the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2024) doi: 10.1111/tpj.16744

RESOURCE

Genome-wide methylation landscape during somatic embryogenesis in *Medicago truncatula* reveals correlation between *Tnt1* retrotransposition and hyperactive methylation regions

Raja Sekhar Nandety[1,2,3,*], Sunhee Oh[1], Hee-Kyung Lee[1], Nick Krom[1], Rajeev Gupta[2,3] and Kirankumar S. Mysore[1,4,5,*]

[1]*Noble Research Institute, Ardmore, Oklahoma 73401, USA,*

[2]*Department of Plant Sciences, North Dakota State University, Fargo, North Dakota 58102, USA,*

[3]*Cereal Crops Research Unit, USDA-ARS, Edward T. Schafer Agricultural Research Center, Fargo, North Dakota 58102, USA,*

[4]*Institute for Agricultural Biosciences, Oklahoma State University, Ardmore, Oklahoma 73401, USA, and*

[5]*Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater, Oklahoma 74078, USA*

Received 26 June 2023; revised 27 February 2024; accepted 14 March 2024.

*For correspondence (e-mail kmysore@okstate.edu; raja.nandety@usda.gov).

SUMMARY

Medicago truncatula **is a model legume for fundamental research on legume biology and symbiotic nitrogen fixation.** ***Tnt1*****, a retrotransposon from tobacco, was used to generate insertion mutants in** ***M. truncatula*** **R108. Approximately 21 000 insertion lines have been generated and publicly available.** ***Tnt1*** **retro-transposition event occurs during somatic embryogenesis (SE), a pivotal process that triggers massive methylation changes. We studied the SE of** ***M. truncatula*** **R108 using leaf explants and explored the dynamic shifts in the methylation landscape from leaf explants to callus formation and finally embryogenesis. Higher cytosine methylation in all three contexts of CG, CHG, and CHH patterns was observed during SE compared to the controls. Higher methylation patterns were observed in assumed promoter regions (~2-kb upstream regions of transcription start site) of the genes, while lowest was recorded in the untranslated regions. Differentially methylated promoter region analysis showed a higher CHH methylation in embryogenesis tissue samples when compared to CG and CHG methylation. Strong correlation (89.71%) was identified between the differentially methylated regions (DMRs) and the site of** ***Tnt1*** **insertions in** ***M. truncatula*** **R108 and stronger hypermethylation of genes correlated with higher number of** ***Tnt1*** **insertions in all contexts of CG, CHG, and CHH methylation. Gene ontology enrichment and KEGG pathway enrichment analysis identified genes and pathways enriched in the signal peptide processing, ATP hydrolysis, RNA polymerase activity, transport, secondary metabolites, and nitrogen metabolism pathways. Combined gene expression analysis and methylation profiling showed an inverse relationship between methylation in the DMRs (regions spanning genes) and the expression of genes. Our results show that a dynamic shift in methylation happens during the SE process in the context of CG, CHH and CHG methylation, and the** ***Tnt1*** **retrotransposition correlates with the hyperactive methylation regions.**

Keywords: legumes, ***Medicago*****, methylation, somatic embryogenesis, tissue culture,** ***Tnt1*****, transposon.**

INTRODUCTION

Legumes play a key role in the sustainable agriculture to fix atmospheric nitrogen through a mutualistic symbiotic process with nitrogen-fixing bacteria, rhizobia (Gepts et al., 2005; Lee et al., 2018; Sun et al., 2019; Young & Udvardi, 2009). *Medicago truncatula* has a relatively small diploid genome of ~400 Mb (Barker et al., 1990; Kaur et al., 2021), high levels of genetic diversity, and has synteny with other legumes with much larger and more complex genomes (Burks et al., 2018). Therefore, *M. truncatula* with two distinct ecotypes, Jemalong A17 and R108 are used as models to study legume biology (Garmier et al., 2017). Currently,

the genome information is available for *M. truncatula* assembly A17 v5.0 (https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR) (Pecrix et al., 2018) and R108, generated by Hi-C technique that assisted in arranging the contigs into a more accurate assembly (MedtrR108_hic) (Kaur et al., 2021; Li et al., 2022; Moll et al., 2017).

Insertional mutagenesis is a method of disrupting gene function by introducing a foreign DNA fragment (T-DNA, transposon, or retrotransposon) into the genome. Tobacco *Tnt1*, a well-characterized plant autonomous long terminal repeat (LTR) retrotransposon (Grandbastien, 1998; Grandbastien et al., 1989), was used in *M. truncatula* R108 ecotype to generate 21 741 *Tnt1* insertion lines through somatic embryogenesis (SE) via tissue culture (d'Erfurth et al., 2003; Lee et al., 2018; Tadege et al., 2008). The overall distribution of *Tnt1* insertions in the *M. truncatula* genome is random and the distribution of *Tnt1* insertions varies in individual lines (Sun et al., 2019). Earlier estimations placed the average number of insertions per line at 25 based on Southern blot analysis and also through recovery of insertions by TAIL-PCR (Cheng et al., 2014, 2017; Tadege et al., 2005, 2008). A more recent analysis using whole genome sequencing and sequence capture approaches revealed a significantly greater number of *Tnt1* insertions (an average of ~80 per line) in *M. truncatula* R108 (Sun et al., 2019). Based on the publicly available limited methylation data on *M. truncatula* A17, Sun et al. (2019) suggested that the *Tnt1* insertion was more frequent in gene groups that are potentially more methylated compared to other gene groups.

Covalent modification by methylation of cytosine residues represents an important epigenetic hallmark (Baubec et al., 2009). In plants, *de novo* methylation is catalyzed by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), a homolog of the DNA METHYL TRANSFERASE 3 (DNMT3), and maintained by three different pathways: CG methylation is maintained by DNA METHYLTRANSFERASE 1 (MET1, also known as DMT1), the plant homolog of DNMT1; CHG methylation is maintained by CHROMOMETHYLASE 3 (CMT3), a plant-specific DNA methyltransferase; and asymmetric CHH methylation is maintained through persistent *de novo* methylation by DRM2 (Law & Jacobsen, 2010). In all the three contexts of cytosine methylation, CG, CHH, and CHG, "H" can be any of the nucleotides "A, T, or C" (Ji et al., 2019; Matzke et al., 2015; Matzke & Mosher, 2014). In Arabidopsis, CHH methylation is maintained by CHROMOMETHYLASE 2 (CMT2), which recognizes the methylation of Lys-9 on histone H3 (Stroud et al., 2014; Wollmann et al., 2017).

Somatic embryogenesis, results in the formation of an embryo from somatic cells (Mendez-Hernandez et al., 2019). Primary embryogenesis uses explants to induce the embryogenesis and secondary embryogenesis uses the existing embryos instead of explants (Gyulai et al., 1993; Tian et al., 2020). It is not clear as to how the cells initiate the embryo formation, but research suggests that an asymmetric distribution of auxins from differential transport leads to the generation of different tissues that will form the embryo (Mendez-Hernandez et al., 2019). SE requires the regulation of several genes with the involvement of transcription factors (Mendez-Hernandez et al., 2019; Tian et al., 2020). Some of the prominent proteins involved in SE include SOMATIC EMBRYOGENIC RECEPTOR KINASES (SERKs) and WUSCHEL (WUS) along with other proteins involved in signal transduction pathways and epigenetic mechanisms during chromatin remodeling (Ji et al., 2019; Mendez-Hernandez et al., 2019; Tian et al., 2020). It has been observed that embryo maturation during SE was marked with gradual increase in the global levels of methylation in *Pinus radiata* (Bravo et al., 2017) and coffee (*Coffee canephora*) plants (Nic-Can et al., 2013) and reviewed by Mendez-Hernandez et al. (2019). The correlative methylation studies during SE in coffee and *P. radiata* identified a crosstalk between methylation and SE (Bravo et al., 2017; Nic-Can et al., 2013). Previously, we observed that lowly expressed genes had more *Tnt1* insertions whereas highly expressed genes had less *Tnt1* insertions and suggested that *Tnt1* insertion frequency in *M. truncatula* R108 positively correlates with the methylation frequency in gene groups (Sun et al., 2019). However, this suggestion was done based on methylation studies in *M. truncatula* A17 somatic and nodule tissues.

In this study, we performed a comprehensive genome-wide investigation of the epigenome during SE in *M. truncatula* R108. We characterized *M. truncatula* epigenome for samples collected during the SE process at four different stages ranging from callus to embryo formation. The results revealed a genome-wide increase in DNA methylation in the differentially methylated promoters (DMPs), which was most apparent for CHH methylation. The correlation of the CHH methylation and the symmetric methylation (CG and CHG) with the *Tnt1* retrotranspositions in the *M. truncatula* R108 genome reveals an active RNA-directed DNA methylation (RdDM) process that dynamically regulates the cytosine methylation in all contexts along with the CMT2 dependent CHH methylation.

