm region and peri-centromere region, respectively.

indicate that the RdDM pathway predominantly regulates euchromatin DNA methylation in soybean, while non-CG methylation in heterochromatin is more controlled by CMT.

Transposable element activation correlates with hypomethylation in *Gmdrm* mutant

To investigate genome-wide patterns of DNA methylation changes in genes and TEs in the *Gmdrm* mutant, we profiled the methylation landscape surrounding genes and TEs. We observed a slight reduction in non-CG methylation across genic regions, including gene bodies, promoters, and terminators. Interestingly, CG methylation (mCG) levels decreased in promoters and terminators but remained unchanged in gene bodies in the *Gmdrm* mutant (Figure 3A). In contrast, the *Gmcm1* mutant showed a significant reduction in non-CG methylation across and around genes and TEs (Figure 3A). A more detailed analysis of the methylation landscape in TEs revealed greater non-CG hypomethylation surrounding TEs in the *Gmcm1* mutant compared to the *Gmdrm* mutant, which aligns with the genome-wide methylation changes observed in both mutants relative to the WT. Additionally, methylation levels in the body of TEs were higher than in gene bodies in the *Gmdrm* mutant (Figure 3A). These findings suggest functional differences between CMT- and DRM-regulated methylation of genes and TEs in soybean.

To further assess the impact of DRM in the RdDM pathway on gene expression and TE activation, we performed transcriptomic analysis on the second trifoliolate leaves of 21-d-old *Gmdrm* mutant and WT plants. This comparison revealed a comprehensive landscape of transcriptional changes: 1,685 differentially expressed genes (DEGs) were up-regulated, 1,405 DEGs were down-regulated, 2,655 differentially expressed TEs (DETs) were up-regulated, and 509 DETs were down-regulated in the *Gmdrm* mutant relative to the WT (Figure 3B, C; Tables S5, S6). We further analyzed the relationship between these DEGs and DETs with methylation alterations by classifying them into methylation-dependent (MD) and non-methylation-dependent (NMD) groups (Figure 3B, C). Although there was no significant difference in gene expression between the MD and NMD categories (Fisher's exact test, P -value = 0.5), TE up-regulation was significantly enriched in the MD group (Fisher's exact test, P -value < 0.001) (Table S7). Notably, *Copia* elements were the most differentially regulated TEs in the *Gmdrm* mutant (Fisher's exact test, P -value < 0.001) (Figure 2D), whereas in our previous analysis of the *Gmcm1* mutant, *Gypsy* elements were more differentially regulated than *Copia* (Xun et al., 2024). This highlights the distinct types of TEs being regulated by the RdDM pathway and CMT in soybean. Additionally, we conducted a methylation analysis on the DETs within and surrounding TEs. We found that the observed differences in methylation distribution patterns at the boundaries of TEs between the WT and *Gmdrm* mutants are primarily driven by the loss of CHH methylation, which is a hallmark feature of DRM-mediated DNA methylation. (Figures 3E, S4). It indicates that TEs regulated by GmDRM-dependent CHH

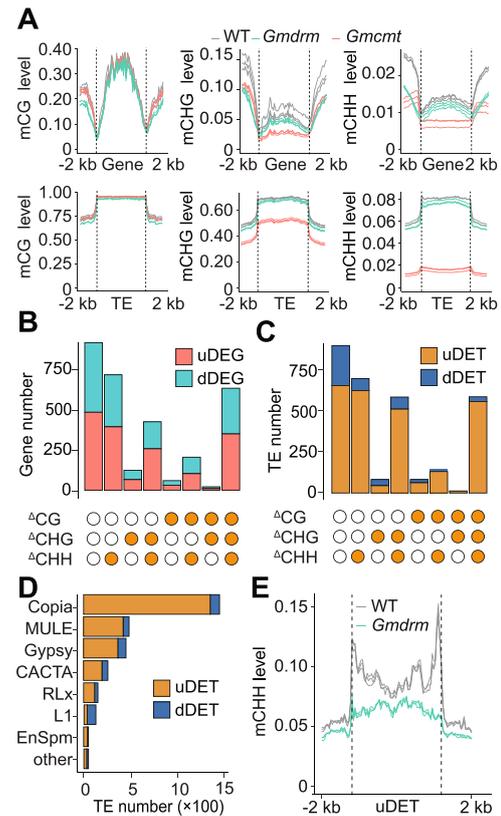


Figure 3. The effects on gene and transposable element (TE) expression in the *Gmdrm* and *Gmcm1* mutants

(A) Methylation landscapes within genes and TEs in *Gmdrm* and *Gmcm1* mutants compared to wild type (WT). This panel illustrates the distribution of mCG, mCHG, and mCHH methylation levels within and surrounding genes and TEs. **(B)** Correlation analysis between methylation status (mC) and differentially expressed genes (DEGs) in the *Gmdrm* mutant. uDEG and dDEG represent up-regulated and down-regulated genes, respectively. Filled circles indicate genes with hypomethylation, while open circles represent genes with no change in methylation status. **(C)** Correlation analysis between mC and differentially expressed TEs (DETs) in the *Gmdrm* mutant. uDET and dDET represent up-regulated and down-regulated TEs, respectively. Filled circles represent TEs with hypomethylation, while open circles represent no change in methylation status. **(D)** The number of TEs in each TE family that are up-regulated or down-regulated in the *Gmdrm* mutant. **(E)** Methylation landscapes of mCHH within up-regulated TEs in the *Gmdrm* mutant compared to WT.

methylation are more likely to become transcriptionally activated in the *Gmdrm* mutant, reflecting a loss of silencing at these regions. This finding reinforces that CHH methylation is a key determinant for GmDRM-mediated TE suppression, especially in euchromatic regions where RdDM activity is predominant.

To explore the activation of TEs and their potential impact on adjacent gene expression, we screened the DEGs surrounding TEs and found that 119 down-regulated TEs (dDTEs) were associated with down-regulated genes (dDEGs) (32.16% of total up-regulated DEGs (uDEGs)), and 255 up-regulated DTEs (uDTEs) were associated with uDEGs (15.5% of total uDEGs) (Table S8). We observed that TE activation was significantly associated with the up-regulation of

The function of GmDRMs in soybean

adjacent genes (Fisher's exact test P -value < 0.001), which supports the hypothesis that TE activation can lead to aberrant expression of adjacent genes.

Twenty-four nucleotide siRNAs are key determinants of DRM-dependent TE regulation

To understand the distinct features of DRM- and CMT-dependent TEs, we classified TEs into three categories: those regulated by *GmDRMs*, comprising 17,896 TEs (8.4%), *GmCMTs*, comprising 184,675 TEs (86.3%), and by both *GmCMT* and *GmDRM* (Both-TE), comprising 11,540 TEs (5.4%), based on their methylation states in the two mutants (Table S9). For each TE, we extracted sequence features, including AT/GC content, the three cytosine contexts, DNA methylation levels, and sRNA abundance in WT plants to evaluate which factors are key to distinguishing DRM-TE from CMT-TE. Using a random forest classification task, we calculated the Gini and accuracy decrease parameters to rank the relative importance of each feature (Figure 4A, B). Our analysis showed that when using all features together, the classification accuracy for distinguishing DRM-TEs from

CMT-TEs reached 0.92 (Figure 4C). The most important features were the abundance of 24-nt small interfering RNAs (siRNAs), AT content, and TE length. When using only 24-nt siRNA as the distinguishing feature, the classification accuracy remained high at 0.74, particularly for DRM-TEs, where accuracy reached 0.74. This highlights that 24-nt siRNA is a primary and direct factor for distinguishing DRM-TEs, consistent with a central role it plays in the RdDM pathway.