RESULTS

Methylation landscape in *M. truncatula* R108

To investigate the overall methylation patterns during tissue culture stress in *M. truncatula* R108 genome, we studied the methylation patterns during callus formation and SE stage. The sterilized leaf explants from the wild-type R108 *M. truncatula* plants were incubated in modified SH (Schenk & Hildebrandt, 1972) or SHMab medium and placed in a growth chamber at 24°C for callus induction as described in the "Materials and Methods" section (Figure S1). The explants that showed initial stages of callus induction were designated as "Before Callus" (BC). The

Table 1 Bisulfite sequencing summary

Samples	Total reads	Mapped reads	Unique mapping rate (%)	Duplication rate[a] (%)	Bisulfite conversion rate (%)
Control-1	49 987 419	38 379 942	76.78	9.35	99.84
Control-2	51 817 441	39 526 720	76.28	9.92	99.875
Control-3	58 706 655	45 109 017	76.84	9.68	99.82
BC1	53 535 640	43 534 067	81.32	9.95	99.88
BC2	53 107 498	43 803 076	82.48	9.54	99.875
BC3	52 163 251	42 803 517	82.06	9.62	99.875
AC1	53 283 822	44 342 596	83.22	9.38	99.865
AC2	57 288 974	47 462 963	82.85	10.34	99.545
AC3	57 453 759	47 612 099	82.87	10.34	99.875

[a]Duplication rate is the percentage of repetitive sequences in all clean sequencing reads.

explants from BC stage were then transferred to a new SHMab medium to obtain a well grown callus and this stage was designated as "After Callus" (AC). We collected the tissues from three different stages: control (CON, leaf explants on medium), BC stage and AC stage in three biological replications. The DNA was extracted from tissue at all three stages (CON, AC, and BC) and treated with bisulfite as described in the "Materials and Methods" section. The bisulfite-treated DNA (Bs-DNA) was used for the generation of next-generation sequencing libraries and whole genome bisulfite sequencing (WGBS) was performed with Illumina HiSeq 2500. The results of WGBS are shown in Table 1. We obtained 160 511 515, 158 806 389, and 168 026 555 high throughput sequencing raw reads from all the three different stages (CON, BC, and AC, respectively) and replicates. The raw reads were trimmed for adapters and preprocessed to remove low-quality reads. Filtered reads were mapped to *M. truncatula* R108 v1.0 reference genome (Kaur et al., 2021) with a unique mapping rate of 76.64% (123 015 679 reads), 81.95% (130 140 660), and 82.97% (139 417 658) in the CON, BC, and AC samples, respectively (Figure 1a). The raw total and mapped reads along with the unique mapping rate per each replication are shown in Table 1. Upon analysis, the observed methylation mapping rate and the mean cytosine methylation percentage in various contexts of cytosine methylation (CG, CHG, and CHH) reveal an altered methylation response during callus development in BC and AC in comparison to CON (Figure 1b). Total cytosine methylation and unique methylation proportions of cytosine methylation in all three contexts of CG, CHG, and CHH methylation were analyzed (Figure 1b; Table 2). Methyl cytosine (average of three replicates) was found higher at CG sites (51.2%), compared to CHG (33.9%) and CHH (24.1%) in the AC samples in comparison to the CON samples (Table 2). Cytosine methylation in CG and CHG context is comparatively higher than CHH methylation in all tissue stages of cytosine methylation (Figure 1b). In comparison to BC samples, AC samples had 2–7% higher rate of cytosine methylation in all contexts of cytosine methylation: CG, CH, and CHG (Table 2).

Correlation analysis was performed between the sample biological replicates to test the sample reliability and rationality of sample selection (Figure S2a). We observed a strong correlation between the samples as presented in the scatter plots (AC1 versus AC2, 0.975; BC1 versus BC2, 0.974; CON1 versus CON2, 0.972) (Figure S2a). Similar correlation between all biological replicates in a tissue stage were in the same ratio leading us to believe that the samples chosen were all true replicates and hence the robustness of the data (Figure S2b). The correlation ratio of all samples chosen for a particular stage was compared and presented as a Pearson correlation table (Figure S2b). Further, we performed motif analysis of the bases around the methylated cytosine in the context of CG, CHH and CHG methylation, to identify the most represented bases (Figure 2). We used Logo Plots (WebLogo) tool to explore the sequence information of a methylated site or its nearby nucleotide bases to help us understand the preference of methylated sites (Crooks et al., 2004). From the motif analysis of the bases around cytosine methylation in *M. truncatula* R108, there appears to be a bias to adenine, followed by thymine and cytosine in the contexts of CHH and CHG methylation and towards guanine in CG methylation (Figure 2a). Interestingly, we found higher methylated cytosines in all contexts of ^{m}CG, ^{m}CHG, ^{m}CHH patterns in the AC samples compared with BC and CON samples (Figure 1b; Table 2). This finding suggests that MLs are very dynamic and correlate with stress induction during tissue culture/SE (Table 2).

DNA methylation patterns in the coding sequences in response to callus formation stress

To better understand the relationship between DNA methylation and genetic structural regions, *M. truncatula* R108 genomic region was divided into several functional regions;

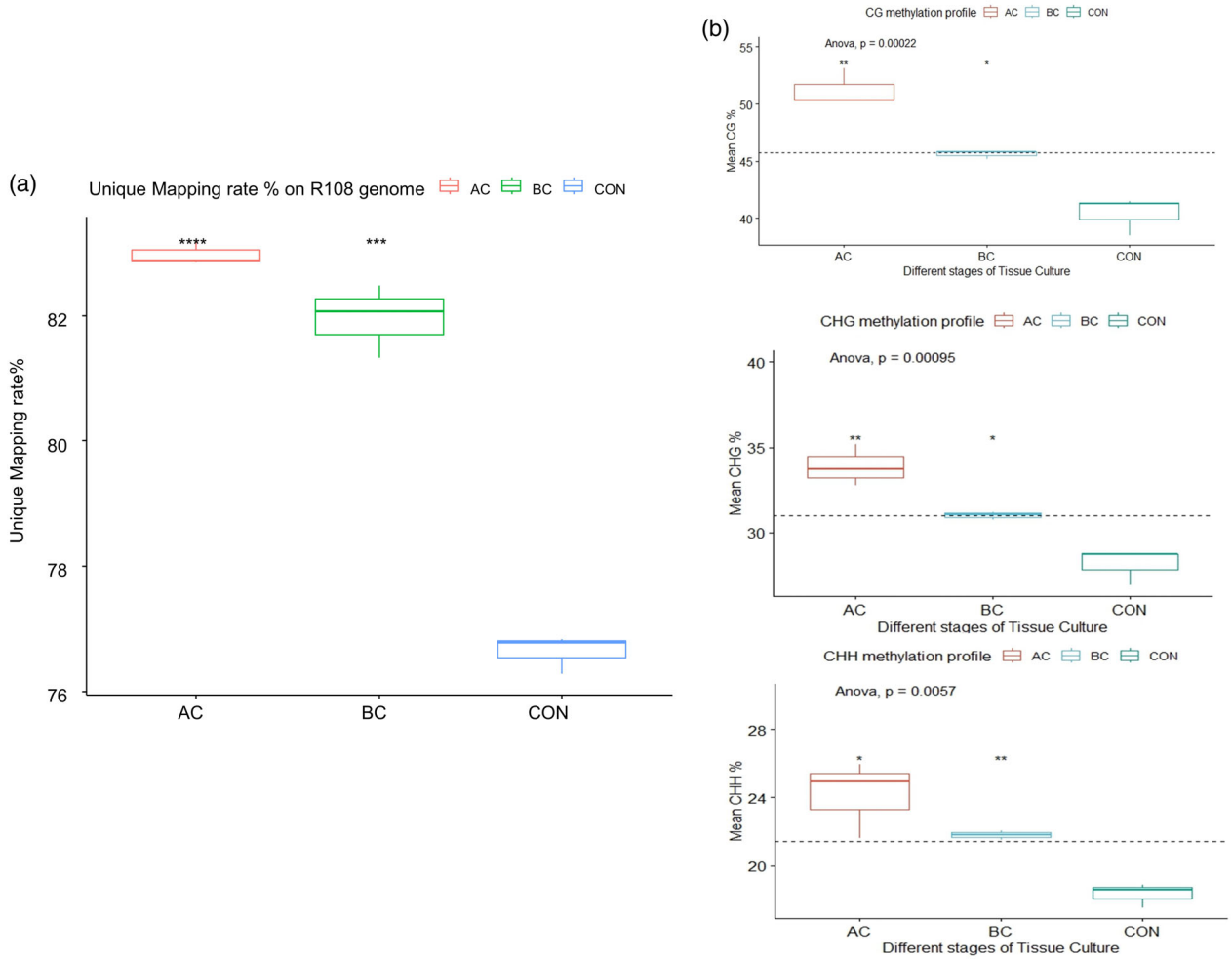

Figure 1. Methylation mapping during various tissue culture stages in *Medicago truncatula* R108 genome.
(a) Unique mapping rate in different stages (AC, BC, and CON) at whole genome level.
(b) Different levels of cytosine methylation in various contexts (CG, CHG, and CHH) in different stages (AC, BC, and CON) are presented. Highest level of cytosine methylation is observed in CG, CHG, and CHH methylation context in AC followed by BC in comparison to the CON. The horizontal line is drawn at base mean. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture. Student's t-test is used for statistical analysis based on three biological replicates (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ and ****$P < 0.0001$).

assumed promoter regions (2 kb region upstream of transcription start sites [TSSs]), 5′ untranslated regions (UTRs), exons, and introns and 3′ UTRs. DNA methylation levels (MLs) were calculated for the methylated sites using the formula ML = mC/(mC + umC), where mC represents methylated cytosines and umC represents unmethylated cytosines in all contexts of cytosine methylation in different stages of tissue culture; CON, BC, and AC (Figure 2b). Methylation profiling within genes in all stages of tissue culture exhibited a higher overall degree of methylation at presumptive promoter regions (2-kb upstream of the TSS) than genic regions (5′ UTR, exon, intron, and 3′ UTR), suggesting that the MLs of cytosine might be playing a role in gene expression. In each gene element, the frequencies of each context, including CG, CHG, and CHH methylation, CG methylation (green-colored graph) accounted for more than 50% of the total methylated cytosine (^{m}C) (Figure 2b). This suggested that CG methylation was the highest methylation pattern observed (Figure 2b). In all the three contexts of cytosine methylation, CG (green-colored graph, Figure 2b), CHH (pink-colored graph, Figure 2b) and CHG (orange-colored graph, Figure 2b), methylation within 5′ UTR region was slightly lower than the coding (CDS) regions (Figure 2b). We inferred that the presumptive promoter regions (2-kb upstream of the TSS) methylation might have been involved in the regulation of gene expression in response to the abiotic stress during callus formation. Increased levels of DNA methylation in the promoter regions and gene bodies inform us about the probable nucleosome compaction that might result in an altered gene expression.

Table 2 Statistical results of methylated cytosines in different contexts

Samples	^{m}C (%)[a]	^{m}CpG (%)[a]	^{m}CHG (%)[a]	^{m}CHH (%)[a]
CON_1	20.62	38.48	26.89	17.53
CON_2	21.97	41.50	28.76	18.61
CON_3	22.15	41.27	28.72	18.87
BC_1	25.17	45.78	31.06	21.81
BC_2	25.41	45.81	31.21	22.08
BC_3	24.88	45.20	30.75	21.55
AC_1	28.39	50.32	33.71	24.95
AC_2	25.67	50.27	32.75	21.64
AC_3	29.61	53.15	35.20	25.92

[a]^{m}C percent (%) is the percentage of methylated cytosines in all cytosines from the whole genome. ^{m}CG, ^{m}CHG, and ^{m}CHH percent (%) are the percentages of methylated ^{m}CG, ^{m}CHG, and ^{m}CHH in all corresponding cytosines context (C, CG, CHG, CHH, and H represents A, C, T) in genome, respectively.

Differentially methylated regions in response to callus formation stress

The DNA methylation profiles at different stages of tissue culture (CON, BC, and AC) showed differentially methylated regions (DMRs) that differ in methylated cytosines, lengths of the regions that were methylated, and group MLs (Table S1). These DMRs are significant as they signal an important epigenetic change. DMRs may be involved in the regulation of differentially expressed genes that might affect the biological processes (Table S1). Each of the observed DMRs were identified in different regions (promoters, exons, introns, and intergenic regions) of genes, with a larger footprint in the promoter region (Figure 3). In comparison between stages of the tissue culture AC versus CON, a total of 15 686 DMR regions were identified with 9207 hypermethylated regions and 6479 hypomethylated regions (Figure 3; Table S1a). In a comparison between different tissue culture stages, AC and CON, a total of 15 686 DMRs were identified, with 13 494 regions encompassing at least 100 bp of the methylated regions (Figure 3; Table S1a). DMR regions with longest regions identified for hypomethylated and hypermethylated regions were 3325 and 2713 bp, respectively (Table S1a). A few of the strong hypomethylated regions by length include a NBS-LRR disease resistance gene (*Medtr6g052750*), a gene encoding cytochrome P450 family protein (*Medtr2g072310*), a gene involved in senescence (*Medtr6g008015*), and a gene-encoding transmembrane protein (*Medtr7g056577*) (Table S1a). A few hypermethylated regions by length include a gene-encoding formin-like 2 domain protein (*Medtr7g080920*) and LRR receptor like kinase proteins (*Medtr4g036575*, *Medtr3g048590*, and *Medtr8g469670*) (Table S1a). Highest hypermethylation ratio between DMR regions of AC versus CON was found in subtilisin-like serine protease (*Medtr4g125180*) followed by chalcone and stilbene synthase family protein (*Medtr2g05870*) (Table S1a). DMRs were analyzed to identify the regions that contain the genetic structural regions. After further analysis of these genic regions that fall under the DMR umbrella, we found that highest number of DMRs were distributed within the promoter regions (10 279) followed by exons (4580) and introns (4071) with lower distributions in the UTR (untranslated) regions (Figure 3b). DMRs were mapped to the new R108 HiC genome (Kaur et al., 2021; Table S2) and DMRs mapped between AC and CON are presented in Table S2a.

In a comparison between different tissue culture stages, BC and AC, a total of 6442 DMRs were identified, with 5322 regions encompassing at least 100 bp of the methylated regions (Figure 3; Table S1b). Strong hypermethylation of genes in BC versus AC was found in extension-like repeat protein (*Medtr4g029600*), C2H2-type zinc finger protein (*Medtr1g090703*) and an ABC transporter protein (*Medtr3g005100*) (Table S1b). Strong hypomethylation of DMR regions was identified in comparison to AC in placenta-specific 8 (PLAC8) family protein (*Medtr3g072040*) followed by lower methylation in a kinase protein (*Medtr2g011640*) (Table S1b). DMR distribution between different tissue culture stages BC and AC were observed as higher in exons (3455) followed by intronic (2521) and promoter regions (2557) (Figure 3). Lowest distributions of DMRs were observed in 5′ UTRs (Figure 3b). The DMR regions between AC and BC were comparatively mapped to the new HiC R108 genome (Table S2b).

Comparisons between BC and CON tissue culture stages resulted in the identification of 13 805 DMRs, with 8387 hypermethylated DMRs and 5418 hypomethylated DMRs (Figure 3; Table S1c). At least there were 11 332 DMRs that circumscribe methylated regions greater than 100 bp (Figure 3; Table S1c). Higher hypermethylation was observed in the intronic region of cytokinin oxidase (*Medtr1g015410*) with methylation ratio of 59 times higher in comparison to CON (Table S1c). The highest number of DMRs were observed in promoters (8409) and lowest in UTRs (Figure 3). The number and distribution of DMRs in the comparison between BC versus CON stages resemble the comparison between AC versus CON stages (Figure 3b). The DMR regions between BC and CON were comparatively mapped to the new HiC R108 genome (Table S2c). Overall, methylation was dominant in the promoter region of the differentially methylated genes (DMGs) when compared to other regions (Figure 3).

DMRs are further analyzed by separating them into sequence context (CG, CHG, and CHH) between different tissue culture stages AC versus CON, BC versus CON, and BC versus AC (Figure S3). We observed higher numbers of hypermethylated DMRs with CHH methylation in AC versus CON and BC versus CON stages (Figure S3). The other types of methylation CG or CHG methylation showed similar kinds of hyper and hypo methylation DMRs in all stages of tissue culture comparisons (Figure S3).