To further confirm the distinctions between CMT-TEs and DRM-TEs, we analyzed the distribution of 24-nt siRNAs, TE length, and AT content. We found that DRM-TEs typically exhibit higher 24-nt siRNA abundance, shorter lengths, and are more AT-rich compared to CMT-TEs (Figure 4D). Additionally, we investigated these features in *Copia* and *Gypsy* TEs. Most *Copia* elements followed the DRM-TEs pattern, while *Gypsy* elements resembled CMT-TEs (Figure S5). The differential activation of *Copia* and *Gypsy* elements in the *Gmdrm* and *Gmcm* mutants is likely influenced by both chromatin context and TE sequence features, particularly the GC versus AT content of these elements (Figure S5). This pattern is consistent with findings in Arabidopsis, where GC-rich TEs exhibit high levels of H3K9me2 and other heterochromatic marks, while AT-rich TEs are more euchromatic with lower repressive histone modifications (Zemach et al., 2013; Sequeira-Mendes et al., 2014). CMT3 mediates non-CG methylation of GC-rich heterochromatic TEs, whereas euchromatic TEs are regulated by the RdDM pathway, with DRM2 controlling CHH methylation. Additionally, Histone H1 restricts the binding of Pol IV to GC-rich TEs but allows it to bind to AT-rich DRM-TEs, leading to the generation of 24-nt siRNAs (Du et al., 2012; Stroud et al., 2014; Choi et al., 2021). This finding may explain why *Copia* elements tend to be activated in the *Gmdrm* mutant, whereas *Gypsy* elements are more commonly activated in the *Gmcm* mutant (Figure 3D).

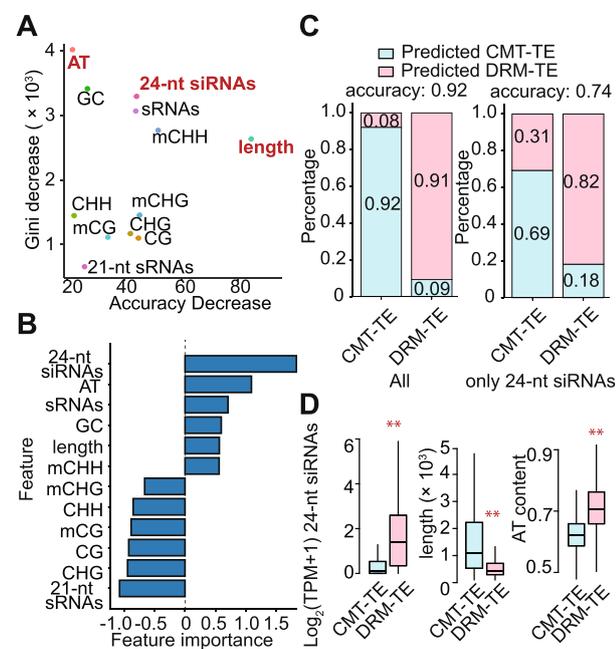


Figure 4. Key features distinguishing chromomethylase (CMT)- and domains rearranged methyltransferase (DRM)-dependent transposable elements (TEs)

(A) The importance of sequence features, DNA methylation, and small RNA (sRNA) abundance in predicting CMT-TE or DRM-TE classes using random forest classification. (B) Feature importance analysis for each factor. The x-axis represents the normalized importance, based on Gini and accuracy decrease scores, with higher values indicating greater importance of a given feature. (C) Prediction of CMT-TE or DRM-TE classes using random forest classification with either all variables or only 24-nt small interfering RNAs (siRNAs). (D) Comparison of 24-nt siRNA abundance, sequence length, and AT content between CMT-TE and DRM-TE classes. KS test was performed to assess statistical significance (** P -value < 0.01).

Disruption of GmDRMs affects 24-nt siRNA biosynthesis and demethylation pathways

As 24-nt siRNAs are crucial components of the RdDM pathway, we sought to investigate whether their biosynthesis is impacted by the loss of the RdDM pathway in the *Gmcm* mutant. It is well established that blocking 24-nt siRNA biosynthesis leads to a reduction in DNA methylation, particularly non-CG methylation (Stroud et al., 2014). To address this question, we performed sRNA sequencing in WT, *Gmcm*, and *Gmdrm* mutants. Across all samples, 21-nt and 24-nt sRNAs were the most abundant sRNA species (Figure 5A; Table S10). Interestingly, 24-nt siRNA levels were dramatically reduced in the *Gmdrm* mutant compared to WT, whereas *Gmcm* showed no significant change in overall sRNA abundance (Figure 5A). This suggests that sRNA production is not affected by GmCMTs but is strongly dependent on the RdDM pathway of GmDRMs.

To further explore the impact of DNA methylation on gene expression, we compared the expression levels of genes involved in the RdDM pathway in WT and the two mutants