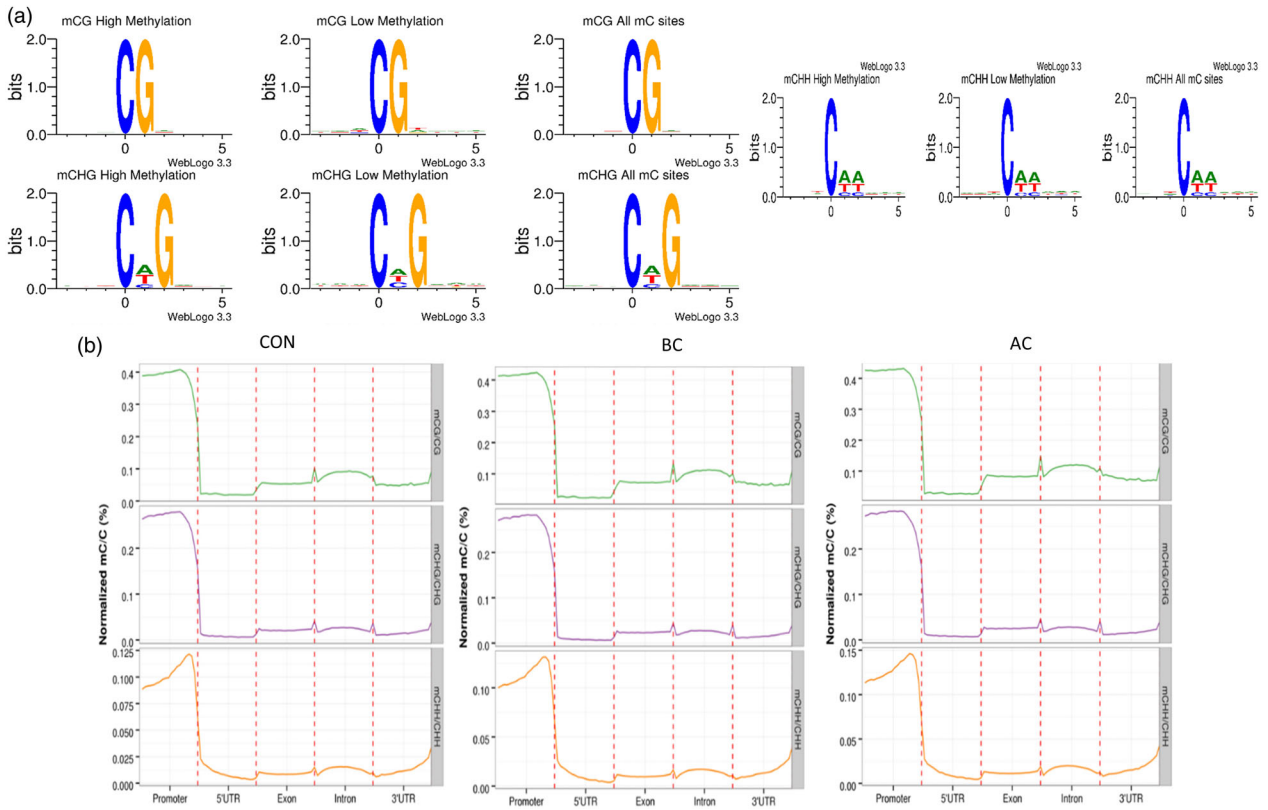

Figure 2. DNA methylation patterns in different *Medicago truncatula* genomic regions.
(a) Weblogo plots show the most commonly occurring patterns of cytosine methylation in the context of CG, CHG, and CHH.
(b) DNA methylation levels for methylated cytosines were plotted against gene structures in different stages (AC, BC, and CON). Highest levels of methylation is observed in the methylated cytosines of promoter regions in all the samples. Each of the panels are sub-divided into three sub-panels for each of CG, CHH, and CHG methylation pattern. The methylated cytosines are plotted as *Y*-axis and different gene structures (promoters, UTRs, exons, and introns) were plotted on the *X*-axis. The graphs in green, pink, and orange represent the CG, CHG, and CGG methylation patterns, respectively. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture; UTRs, untranslated regions.

DMPs in response to callus formation stress

Differentially methylated promoters (DMPs) refer to promoter regions with differential methylation. Though we do not know the exact promoter regions of every gene, the upstream 2 Kb region is generally regarded as promoter region for the gene in consideration. Promoter methylation represses the gene expression, and this process plays an important role in the regulation of gene expression. DMPs were identified by testing for each cytosine in each context (CG, CHG, and CHH) of the promoter region using Fisher's exact test and false discovery rate (FDR) was employed to correct the *P*-value. DMPs were filtered based on the FDR value <0.05 and with absolute difference of the MLs >0.2 between the samples. The FDR value estimation based on multiple hypotheses testing and *P*-values are based on statistical hypothesis test. Briefly, the absolute methylation data at each cytosine is taken and cumulatively added for each stage (methylated versus non-methylated) and ratio of methylated over total depth is calculated for each stage. The normalized difference between stages is attributed to the methylation ratio differences. All the samples with a negative methylation ratio difference of less than −0.20 is considered as hypomethylated and the samples with positive methylation ratio difference of greater than 0.20 are considered hypermethylated (Table S3). All the DMPs identified were clustered in different contexts of cytosine methylation (CG, CHH, and CHG) and are presented as heat maps for visualization (Figure S4). All the DMP MLs for union of clusters in different stages; AC, BC, and CON (151 for CG; 247 for CHH and 120 for CG) were clustered in different contexts of cytosine methylation (CG, CHH, and CHG) and are also presented as heat maps for visualization (Figure S4). The MLs of promoter region were plotted for the 2-Kb distance upstream from the TSS for all contexts of cytosine methylation (Figure 4). DMPs with hypermethylation (>0.6, Figure S4) and hypomethylation levels (<0.2; Figure S4) were observed in all the contexts of cytosine methylation (CG, CHG, and CHH) clusters. In the case of CHG methylation context, presumed promoter methylation (~2 Kb upstream of start codon) was lower in the AC and BC tissue culture stages in comparison to the CON stage

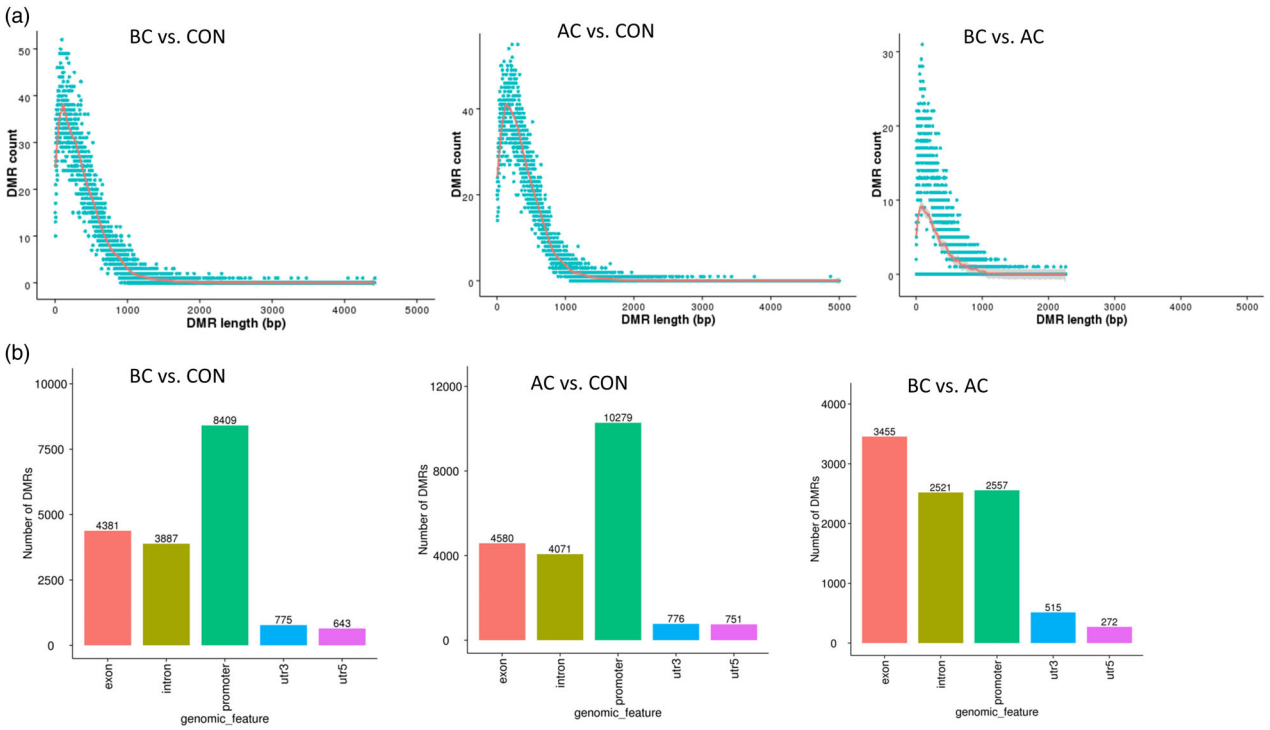

Figure 3. Differentially methylated region (DMR) analysis during various tissue culture stages in *M. truncatula* R108 genome.
(a) In different stages of tissue culture (AC, BC, and CON) presented, DMRs are analyzed and graphs are plotted with DMR length on the *X*-axis and the corresponding DMR count on the Y-axis.
(b) In different stages of tissue culture (AC, BC, and CON) presented, DMRs are analyzed for their distribution in different structural regions of the gene. The analysis of DMR distributions were plotted as bar graphs with gene structural regions (promoters, exons, introns, 5′ UTRs, and 3′ UTRS) on the *X*-axis and the number of DMRs on the *Y*-axis. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture.

(Figure 4). In Figure 4, the methyl cytosine levels are mean MLs for each context pertaining to genes with promoter regions labeled as "DMPs" and are compared across the stages of tissue culture. It is particularly interesting that promoters were less methylated in AC stage compared to BC stage in CHG context (Figure 4). Quite a contrasting pattern of promoter methylation was observed in CHH methylation context, with higher methylation pattern in tissue culture stages of AC and BC in comparison to CON stage (Figure 4).