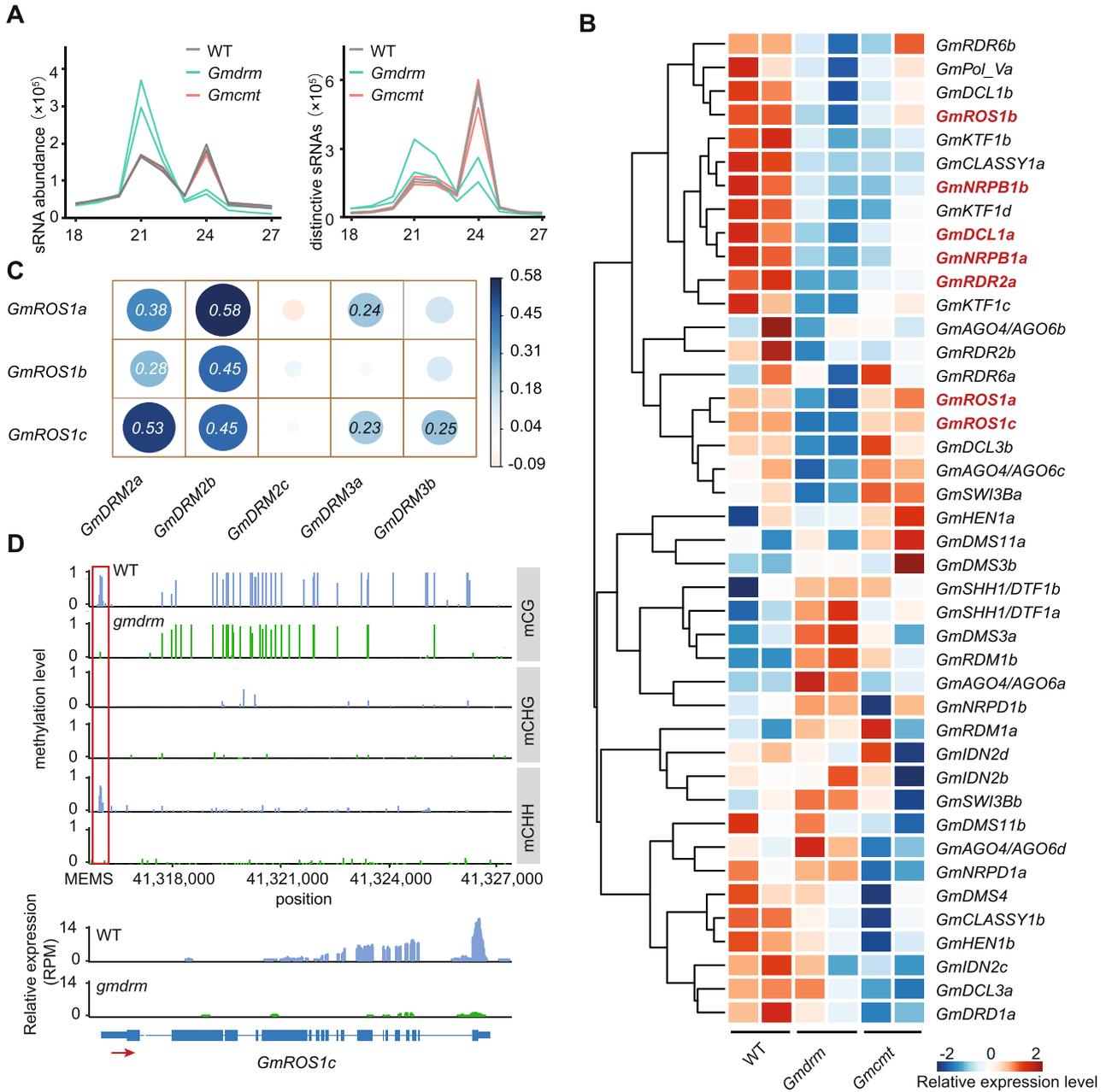


Figure 5. The effects on the expression of RNA-directed DNA methylation (RdDM)-related and DNA demethylation genes in *Gmcm1* and *Gmdrm* mutants compared to wild type (WT)

(A) Small RNA (sRNA) abundance and distribution of sRNAs of different lengths (18–27 nt) in WT, *Gmcm1*, and *Gmdrm* mutants. **(B)** Relative expression levels of genes involved in the RdDM pathway in WT, *Gmdrm*, and *Gmcm1* mutants. Heatmap colors are scaled for each row, with red gene names indicating significantly altered expression between WT and *Gmdrm* mutants. **(C)** Correlation analysis of the expression levels of five *GmDRM* genes and the three *GmROS1* genes using 164 public transcriptome data sets. The colors represent the correlation coefficient (r), and the circle size reflects the P -value (cor.test); larger circles indicate more significant correlations. The significant correlations are marked with r value on the circles. Each dot represents the expression level of the corresponding gene in one RNA sequencing sample. **(D)** Methylation levels and relative expression of *GmROS1c* in WT and *Gmdrm* mutants. The region in the red box represents the methylation monitoring sequence (MEMS) motif of *ROS1*.

(Figure 5B). We found that *GmNRPB1a/b* and *GmDCL1a* were down-regulated in both mutants, but this did not affect 24-nt siRNA levels in the *Gmcm1* background (Figure 5B). This indicates that *GmNRPB1a/b* and *GmDCL1* do not have a significant role in 24-nt siRNA biosynthesis in these mutants (Figure 5B). However, *GmRDR2a* was specifically down-regulated in the *Gmdrm* mutant, which may explain the

substantial reduction in 24-nt siRNA levels in this mutant (Figure 5B). Additionally, we observed that all three copies of *GmROS1* were down-regulated in the *Gmdrm* mutant. By analyzing public RNA-seq datasets (Table S11), *GmROS1* genes are indeed co-expressed with some of the *GmDRM* genes, particularly *GmDRM2a* and *GmDRM2b* (Figure 5C). This suggests that *GmDRMs* influence demethylation

regulated by *GmROS1s*. For example, we identified changes in CG and CHH methylation in the methylation monitoring sequence (MEMS) motif upstream of *GmROS1c*, which led to the down-regulation of *GmROS1c*, blocking demethylation (Figure 5D). Collectively, the loss of *GmDRMs* reduces DNA methylation, which in turn reduces *GmROS1*-mediated demethylation, indicating an antagonistic feedback loop between methylation and demethylation pathways.

Although 21-nt sRNAs showed no significant overall changes in the mutants, we did observe differential abundances of several micro RNAs (miRNAs) between the *Gmdrm* mutant and WT (Figure S6). For example, *gma-miR1510b-3p* and *gma-miR171* were down-regulated in the mutant, while TE-related miRNAs such as *miR408*, *miR4996*, and *miR5770* were up-regulated (Figure S6). This suggests that miRNAs are directly or indirectly affected by the RdDM pathway.

Up-regulation of photosynthesis and *R* genes in *Gmdrm* mutant lines

To investigate the biological processes affected in the *Gmdrm* mutant, we conducted a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis focusing on DEGs. This analysis revealed significant enrichment in pathways related to carbon fixation, hormone signal transduction, photosynthesis, and the photosynthetic antenna protein pathway (Figure 6A, B; Table S5). Previous studies suggest that optimizing chloroplast ultrastructure and enhancing photosynthesis rates, potentially driven by non-CG hypomethylation, can improve soybean yield, as seen in the *Gmcm1* mutant (Xun et al., 2024). Similarly, we observed significant up-regulation of expression for three GLK transcription factors (GmGLK4, GmGLK10, and GmGLK66) in the *Gmdrm* mutant (Figure S7). To explore potential ultrastructural changes in the *Gmdrm* mutant, we utilized scanning electron microscopy (SEM) for surface examination and transmission electron microscopy (TEM) for deeper analysis of cell structures. The *Gmdrm* mutant exhibited reduced cell size and increased wax accumulation on leaves (Student's *t*-test, *P*-value < 0.001) (Figures 6C, S8). Furthermore, we observed a significant increase in both the number and size of starch granules, as well as a greater number of granal thylakoid layers, compared to WT (Student's *t*-test, *P*-value < 0.001) (Figures 6D, S8). However, due to the smaller, shrunken leaves of the dwarf plants, direct measurements of starch accumulation and photosynthesis rates were not feasible in this study. To further understand hormonal changes in the *Gmdrm* mutant, we conducted a non-targeted metabolome analysis to assess the relative abundance of various phytohormones (Table S12). We found that several hormones, including abscisic acid (ABA), cytokinins (CK), and auxin, were altered between *Gmdrm* and WT (Table S11). The imbalance in phytohormone levels may contribute to the dwarf phenotype observed in the *Gmdrm* mutant (Figure 1D). Further, we performed Gene Ontology (GO) enrichment analysis for the DEGs between the *Gmdrm* mutant and WT. We found that the “response to auxin” term

was significantly enriched among the up-regulated genes. When integrating this with the DMR information, we identified that 27 out of 57 up-regulated genes exhibited changes in at least one context of DNA methylation, suggesting that these genes might be directly or indirectly regulated by DRM-dependent methylation (Tables S13, S14). For example, *Glyma.01G167000*, encoding a member of the SAUR-like auxin-responsive protein family, was regulated by hypomethylation across all three cytosine contexts (Table S14). Together, these results suggest that DRM-mediated DNA methylation influences phytohormone regulation and which in turn modulates important agronomic traits in soybean.