Gene ontology analysis of genes with DMRs and DMPs in response to callus formation stress

In order to understand the specific gene groups that were enriched within the DMRs and genes with differential methylation in promoters, we performed Gene Ontology (GO) enrichment and KEGG pathway enrichment. In the comparison of tissue culture stages AC and CON, genes involved in the metabolic processes and phosphorylation were significantly enriched (Figure S5; Table S4a). GO enrichment profiles are presented for tissue culture stages BC versus CON (Table S4b) and for AC versus BC (Table S4c). This pattern of enrichment was conserved between stages AC and BC in comparison to CON that leads us to conclude that several genes involved in the metabolic and catalytic activity are activated or affected during the callus formation in comparison to the CON stage and that the activity of genes could have been regulated by methylation either directly or indirectly (Table S4). A similar approach was taken to identify the most enriched GO terms in DMP genes (Figure S6) and DMR genes (Figure S7). A significant number of genes involved in signal peptide processing, ATP hydrolysis and transport-related biological processes were enriched (Figure S6a). Further examination of DMPs in the hypermethylated regions identified genes involved in lipid glycosylation, photosynthesis, transport of ions, nucleotide binding, and RNA polymerase activity (Figure S6b; Table S5a–c). Significantly enriched genes from DMPs and DMRs are calculated based on *P*-values using hypergeometric tests with *P*-values less than 0.05.

In order to analyze the specific pathways that were enriched in genes with DMRs and DMPs, we performed KEGG pathway mapping for DMRs (Table S6) and DMPs (Table S7). KEGG pathway enrichment for DMGs in the tissue culture stages of AC and CON showed an enrichment in 128 metabolic pathways (Table S7a). Significant enrichment in the secondary metabolite pathways including linolenic acid and nitrogen metabolism pathways were identified in the callus forming stages of AC and BC in comparison to CON stage (Figure S8a). Similarly, in the comparison between BC and AC stages, strong enrichment

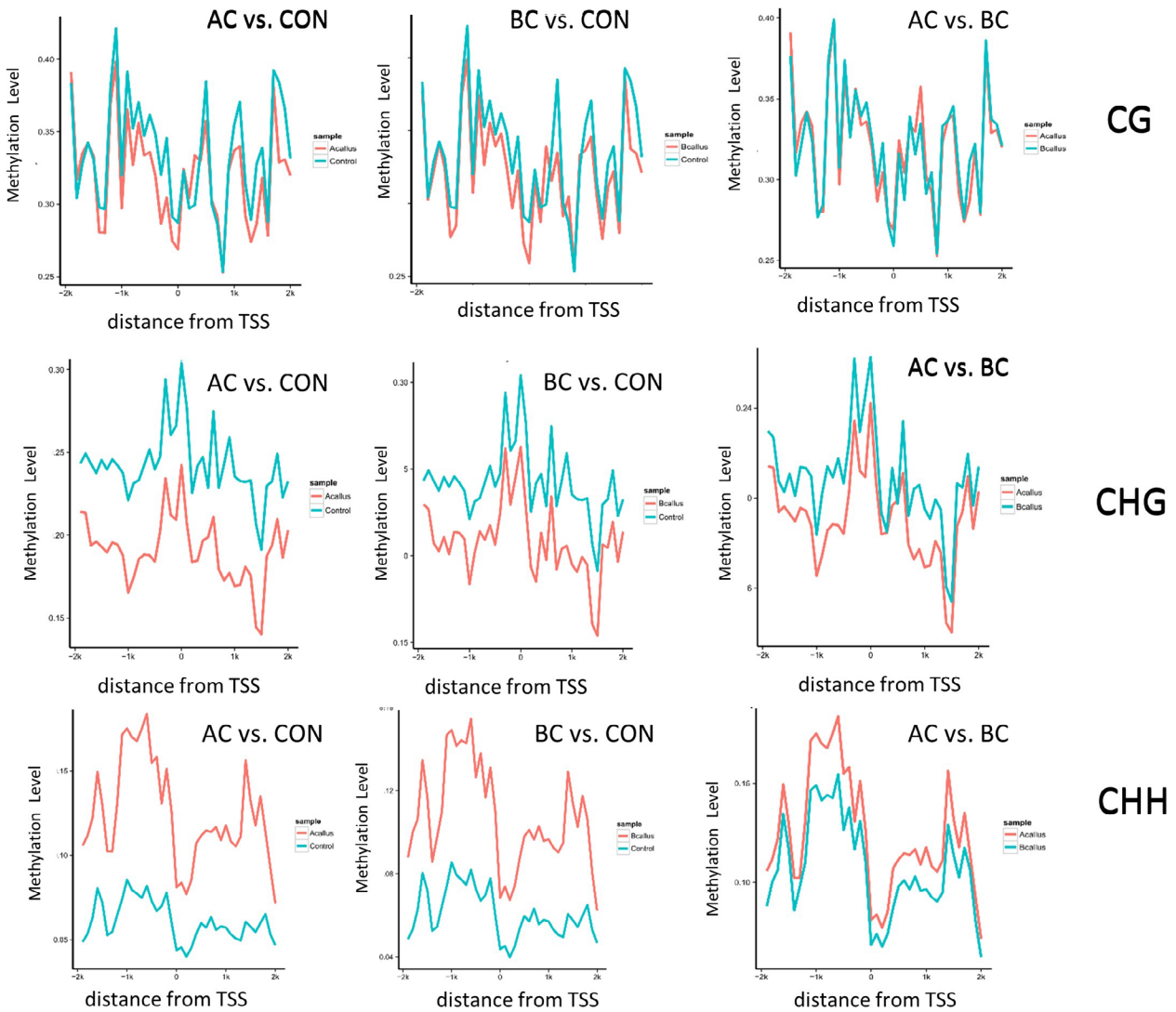

Figure 4. Differentially methylated promoter analysis during various tissue culture stages in *Medicago truncatula* R108 genome in the context of CG, CHH, and CHG methylation.
In each of the panels, CG, CHH, and CHG context, promoter methylation is mapped and up to 2 Kb from transcription state site (TSS) is mapped. The distance from the "TSS" is plotted on the *X*-axis and the methylation levels are plotted on the *Y*-axis. Higher amounts of DMP methylation levels were observed in the CHH context. In the AC versus CON, AC is represented by red-barred graph and CON is represented by blue-colored graph. In BC versus CON, BC is represented by red-barred graph and CON is represented by blue-colored graph. Each of the figures inset shows the color code used (Acallus: AC; Bcallus: BC, or Control: CON). In AC versus BC, AC is represented by red-barred graph and BC is represented by blue-colored graph. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture.

of pathways involved in amino acids, phosphate, and lipid metabolism/synthesis were identified (Figure S8b; Table S6c). Enrichment *P*-values were calculated using the hypergeometric test and the FDR correction method was applied using Benjamini and Hochberg test (Figure S8; Tables S6 and S7). In the comparison between AC and CON stages, genes involved in secondary metabolites, plant–pathogen interactions, photosynthetic, and metabolic pathways besides other important pathways were strongly enriched (Table S7).

Correlation between DNA methylation status and *Tnt1* insertion frequency

To understand the mechanism of *Tnt1* insertion frequency and its correlation to the methylation changes in the genome during callus formation, we compared *Tnt1* insertion frequency in the genes and physically mapped those regions to the cytosine-methylated regions. Briefly, the *Tnt1* insertion frequencies were compared against the cytosine methylation in DMRs between AC and CON