In addition, plants have evolved the capability to perceive pathogen effectors, either directly or indirectly, by utilizing proteins that are encoded by resistance (*R*) genes (Deng et al., 2017; Richard et al., 2018). Our analysis revealed significant enrichment of biological processes related to plant–pathogen interactions. Specifically, we observed up-regulation of an *R* gene cluster comprising 44 genes in the *Gmdrm* mutant (Figure 6B). To explore this further, we examined the methylation landscape surrounding these *R* genes and found a significant reduction in non-CG methylation across all gene regions, including gene bodies, promoters, and terminators. Importantly, CG methylation (mCG) levels within these regions remained relatively stable (Figure S3). This suggests that the reduction in non-CG methylation mediated by DRM may facilitate the up-regulation of *R* gene clusters (Figure S3). To assess the potential resistance to pathogens, we inoculated WT and *Gmdrm* mutant plants (excluding the quintuple mutant that is morphologically too malformed to enable reliable assaying) with soybean mosaic virus (SMV). Our screening identified a specific genotype, *Gmdrm2c/3a/3b*, which exhibited stronger resistance to SMV compared to the WT under greenhouse conditions (Figure S9), although further field evaluations are needed. These results suggest that DRM-mediated non-CG methylation may play a role in modulating soybean resistance to biotic stress.

DISCUSSION

In this study, we investigated the role of GmDRMs in soybean by generating a quintuple mutant using CRISPR/Cas9 technology. Our primary focus was to explore how RdDM influences DNA methylation, TE silencing, gene expression, and overall plant development. WGBS revealed a global reduction in DNA methylation levels in the *Gmdrm* mutant, affecting all cytosine content, CG, CHG, and CHH. This reduction in methylation was associated with widespread transcriptional activation of TEs, particularly in euchromatic regions. The observed reduction in 24-nt siRNAs in the *Gmdrm* mutant, highlights the importance of DRMs in RdDM-mediated siRNA biogenesis. The correlation between reduced non-CG methylation and changes in gene expression was also evident, with up-regulation of photosynthesis-related genes and an *R*

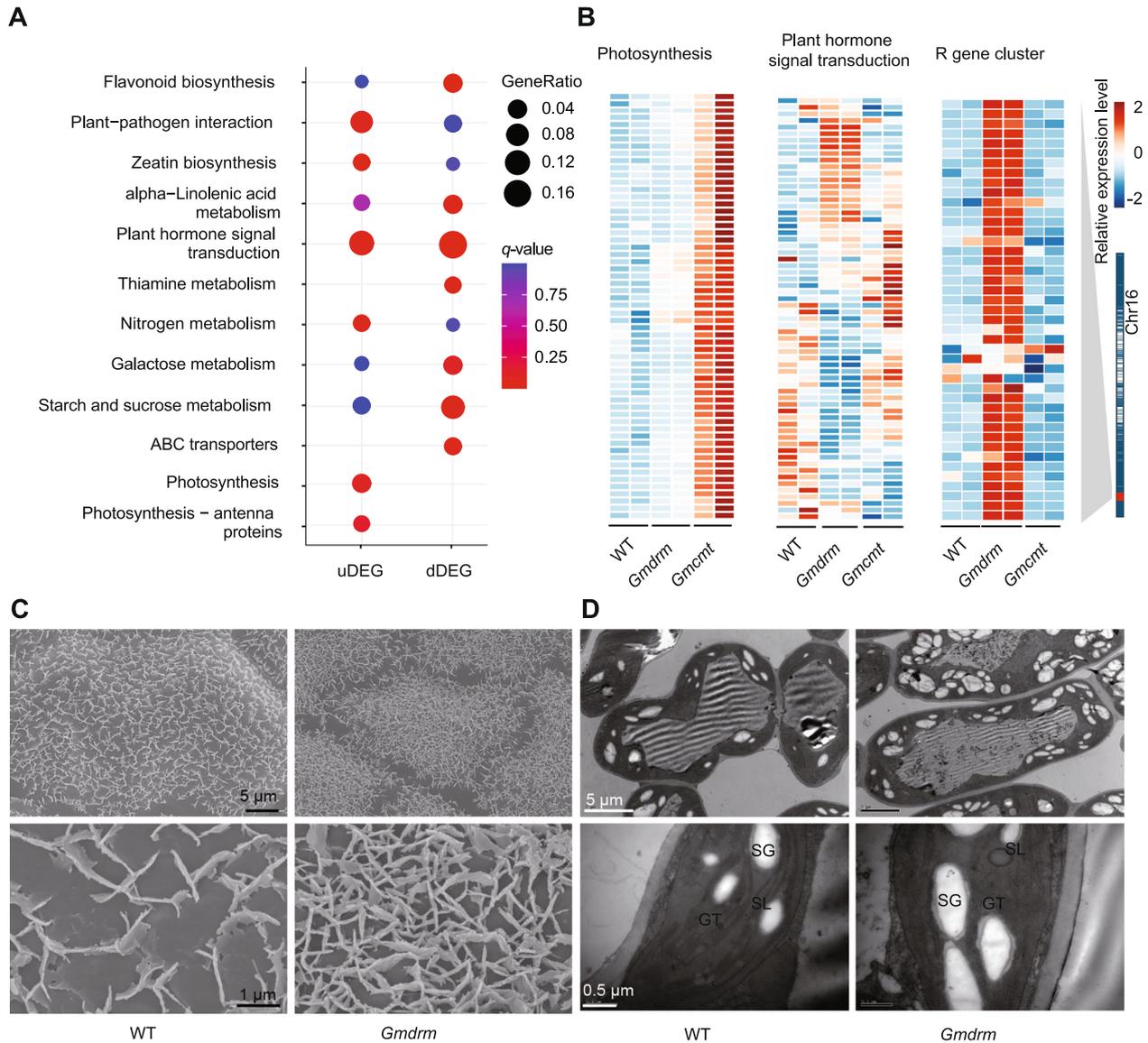


Figure 6. Impacts on biological processes related to the photosynthetic system and resistance (R) genes in the *Gmdrm* mutant

(A) Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis of up-regulated differentially expressed genes (uDEG) and down-regulated differentially expressed genes (dDEG) in the *Gmdrm* mutant. **(B)** Relative expression levels of DEGs involved in photosynthesis, plant hormone signal transduction, and *R* genes in the *Gmdrm* and *Gmcm1* mutants compared to wild type (WT). **(C)** Increased wax accumulation in the leaves of the *Gmdrm* mutant. Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) images show the upper leaf surface at the V2 stage, sampled at 08:00 hours. **(D)** Altered chloroplast ultrastructure in the leaves of the *Gmdrm* mutant. SEM and TEM images display the chloroplast ultrastructure of leaves at the V2 stage, sampled at 08:00 hours, highlighting granum thylakoids (GT), starch granules (SG), and stroma lamellae.

gene cluster. These findings underscore the pivotal role of DRM-mediated DNA methylation in maintaining epigenetic silencing of TEs and regulating key gene networks in soybean.

The *Gmdrm* quintuple mutant displayed noticeable developmental abnormalities, including dwarfism, growth retardation, and impaired reproductive capability in the T2 generation following self-crossing. Several factors may contribute to the fertility issues observed in the mutant lines. First, the loss of DRM-mediated methylation resulted in widespread TE transcriptional activation, which could lead to genome instability via ectopic recombination, and cause

aberrant expression of adjacent genes; both may lead to developmental defects over successive generations. As observed in *Setaria viridis*, where the *drm1a drm1b* double mutant could only sustain a homozygous state for a limited number of generations (Read et al., 2022), genome instability likely plays a role in the reduced fertility and gradual decline in plant health seen in the *Gmdrm* mutants. Second, the reduction in 24-nt siRNAs observed in the *Gmdrm* mutant could impede the proper establishment of DNA methylation marks required for reproductive development. In *Arabidopsis*, 24-nt siRNAs are essential for transmitting DNA methylation patterns to progeny, particularly in the regulation of TEs and gene imprinting during

seed development (Long et al., 2021). Similarly, in rapeseed, these siRNAs are transmitted from the maternal tissues to the progeny, helping to establish DNA methylation and ensure proper seed development (Grover et al., 2020). The disruption of this process in the *Gmdrm* mutant could explain the sterility observed after multiple rounds of self-fertilization. The absence or reduction of these siRNAs in the *Gmdrm* mutant likely compromises the reproductive cell's ability to maintain genome integrity, contributing further to the fertility issues in the mutant. We also found that the *Gmcm1* mutant did not affect the abundance of 24-nt siRNAs. This result is consistent with findings in rice in that the majority of 24-nt siRNAs are lost in the *Osdrm2* single mutant, but not in *Oscmt2* or *Oscmt3* single mutant (Hu et al., 2021), suggesting that DRM-dependent DNA methylation is necessary for the biogenesis of 24-nt siRNAs in both soybean and rice. A third potential cause of the fertility defects in the *Gmdrm* mutant is the disordered phytohormone regulation. In the *Gmdrm* mutant, we found significant alterations in the levels of key phytohormones, including gibberellin (GA) which was elevated, while auxin and jasmonic acid levels were significantly reduced. Gibberellins are known to promote growth and flowering, but elevated GA levels can also lead to abnormal development, particularly affecting reproductive tissues (Jiang et al., 2022). The imbalance between GA and auxin, which plays a key role in regulating meristem activity and floral organ formation, could disrupt the normal development of reproductive structures in the *Gmdrm* mutant.

Our findings have broadened our understanding of the roles played by *GmDRMs* in DNA methylation, gene expression, and plant development. The ability of GmDRM to regulate both photosynthetic efficiency and disease resistance suggests exciting opportunities for crop improvement through epigenetic modifications. By leveraging the distinct epigenetic mechanisms mediated by GmCMT and GmDRM, it may be possible to develop soybean varieties that exhibit enhanced productivity and resilience to environmental stressors. In Arabidopsis, Epi-RILs (epigenetic recombinant inbred lines) exhibited a wide range of phenotypic diversity, encompassing plant height, biomass, fruit size, fruit number, flowering time, and response to pathogen infection (Johannes et al., 2009; Reinders et al., 2009; Furci et al., 2019). In soybean, when *MSH1*-RNAi was crossed with WT plants, the F2 generation population exhibited a rich phenotypic diversity, encompassing traits such as pods per plant, seed number, 100-seed weight, flowering time (Raju et al., 2018). Given the promising results from these studies, future research in soybean should explore the potential of creating Epi-RILs by combining favorable traits from *Gmcm1*, *Gmdrm*, and WT plants.

In summary, our study highlights a central role of DRM-mediated DNA methylation in maintaining chromatin stability, regulating TE silencing, and controlling key gene networks in soybean. The full disruption of DRM function results in widespread changes in DNA methylation, leading to transcriptional activation of TEs, altered gene expression, and significant phenotypic abnormalities, including sterility and growth retardation. The findings of this study

demonstrate the complementary roles of DRM- and CMT-mediated methylation in regulating euchromatic and heterochromatic regions, respectively. The differential regulation of TEs and genes by these two pathways provides new insights into the complex interplay between DNA methylation, gene expression, and chromatin stability in plants. Collectively, these findings provide a foundation for further research into the role of DNA methylation in plant development and stress responses, with the ultimate goal of improving soybean yield and resistance to environmental challenges through epigenetic breeding strategies.

MATERIALS AND METHODS

Plant materials and growth conditions

Wild-type soybean of Williams 82 (Wms82) was used for plant transformation and as a control for all experiments. The *Gmdrm* quintuple mutants (*Gmdrm2a*^{-/-}*2b*^{-/-}*2c*^{-/-}*3a*^{-/-}*3b*^{-/-}) and partial mutants (*Gmdrm2c*^{-/-}*3a*^{-/-}*3b*^{-/-}, *Gmdrm2a*^{-/-}*2b*^{-/-}*2c*^{-/-}*3a*^{-/-}*3b*^{+/-}, and *Gmdrm2b*^{-/-}*2c*^{-/-}*3a*^{-/-}*3b*^{-/-}*2a*^{+/-}) were generated by CRISPR/Cas9 gene editing in the Wms82 background. Growth room conditions were set at 25°C and a 16-h light/8-h dark light cycle.

Phylogenetic analysis

To conduct phylogenetic analysis, we obtained the full-length amino acid sequences of all protein sequences from soybean, Arabidopsis, rice, and *Solanum lycopersicum* from Phytozome. The OrthoFinder program (v2.2.6) was employed to categorize ortholog groups for each family. Specifically, we focused on the *DRM2*, *DRM3* gene families in Arabidopsis and extracted the relevant gene tree from the generated database. The gene tree was constructed using the “One Click” mode on Phylogeny (www.phylogeny.fr).

Guide RNA design, vector construction, and gene cloning

For guide RNA (gRNA) design, vector construction, and gene cloning, the CRISPR-P tool (<http://crispr.hzau.edu.cn/CRISPR2/>), selecting targets with high target scores, low off-target rates, and appropriate positions, with U6 promoter and the enhanced CaMV 35S promoter were employed to drive gRNA and Cas9 expression (Xun et al., 2024). Three gRNAs were selected for vector construction: gRNA-1: targeting GmDRM2a and GmDRM2b, sequence: CTGAGAATGATACCTTACGG. gRNA-2: targeting GmDRM3a, sequence: CTGTCAAAGCCTCCTCAAAGGG. gRNA-3: targeting GmDRM3b and GmDRM2c, sequence: TGAAGCATCAATTCCTGAGCTGG.