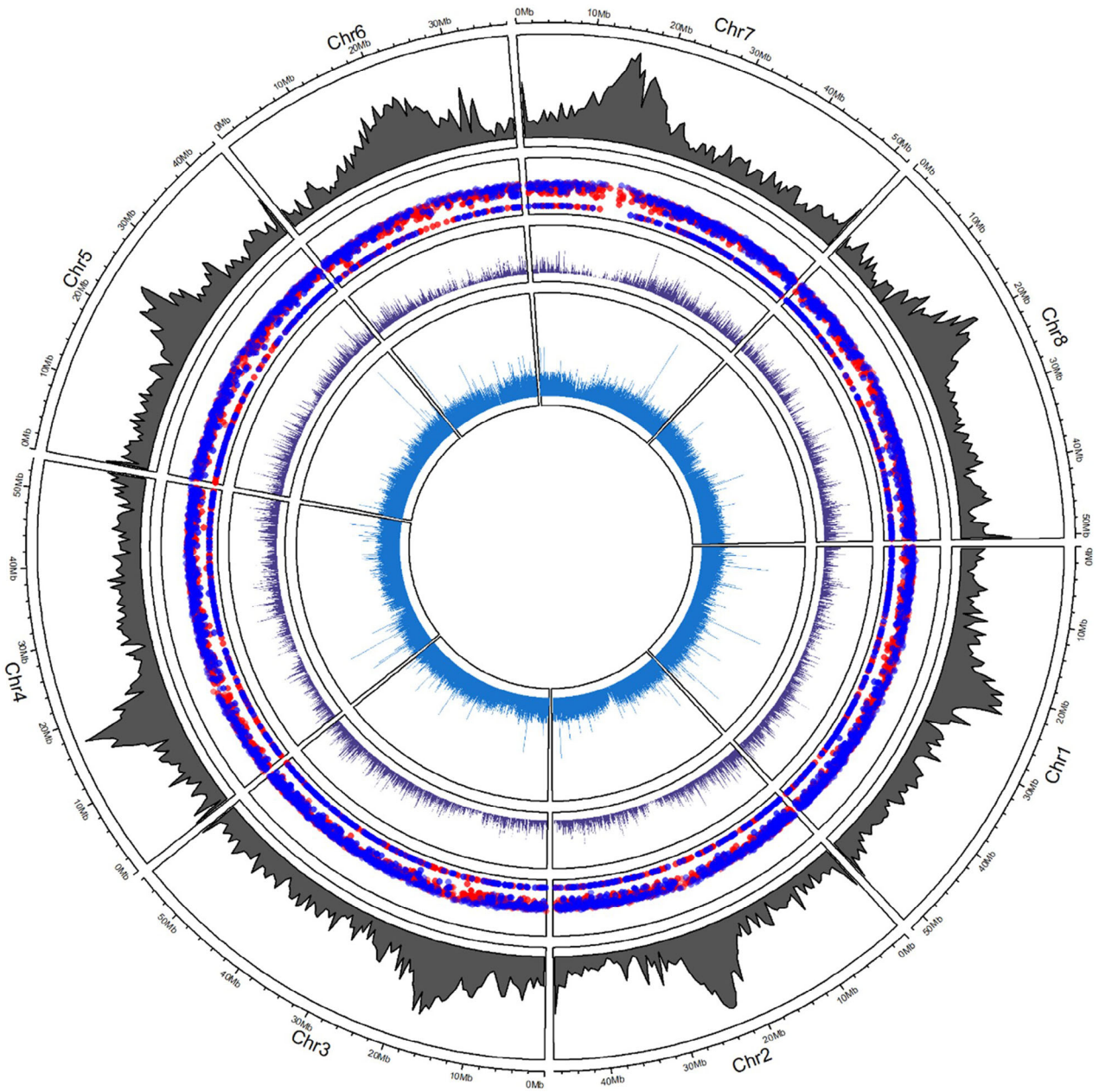

Figure 5. Circular genome visualization showing correlation between Differentially methylated regions (DMRs) and *Tnt1* insertion frequency in *Medicago truncatula* R108 genome between control and somatic embryogenesis.
Circular visualization map generated by R circlize program. The outer circle represents the eight chromosomes and the physical positions. The second circle shows the GC percent of the chromosomes shown as plots. The third circle shows the hypermethylated DMRs in blue and hypomethylated DMRs in red presented as rainfall plot. The fourth circle with dark blue bar diagrams represents the DMRs between CON and AC tissue culture stages and presented as density plot. The fifth circle shows *Tnt1* insertions as FST size at the corresponding positions as their insertion sites in the genome. AC, after callus; CON, controls represent different stages of tissue culture.

samples (Figure 5). We analyzed data from 7803 DMRs for which the *Tnt1* frequency insertion data exist. Correlation analysis between the *Tnt1* insertion frequency and DMRs were analyzed and presented as a circular genomic visualization map (Figure 5). In the circular visualization map, all the *M. truncatula* R108 genome chromosomes (1–8) are mapped on the outermost circle and the GC percentage of the chromosomes are plotted in the second circle from the outer circle (Figure 5). Each of the GC percentage of the chromosomes is plotted based on 500 Kb windows (Figure 5, second circle from outside). The third circle was plotted as a rainfall plot with hyper (blue) and hypo (red)

DMRs (Figure 5, third band). All the DMRs between AC and CON are presented as dark blue histograms (Figure 5, fourth circle from outside). The inner most circle shows *Tnt1* insertions as FST size at the corresponding physical positions in the genome representing their insertion sites (Figure 5, fifth circle from outside).

For each class of *Tnt1* insertion frequency ranging from one to 38 inserts per gene, DMRs were identified in AC stage in comparison to the CON stage (Figure S9; Table S8a,b). DMRs were further classified into hypermethylated regions and hypomethylated regions and the correlation analysis with *Tnt1* frequency revealed that there is a direct correlation of hypermethylated DMRs to the *Tnt1* insertion frequency (Figure S9a; Table S8a,b). Total DMRs that were identified in the *Tnt1* inserted regions corresponded to 89.8% (7803 DMRs out of 8697 DMRs) compared to 10.3% (894 DMRs) in the genes with no *Tnt1* insertions (Table S8b). Among the 7803 DMRs that correlated directly to the *Tnt1* mapping frequency, strong correlation exists for hypermethylation (51.5%; 4477 DMRs) compared to hypomethylated DMRs (38.2%; 3326 DMRs) (Table S8b; Figure 5). The genes with higher frequency of *Tnt1* inserts (37 *Tnt1* insertions) and hypermethylated DMRs were found to be Serine/Threonine kinase encoding gene (*Medtr3g102400*), followed by a receptor-like protein-encoding gene (*Medtr3g027330*) with 33 *Tnt1* insertions (Table S8a,b). In case of hypomethylated DMRs, genes with highest frequency of *Tnt1* insertions (55 insertions) encoded a helicase protein (*Medtr4g011450*) and 1,3-beta--glucan synthase component-like protein (*Medtr2g090375*; 54 *Tnt1* insertions) (Table S8a). Enrichment of genes in the DMR regions with *Tnt1* insertions is calculated with hypergeometric tests with 95% confidence interval with odds ratio greater than 1, with *P* value less than 0.05.

We further compared datasets between tissue culture stages AC versus BC (Table S8c,d) and BC versus CON (Table S8e,f) to verify the methylation patterns and their correlations with *Tnt1* insertions (Figure S9). In the comparison between BC and AC, we identified 3569 DMRs among which, 90% DMRs correlated well with having at least one *Tnt1* insertion (Table S8c,d), compared to 72% of DMRs (6836 DMRs) with at least one *Tnt1* insertion between BC and CON (Table S8e,f). DMRs with attributable *Tnt1* insertion in BC versus AC stages show a significant hypomethylation (63.8%; 2276 DMRs) compared to 26.7% or 953 hyper methylation DMRs (Figure S9b; Table S8d), while BC versus CON shows a presence of more hyper methylation DMRs (4132; 43.5%) compared to hypomethylation (2706; 28.5) DMRs (Figure S9c; Table S8f). Similar correlations were observed in the stages between AC and CON tissues, and between AC and BC stages. Among the hypomethylated DMRs between AC and BC stages, DMRs with highest *Tnt1* insertions include genes encoding 1,3-beta-glucan synthase component-like protein (*Medtr2g090375*, 54 *Tnt1* insertions), and gene encoding no exine formation protein, (*Medtr7g006760*, 36 *Tnt1* insertions; Table S8c). Hypermethylated DMRs with largest *Tnt1* insertions include genes encoding translation elongation factor (EF) protein (*Medtr2g020660*, 34 *Tnt1* insertions) and a receptor-like kinase (*Medtr3g116450* with 28 *Tnt1* insertions; Table S8c). Enrichment of genes in the DMR regions with *Tnt1* insertions is calculated with hypergeometric tests with 95% confidence interval with odds ratio greater than 1, with *P*-value less than 0.05. Among the hypermethylated DMRs between BC and CON stages, DMRs with highest *Tnt1* insertions include genes encoding U5 small nuclear ribonucleic helicase protein (*Medtr4g011450*, 66 *Tnt1* insertions), and gene-encoding beta amyrin synthase, (*Medtr4g005270*, 46 *Tnt1* insertions; Table S8e). Hypomethylated DMRs with relatively larger number of *Tnt1* insertions encodes a nucleotidyl transferase family protein (*Medtr4g027425*), with 49 *Tnt1* insertions (Table S8f).

Additionally, we performed the differential methylation site analysis at the *Tnt1* insertion junctions. Since the *Tnt1* is flanked by LTRs, we have included the 50-bp inner flanking region of the LTR regions from the insertion junctions in our analyses (Figure S10). We have plotted the differential methylated sites across the first 50-bp regions of the *Tnt1* junctions for each of the stages in the tissue culture AC versus CON, BC versus CON, and BC versus AC (Figure S10). Analysis was further focused on different types of methylation context; overall cytosine context, CG, CHG, and CHH methylation patterns (Figure S10a–c). We observed that CHH type of methylation is predominant at the junctions that span the first 50 bp of *Tnt1* insertions at the insertion junctions (Figure S10a). Absolute methylation values in the R108 line were calculated as the average methylation values and have correlated with the *Tnt1* insertion sites. For plotting the data, we took the absolute cytosine methylation values for the control R108 using *Tnt1* insertion sites (~10 bp) and have plotted the boxplots (Figure S10d). Methylation patterns in *M. truncatula* R108 genome between various stages of tissue culture is plotted on to the genome using bigwig files from the alignments and are viewed through the integrated genome viewer (IGV) (Figure S11). We present some examples of such visualization (Figure S11). All bigwig files will be made available upon request.

Differential gene expression and the corresponding methylation pattern changes in response to callus stress

Differences in gene expression and corresponding MLs were examined in three tissue culture stages CON, BC, and AC of *M. truncatula* (Figure 6). We performed the RNAseq experiment to understand the expression levels of genes during CON, BC, and AC stages. We used the same leaf/-callus tissue used for the methylation experiment and the experiment was performed with three biological replicates for each tissue sample type. The datasets from differential

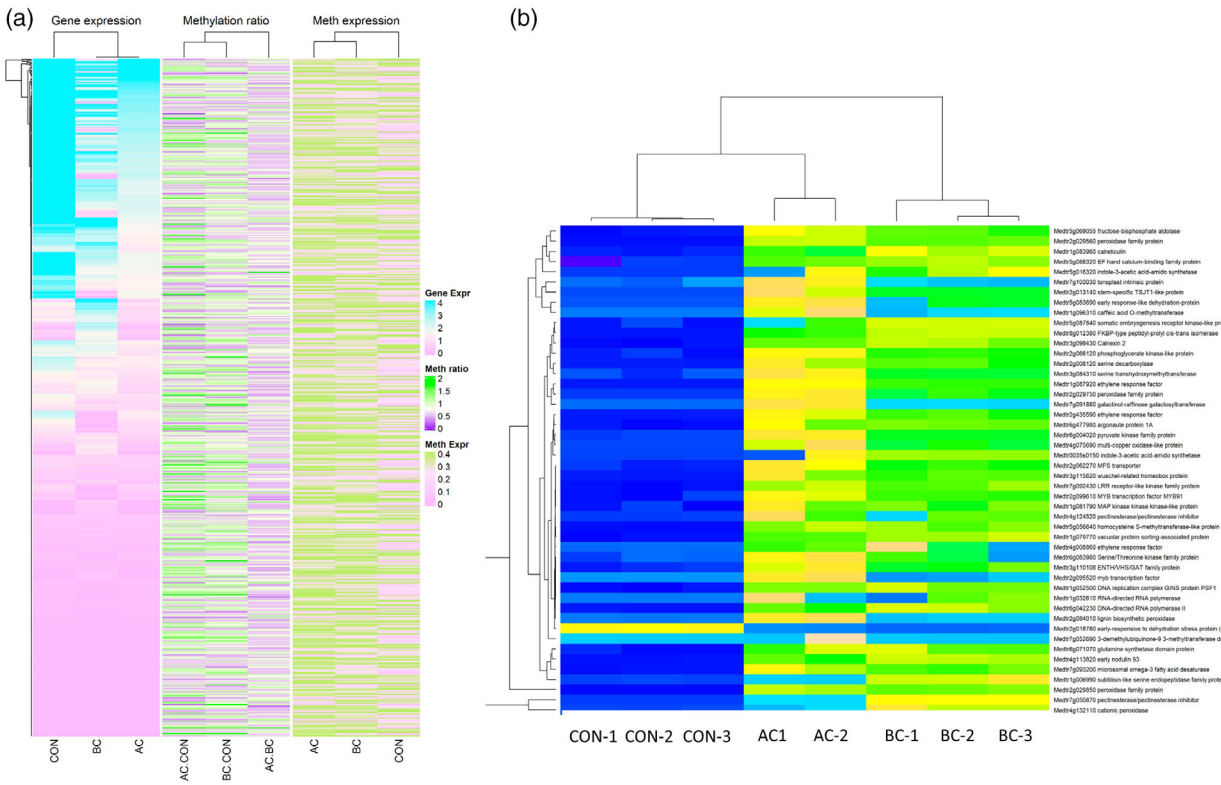

Figure 6. Differential gene expression between different tissue culture stages and correlation with gene methylation.
(a) RNAseq studies reveal the expression of genes that are differentially expressed between samples (AC, BC, and CON) in *Medicago truncatula* genome. Genes that are differentially expressed are presented as heat maps with log transformed values on *Y*-axis. The corresponding methylation ratios and methylation expression heatmaps for the similar genes were presented next to the gene expression profiles.
(b) The top 50 differentially expressed genes are presented. Samples are represented on *X*-axis and gene IDs along with the gene description are presented on *Y*-axis. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture.

gene expression (RNAseq data) and MLs were compared to identify patterns of methylation and their correlation with gene expression. All the methylation and gene expression values were log transformed and correlation studies are presented as heatmaps (Figure 6a). In a comparative analysis of DMRs and the corresponding gene expression patterns within those DMRs, higher levels of methylation in the DMRs correlated inversely to the corresponding gene expression (Figure 6a). Quantification of gene expression was performed by calculating the numbers of mapped reads and normalizing the results to FPKM (expected number of Fragments Per Kilobase of transcript sequence per millions of base pairs sequenced). Based on the criteria preselected for differentially expressed genes (DEGs) of *P*-value ≤0.05 a Log_2 fold change of either ≥"2" or ≤"2" and the presence of corresponding *M. truncatula* gene annotations, a significant number of DEGs (3515) exist between tissue culture stages AC versus CON (Table S9a); 3106 DEGs between BC versus CON (Table S9b) and comparatively lower number of 800 DEGs were identified between AC and BC stages (Table S9c). Analysis of DEGs in AC stage against the CON stage revealed that 172 protein kinase-related genes (e.g., *LRR receptor-like kinase*, *Medtr8g023720*: 20 fold change; *Osmosensor histidine kinase*, *Medtr8g075340*: 14-fold change); 123 transcription factors (e.g., *WRKY family transcription factor*, *Medtr1g013760*: 11.8-fold change; *myb transcription factor*, *Medtr2g023100*: 11.13 fold change), 10 embryogenesis related proteins (e.g., *late embryogenesis abundant hydroxyproline-rich glycoprotein*, *Medtr1g057420*: 12.12-fold change; *maternal effect embryo arrest protein, putative*, *Medtr3g462220*: 8.4-fold change), 39 histone and methyltransferase protein-related genes [e.g., *phosphoethanolamine N-methyltransferase*, *Medtr6g069600*: 16.65-fold change; *DNA (cytosine-5)-methyltransferase, putative*, *Medtr2g046020*: 8.98-fold change] and genes involved in the active methylation pathway (Table S9). DEGs that were expressed in AC stage in comparison to BC and CON tissue culture stages were analyzed and the top 50 DEGs expressed as FPKM counts were used to generate a heatmap (Figure 6b). Using a *P*-value <0.05 and log_2-fold change >2 in one sampling point as the higher criteria, we have plotted each of the DEGs as volcano plots at each pairwise comparison AC versus CON, BC versus CON, and AC versus BC

(Figure S12). DEGs are color coded in the volcano plots to represent the genes that are differentially expressed up to 2-fold change (blue), 2- to 3-fold change (red) and greater than 3-fold change (green) (Figure S12). Correlation between the gene expression data and the methylation expression was calculated using Pearson correlation and since we used the entire gene sets, a weak negative correlation was observed in AC versus CON stages (with *R* value of −0.08 and *P*-value of 2e-05) and in BC versus CON stages (with *R* value of −0.072 and *P*-value of 8.9e-08) of tissue culture (Figure S13). All the other DEGs expressed in all the libraries are presented as Supporting Information (Table S9d).

Methylation changes during SE

Tnt1 transposition occurs at the very early stage of regeneration phase in *M. truncatula* or during early SE (d'Erfurth et al., 2003). In order to test the epigenetic changes through the callus formation and SE stage, we extended the experiment to study methylation changes during SE (Figure S1; Figure 7; Table S10). The explants from SHMab medium, which were represented by bright yellowish callus, were transferred to SHMb medium for SE. To test the effect of methylation specifically at SE, we used azacytidine (5-azacytidine [5-azaC], a cytidine analog that can effectively suppress methylation) as a negative control for methylation. Callus transfers to SHM medium were done either in the presence or absence of 50 μM 5-azaC for embryo development. 5-azaC was used as an inhibitor of DNA methyl transferases (Santos, 2002). Genomic DNA was extracted from the embryogenic calli after 2 weeks and WGBS was performed in a similar manner as described above (Figure S1). The resulting bisulfite-converted reads were mapped on to the R108 *M. truncatula* genome (Kaur et al., 2021) and the cytosine methylation was assessed in the context of CG, CHH, and CHG methylation (Table S10). Absolute methylation values were obtained for all the cytosines in the treated and untreated samples and the methylation values were normalized based on the 5-azaC removal of methylation (Table S10). Normalized methylation values were used to calculate the percent methylation and are then classified into hypomethylation or hypermethylation based on their comparisons to the untreated calli. The samples were considered hypermethylated if the percent of methylation falls between 70 and 100% in comparison to hypomethylated samples (1–20%). This was done for all of the CG, CHH, and CHG forms of cytosine methylation (Table S10). Based on our analysis of the bisulfite reads and the cytosine methylation patterns during the SE stage, we estimated that CG methylation contributes to majority of methylation during SE stage (Figure 7; Figures S14 and S15) as seen in the earlier and later stages of callus formation (Figure 1). Hence, we compared the CG methylation and observed that at least 1008 genes were significantly methylated over 70% compared to the SE samples treated with 5-azaC (Table S10a,b; Figure 7a; Figure S15). Among the 1008 genes that were strongly methylated during SE for CG methylation, some of them include genes encoding transcription factors, transporter proteins, and kinases (Table S10b). Similar patterns were observed with the CG methylation in the promoter regions and strong methylation was observed in the promoter regions of genes encoding receptor-like kinases (*Medtr3g028650*, *Medtr2g036480*) and NBS-LRR protein (*Medtr2g038900*) (Table S10a,b). Correlation analysis of the SE methylation with *Tnt1* insertion frequency was done as described in the "Materials and Methods" section. Strong hypomethylation of genes (10–20% SE methylated genes) in CG context was observed with lower rates of *Tnt1* insertions and stronger hypermethylation of genes (70–100% methylated genes) for CG methylation was observed with the higher numbers of *Tnt1* insertions per gene (Figure 7a; Table S10b,c). This pattern is similar to the previously observed pattern of *Tnt1* insertion correlation in AC and BC tissue culture stages.