Soybean transformation, genotyping, and transgenic plant screening

We used the *Agrobacterium tumefaciens*-mediated cotyledonary node transformation method and the GMO DETECT kit (specific for bar/pat genes, Artron Laboratory Inc., Beijing) to identify transgenic plants. PCR amplification of *GmDRM* gene fragments was performed, followed by sequencing to

assess the status of gene editing (Xun et al., 2024). A list of all PCR primers used can be found in Table S15.

Measurement of plant hormones content

The inoculated leaflets of WT and the mutant plants were collected. Three biological replicates were performed and the averages of these three replicates of both *Gmdrm* and WT were compared. Endogenous plant hormones were extracted by using a reverse-phase column (C18; Kromasil). Then, liquid chromatography – tandem mass spectrometry (LC-MS/MS) was performed by coupling a Dionex Ultimate 3000 UPLC system (Thermo Fisher Scientific) to a TSQ Altis triple-quadrupole mass spectrometer (Thermo Fisher Scientific, TSQ Altis™, USA). Metabolite separation was performed on a Waters Acquity HSS T3 column (100 mm × 2.1 mm, 1.8 μm) at a flow rate of 0.40 mL/min using a 10 min reversed-phase gradient: starting from 2% of solvent B (solvent A: 0.04% acetic acid in high-performance LC (HPLC) grade water; solvent B: 0.04% acetic acid in HPLC grade acetonitrile) and gradually changing to 98% B over 6 min, 98% of B for 2 min, return to 2% B over 0.1 min, 2% of B for 1.9 min. The column temperature was held at 40°C. The instrument was operated in selected/multiple reaction monitoring (SRM/MRM) mode. Phytohormones were detected by optimized MRM transitions.

Whole genome bisulfite sequencing and data processing

Genomic DNA was extracted from second trifoliolate leaves of WT and *Gmdrm* mutant plants at 21 d old. Three biological replicates were performed for *Gmdrm* and WT samples. And the WGBS and data processing were conducted as previously described (Xun et al., 2024). Briefly, for library preparation, 0.8 μg of genomic DNA was sonicated using a Covaris S220 system to generate fragments of approximately 500 bp. These fragments were then bisulfite-converted using the EZ DNA Methylation-Lightning Kit (ZYMO, Irvine, CA, USA). End-repair, 5' phosphorylation, and dA-tailing were performed in a single reaction using Vazyme (Nanjing, China) dA-Tailing Enzyme Mix (NR602-02-AI), followed by T-A ligation of methylated adapters with the VAHTS Universal Adapter Ligation Module (N204-02; Vazyme). Size selection of adaptor-ligated DNA fragments (~410 bp) was conducted using VAHTS DNA Clean Beads (Vazyme). The constructed libraries were sequenced on the Illumina X Plus platform. Quality control and sequence cleaning were performed using Trimmomatic to remove low-quality and contaminated reads. Clean reads were aligned to the bisulfite-converted reference genome (Gmax v4; Phytozome) using Bismark (v0.20.0). DNA methylation patterns in CG, CHG, and CHH contexts were analyzed and visualized using ViewBS and deeptools. Differentially methylated regions (DMRs) were identified with CGmappro, with the following cutoff criteria: an adjusted *P*-value (*q*-value) < 0.01 and a methylation level difference greater than 0.1. To classify TEs as DRM-TE, CMT-TE, or Both-TE, we determined these categories computationally based on methylation changes observed in the WT, *Gmcm*t,

and *Gmdrm* mutants. Specifically, we first calculated the number of methylated and unmethylated cytosines across the three cytosine contexts (CG, CHG, and CHH) in each of the mutants. We then applied Fisher's exact test and *P*-value adjustment (false discovery rate) method, selecting differential TEs with a *q*-value less than 0.05 and a TE length greater than 200 bp. TEs showing hypomethylation only in the *Gmcm*t mutant were classified as CMT-TEs. TEs showing hypomethylation only in the *Gmdrm* mutant were classified as DRM-TEs. TEs showing hypomethylation in both mutants were classified as Both-TEs.

Transcriptome sequencing (RNA-seq) and small RNA sequencing data processing

RNA was extracted from the second trifoliolate leaves of WT and *Gmdrm* mutant plants at 21 d old, with two biological replicates. RNA sequencing was performed on the HiSeq. 2500 platform (Illumina, San Diego, CA, USA). Sequence cleaning was performed using Trimmomatic to remove sequencing adaptors and low-quality reads. Clean data were aligned to the soybean reference genome (Gmax v4; Phytozome) using Hisat2 (v2.1.0) with default settings. Alignment files (SAM) were processed using SAMtools with a parameter (“-q 30”) to remove multi-mapped reads, and converted to BAM format. Gene expression levels were quantified using transcripts per million (TPM) values via the EdgeR program. Differentially expressed genes were identified using EdgeR, with *q*-value ≤ 0.05 and fold change ≥ 2 as thresholds. Over-represented GO terms and KEGG pathways were identified using a custom Perl script based on soybean annotations (Gmax v4), followed by a hypergeometric test with the R package clusterProfiler.

For sRNA sequencing (sRNA-seq), libraries were constructed using the NEBNext Multiplex Small RNA Library Prep Set (NEB) and sequenced on the Illumina HiSeq. 2500 platform to generate 50-bp single-end reads. Raw reads were processed using the fastx-toolkit to remove low-quality reads and adapter sequences. The processed reads were aligned to the soybean and common bean reference genomes (v12.1; Phytozome) using ShortStack (v4.0.0). Shortstack employs an advanced method called the local-weighting method to handle multiple alignments of sRNAs (Johnson et al., 2016). Small RNA abundance was normalized to counts per million (cpm) based on mapping results using in-house Perl scripts. While we acknowledge that there appears to be slightly lower reproducibility among the biological replicates in some cases, we also calculated the reproducibility of length distribution, which yielded a high correlation of ~0.99. Since our primary focus is on the distribution and abundance of 24-nt siRNAs, we believe that the observed variation does not significantly impact the reliability of our conclusions.

Random forest classification and prediction

To assess the importance of each sequence feature and epigenetic marker for classifying DRM- and CMT-regulated TEs, we first calculated the average enrichment of sRNA

expression and DNA methylation levels for each TE using custom Perl scripts. We also included the density of various cytosine sequence contexts and AT/GC contents. The significance of each variable was evaluated using the “randomForest” and “measure_importance” functions in the RandomForestExplainer R package (Ishwaran et al., 2012). The importance matrices were visualized using the “plot_multi_way_importance” function from the same package. To evaluate the predictive power of each variable, TEs were randomly divided into training and validation sets. A random forest classifier was built using the training set, with the specified variables and TE classification (DRM-TE or CMT-TE). The trained model was then applied to the validation set to predict TE categories. Accuracy was calculated by comparing the predicted classifications with the actual classifications. We used the “randomForest” and “predict” functions from the randomForest R package to perform this analysis. The RandomForestClassifier was implemented using its default parameters, with the exception of $n_estimators = 100$. The complete list of default parameters for the RandomForestClassifier is as follows: $n_estimators$: 100, criterion: “gini,” $min_samples_split$: 2, $min_samples_leaf$: 1, $min_weight_fraction_leaf$: 0.0.

Scanning electron microscopy and TEM

The morphology of WT and Gmdrm mutant leaves was observed using field emission SEM (FE-SEM; SU8010; Hitachi, Japan) and TEM (H-7650; Hitachi, Japan). The subsequent operation procedures were conducted the same as previously described (Xun et al., 2024).

Quantitative RT-PCR

Total RNA from Gmdrm mutant and WT plants was extracted using EasyPure Plant RNA Kit (Transgen Biotech, Beijing, China). Subsequently, complementary DNA (cDNA) was synthesized using the ThermoScript RT-PCR System (Invitrogen, Carlsbad, CA, USA), and both kits were operated according to the manufacturer's instructions. Quantitative RT-PCR was conducted the same as previously described (Xun et al., 2024). A comprehensive list of qRT-PCR primers used can be found in Table S15.

Soybean mosaic virus resistance assay

The Gmdrm mutant and Williams 82 were grown together in a 25°C greenhouse with a light period of 16 h and a dark period of 8 h. Fourteen-d-old plants were used to examine virus resistance and inoculated with a mixture of SMV SC1 and SMV SC3 and 0.01 mol/L phosphate solution as a mock control treatment. The phenotype was observed 21 d post-inoculation. Newly emerged leaves were then collected for RNA extraction, followed by qRT-PCR to detect the expression level of the CP gene in the leaves.

Data availability statement

The raw WGBS, RNA-seq and sRNA-seq read sequences produced in this study have been submitted to the National Genomics Data Center (<https://ngdc.cncb.ac.cn/>) and can be accessed under the accession number PRJCA021234.

ACKNOWLEDGEMENTS

This work was mainly supported by the National Key Research and Development Program of China (2022YFF1003303, 2022YFD1201400); the National Natural Science Foundation of China (32301796, 32272064); the Fundamental Research Funds for the Central Universities (2412023QD021, 2412023YQ005); the China Postdoctoral Science Foundation grant (2019M651236). These funds played an important role in experimental reagents, consumables, and also provided strong support for preparing the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interests.

AUTHOR CONTRIBUTIONS

X.W., B.L., and H.X. conceived and designed the experiments. H.X., L.L., J.Y., J.H., S.H., H.Z., S.L., W.F., H.Y. performed the experiments. X.W., H.X., and W.F. analyzed the data. J.Y. performed the electronic microscope image analysis. X.W., B.L., and H.X. wrote the manuscript. All authors read and approved the contents of this paper.

Edited by: Zhaobo Lang, Southern University of Science and Technology, China

Received Sep. 29, 2024; **Accepted** Feb. 13, 2025

REFERENCES

- Cao, X., Aufsatz, W., Aufsatz, D., Mette, M., Huang, M.S., Matzke, M., and Jacobsen, S.E. (2003). Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr. Biol.* **13**: 2212–2217.
- Cao, X., and Jacobsen, S.E. (2002a). Role of the *Arabidopsis* DRM methyltransferases in *de novo* DNA methylation and gene silencing. *Curr. Biol.* **12**: 1138–1144.
- Cao, X., and Jacobsen, S.E. (2002b). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proc. Natl. Acad. Sci. U.S.A.* **99**: 16491–16498.
- Choi, J., Lyons, D.B., and Zilberman, D. (2021). Histone H1 prevents non-CG methylation-mediated small RNA biogenesis in *Arabidopsis* heterochromatin. *eLife* **10**: e72676.
- Costa-Nunes, P., Kim, J.Y., Hong, E., and Pontes, O. (2014). The cytological and molecular role of domains rearranged methyltransferase3 in RNA dependent DNA methylation of *Arabidopsis thaliana*. *BMC Res. Notes* **7**: 721.
- Deng, Y., Zhai, K., Xie, Z., Yang, D., Zhu, X., Liu, J., Wang, X., Qin, P., Yang, Y., Zhang, G., et al. (2017). Epigenetic regulation of antagonistic receptors confers rice blast resistance with yield balance. *Science* **355**: 962–965.
- Du, J., Zhong, X., Bernatavichute, Y.V., Stroud, H., Feng, S., Caro, E., Vashisht, A.A., Terragni, J., Chin, H.G., Tu, A., et al. (2012). Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* **151**: 167–180.

- Erdmann, R.M., and Picard, C.L. (2020). RNA-directed DNA methylation. *PLoS Genet.* **16**: e1009034.
- Furci, L., Jain, R., Stassen, J., Berkowitz, O., Whelan, J., Roquis, D., Baillet, V., Colot, V., Johannes, F., and Ton, J. (2019). Identification and characterisation of hypomethylated DNA loci controlling quantitative resistance in *Arabidopsis*. *eLife* **8**: e40655.
- Garcia-Aguilar, M., Michaud, C., Leblanc, O., and Grimanelli, D. (2010). Inactivation of a DNA methylation pathway in maize reproductive organs results in apomixis-like phenotypes. *Plant Cell* **22**: 3249–3267.
- Grover, J.W., Burgess, D., Kendall, T., Baten, A., Pokhrel, S., King, G.J., Meyers, B.C., Freeling, M., and Mosher, R.A. (2020). Abundant expression of maternal siRNAs is a conserved feature of seed development. *Proc. Natl. Acad. Sci. U.S.A.* **117**: 15305–15315.
- Henderson, I.R., Deleris, A., Wong, W., Zhong, X., Chin, H.-G., Horwitz, G.-A., Kelly, K.A., Pradhan, S., and Jacobsen, S.E. (2010). The de novo cytosine methyltransferase DRM2 requires intact UBA domains and a catalytically mutated paralog DRM3 during RNA-directed DNA methylation in *Arabidopsis thaliana*. *PLoS Genet.* **6**: e1001182.
- Henderson, I.R., and Jacobsen, S.E. (2007). Epigenetic inheritance in plants. *Nature* **447**: 418–424.
- Hu, D., Yu, Y., Wang, C., Long, Y., Liu, Y., Feng, L., Lu, D., Liu, B., Jia, J., Xia, R., et al. (2021). Multiplex CRISPR-Cas9 editing of DNA methyltransferases in rice uncovers a class of non-CG methylation specific for GC-rich regions. *Plant Cell* **33**: 2950–2964.
- Ishwaran, H., Kogalur, U.B., Gorodeski, E.Z., Minn, A.J., and Lauer, M.S. (2012). High-dimensional variable selection for survival data. *J. Am. Stat. Assoc.* **105**: 205–217.
- Jiang, K., Guo, H., and Zhai, J. (2022). Interplay of phytohormones and epigenetic regulation: A recipe for plant development and plasticity. *J. Integr. Plant Biol.* **65**: 381–398.
- 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.
- Johannes, F., Porcher, E., Teixeira, F.K., Saliba-Colombani, V., Simon, M., Agier, N., Bulski, A., Albuissou, J., Heredia, F., Audigier, P., et al. (2009). Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* **5**: e1000530.
- Lindroth, A.M., Cao, X., Jackson, J.P., Zilberman, D., McCallum, C.M., Henikoff, S., and Jacobsen, S.E. (2001). Requirement of CHROMO-METHY LASE3 for maintenance of CpXpG methylation. *Science* **292**: 2077–2080.
- Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., Zhou, G.A., Zhang, H., Liu, Z., Shi, M., et al. (2020). Pan-genome of wild and cultivated soybeans. *Cell* **182**: 162–176.
- Long, J., Walker, J., She, W., Aldridge, B., Gao, H., Deans, S., Vickers, M., and Feng, X. (2021). Nurse cell-derived small RNAs define paternal epigenetic inheritance in *Arabidopsis*. *Science* **373**: eabh0556.
- Moritoh, S., Eun, C.H., Ono, A., Asao, H., Okano, Y., Yamaguchi, K., Shimatani, Z., Koizumi, A., and Terada, R. (2012). Targeted disruption of an orthologue of DOMAINS REARRANGED METHYLASE 2, OsDRM2, impairs the growth of rice plants by abnormal DNA methylation. *Plant J.* **71**: 85–98.
- Raju, S.K.K., Shao, M., Sanchez, R., Xu, Y., Sandhu, A., Graef, G., and Mackenzie, S. (2018). An epigenetic breeding system in soybean for increased yield and stability. *Plant Biotechnol. J.* **16**: 1836–1847.
- Read, A., Weiss, T., Crisp, P.A., Zhikai, Z.K., Noshay, J., Menard, C.C., Wang, C., Song, M., Hirsch, C.N., Springer, N.M., et al. (2022). Genome-wide loss of CHH methylation with limited transcriptome changes in *Setaria viridis* DOMAINS REARRANGED METHYLTRANSFERASE (DRM) mutants. *Plant J.* **111**: 103–116.
- Reinders, J., Wulff, B.B.H., Mirouze, M., Mari-Ordóñez, A., Dapp, M., Rozhon, W., Bucher, E., Theiler, G., and Paszkowski, J. (2009). Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes. *Genes Dev.* **23**: 939–950.
- Richard, M.M.S., Gratias, A., Thareau, V., Kim, K.D., Balzergue, S., Joets, J., Jackson, S.A., and Geffroy, V. (2018). Genomic and epigenomic immunity in common bean: the unusual features of NB-LRR gene family. *DNA Res.* **25**: 161–172.
- Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, H., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* **463**: 178–183.
- Sequeira-Mendes, J., Aragüez, I., Peiró, R., Mendez-Giraldez, R., Zhang, X., Jacobsen, S.E., Bastolla, U., and Gutierrez, C. (2014). The functional topography of the *Arabidopsis* genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell* **26**: 2351–2366.
- 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.
- 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.
- Tran, R.K., Zilberman, D., de Bustos, C., Ditt, R.F., Henikoff, J.G., Lindroth, A.M., Delrow, J., Boyle, T., Kwong, S., Bryson, T.D., et al. (2005). Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in *Arabidopsis*. *Genome Biol.* **6**: R90.
- Xun, H., Wang, Y., Yuan, J., Lian, L., Feng, W., Liu, S., Hong, J., Liu, B., Ma, J., and Wang, X. (2024). Non-CG DNA hypomethylation promotes photosynthesis and nitrogen fixation in soybean. *Proc. Natl. Acad. Sci. U.S.A.* **121**: e2402946121.
- Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D. (2013). The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**: 193–205.
- Zhong, X., Hale, C.J., Nguyen, M., Ausin, I., Groth, M., Hetzel, J., Vashisht, A.A., Henderson, I.R., Wohlschlegel, J.A., and Jacobsen, S.E. (2015). Domains rearranged methyltransferase3 controls DNA methylation and regulates RNA polymerase V transcript abundance in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **112**: 911–916.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article: <http://onlinelibrary.wiley.com/doi/10.1111/jipb.13883/supinfo>

Figure S1. Visualization of Sanger sequencing results for five genes in the *Gmdrm* mutant

Figure S2. The phenotype of seed size in the *Gmdrm* mutant and wild type (WT)

Figure S3. Comparison of phenotypic, DNA methylation levels, and types of gene editing between different *Gmdrm* mutants compared to wild-type (WT) plants

Figure S4. Analysis of methylation landscape surrounding resistance (*R*) genes and differentially expressed transposable elements (DETs)

Figure S5. Analysis of sequence pattern between *Copia* and *Gypsy* in soybean

Figure S6. Relative expression levels of micro RNAs (miRNAs) in wild-type (WT), *Gmdrm*, and *Gmcm1* mutants

Figure S7. Relative expression levels of four *GLK* transcription factors in the *Gmdrm* mutant compared to wild type (WT)

Figure S8. Statistical analysis of cell structures on leaves between *Gmdrm* mutant and wild type (WT)

Figure S9. The *Gmdrm2c/3a/3b* variants exhibited enhanced resistance to soybean mosaic virus (SMV) under greenhouse inoculation conditions

Table S1. Different methylation levels of chromosomes in chromatin between the *Gmdrm* mutant, the *Gmcm1* mutant and wild type (WT)

Table S2. The statistical information of DNA methylation change between mutants and wild type (WT)

Table S3. Differentially methylated regions (DMR) information in *Gmdrm* mutant compared with wild-type

Table S4. Differentially methylated regions (DMR) information in *Gmcmt* mutant compared with wild-type

Table S5. Differentially expressed genes among the *Gmdrm*, *Gmcmt* mutants and wild type (WT) and their annotation information

Table S6. Differentially expressed transposable elements (TEs) among the *Gmdrm*, *Gmcmt* mutants and wild type (WT)

Table S7. Statistical overview of differentially expressed genes and transposable elements (TEs) in methylation-dependent (MD) and non-methylation-dependent (NMD) groups

Table S8. The differential expression status of differentially expressed genes (DEGs) with adjacent differentially expressed transposable elements (DTEs)

Table S9. The information of sequence, small RNA (sRNA), and DNA methylation features and transposable element (TE) categories in mutants and wild type (WT)

Table S10. The abundance of small RNAs (sRNAs) among the two mutants and wild type (WT)

Table S11. The expression levels of *GmDRMs* and *GmROS1a/b/c* genes were calculated using FPKM (fragments per kilobase of exon model per million mapped reads) in various tissues

Table S12. The average content of hormones in *Gmcmt*, *Gmdrm* and wild type (WT)

Table S13. The Gene Ontology (GO) enrichment analysis of differentially expressed genes between *Gmdrm* and wild type (WT)

Table S14. The differential DNA methylation information of up-regulated genes related to auxin response

Table S15. Main primers used in this study



Scan the QR code to view JIPB on WeChat (WeChat: [jipb1952](#))



Scan the QR code to view JIPB on Twitter (Twitter: [@JIPBio](#))