In the CHG methylation context, which is approximately 20% of the total cytosine methylation (Figure S14), strong methylation was observed for 1364 genes (Table S10e) coding for proteins involved in transport (*Medtr1g059080*, *Medtr3g093270*, MFS transporters; *Medtr1g029380*, bidirectional sugar transporter), ubiquitination pathways (*Medtr7g094280*, SCF Ubiquitin ligase SKP1 component; *Medtr7g094290*, E3 Ubiquitin ligase, *Medtr3g114990*, Ubiquitin ligase), kinases (*Medtr2g022810*, Serine/Threonine receptor kinase; *Medtr2g055690*, LRR receptor kinase), methyl transferases (*Medtr3g095830*; *Medtr4g013290*), and other important plant metabolic pathway proteins (Table S10d,e; Figure 7b). Similarly, promoters of genes involved in receptor kinases (*Medtr1g027840*, stress-induced receptor-like kinase; *Medtr1g018910*, Serine/Threonine receptor kinase), transferases (*Medtr7g445710*, histone-lysine N-methyltransferase; *Medtr5g072825*, *Medtr6g014240*, UDP-glycosyltransferases), late embryogenesis proteins (*Medtr7g118270*, *Medtr2g100060*, late embryogenesis abundant protein; *Medtr2g081080*, *Medtr1g075530*, embryosac development arrest protein), and other significant protein classes in the embryogenesis were highly methylated relative to 5-azaC controls (Table S10e). Correlation analysis of the SE methylation with *Tnt1* insertion in CHG context displayed a contrasting pattern to the correlations observed in CG methylation to *Tnt1* insertions during SE (compare Figure 7a and 7b; Table S10d,f). Strong hypermethylation of genes (70–100% SE methylation) was observed with lower rates of *Tnt1* insertions in CHG context and stronger hypomethylation of genes (1–20% methylation) was observed in CHG context with the higher numbers of *Tnt1* insertions per gene (Figure 7b; Table S9d–f). In the context of CHH methylation, strong methylation was observed for genes involved in stress responses (*Medtr1g028100*, *Medtr2g084315*) among 2650 genes that were differentially methylated (Table S10g,h; Figure 7c).

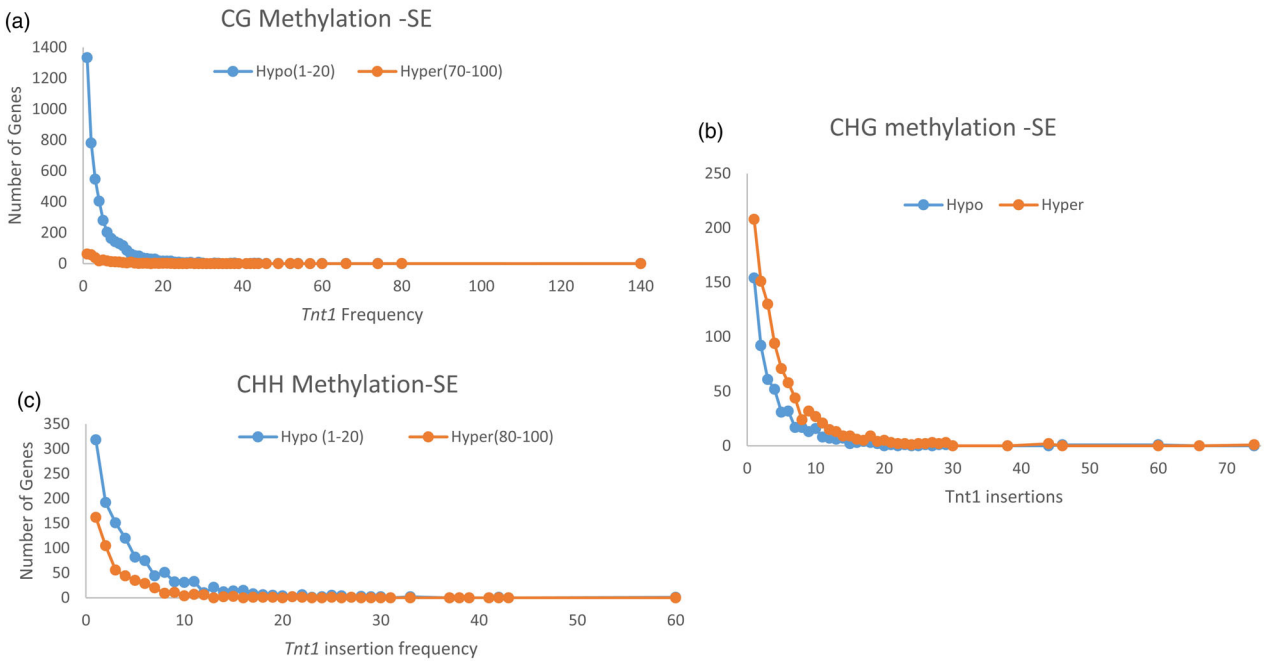

Figure 7. Correlation of *Tnt1* insertions with gene methylation during somatic embryogenesis (SE).
(a) CG methylation in SE showing the correlation of hypermethylated and hypomethylated genes with *Tnt1* insertion frequency.
(b) CHG CG methylation in SE showing the correlation of hyper and hypomethylated genes with *Tnt1* insertion frequency.
(c) CHH CG methylation in SE showing the correlation of hyper and hypomethylated genes with *Tnt1* insertion frequency. Hypo methylated genes are represented by blue-colored graph and hypermethylated genes are plotted using red-colored graph. The numbers of genes are plotted on *Y*-axis and *Tnt1* insertion frequency on *X*-axis. All the values are log transformed and the legends represent the range of their expression.

Correlation analysis of the SE methylation with *Tnt1* insertions in CHH context displayed a pattern similar to the CG methylation during SE (Figure 7c; Table S10h,i).

DISCUSSION

We performed a comprehensive genome-wide investigation of the methylation profiling during early embryogenesis/callus generation stage in *M. truncatula*. Methylation profiling during the callus generative process of *M. truncatula* revealed a genome-wide increase in DNA methylation in the DMPs, which was most apparent for CHH methylation. We checked to see the correspondence of the CHH methylation and the symmetric methylation (CG and CHG) with the *Tnt1* retrotransposition and hypothesized that an active RdDM regulation process might be in play during this process. We explored the methylation patterns and the corresponding gene expression in the DMR regions and the frequency of the *Tnt1* insertions for those regions.

Based on our methylation profiling in the genes, we found that assumed promoter regions (~2-Kb upstream regions of TSS) displayed significantly higher levels of methylation than exons and introns, suggesting that the cytosine methylation may be correlated with gene expression. This is in accordance with the previously published work that showed methylation most frequently occurs in the so-called CpG islands in the 5′ regulatory gene regions (promoters) (Eprintsev et al., 2017, 2020; Lu et al., 2017). Gene promoter methylation in AC and BC tissue culture stages could be due to the *de novo* methylation to regulate gene expression in response to the abiotic stress during SE. This correlation was observed between the methylated regions and the suppression of gene expression in the highly methylated regions (Figures 6 and 7). In the DMP analysis, we report the active site cytosine hypermethylation in the CHH context compared to CG and CHG methylation which is associated with CMT2/DDM1 and RdDM components of methylation. This is associated with the enrichment of 24-nt small RNA species and gene expression along with the transgene silencing (Stroud et al., 2014). This kind of hypermethylation of cytosine methylation was also observed in other abiotic stress responses such as salinity stress in *M. truncatula* (Yaish et al., 2018). Strong hypermethylation of genes (70–100% methylated genes) for CG and CHH SE methylation was observed with the higher numbers of *Tnt1* insertions per gene (Figure 7), while the pattern displayed an opposite trend in CHG context during SE (Figure 7). This continuous reconfiguring of methylation patterns could occur during SE enabling the silencing of transposable elements as observed in rice shoot apical meristem (Higo et al., 2020).

Based on the data we observed, there has been an increase in methylated cytosines in the context of CG,

CHG, and CHH methylation from tissue culture progression CON stage to AC stage suggesting the increasing role of MET1, CMT3, and CMT2 and a synergistic role of RdDM pathway. In the RNAseq experiments, top 50 DEGs from CON to AC stages suggest that the early embryogenesis is triggering the expression of genes encoding several transcription factors (WUSCHEL, MYB), kinases (SE receptor kinases, MAP kinases, S/T kinases), ethylene responsive genes, oxidases, carboxylases and transporters, and genes related to small RNA regulation and methylation pathways (Figure 6). The corresponding MLs for all these genes seemed to be at a very low level, perhaps hypomethylated in some cases (Table S10). This is expected since early embryogenesis events are auxin-dependent and require the expression of several transcription factors (Grzybkowska et al., 2020; Tian et al., 2020). Several transcription factors were shown to be expressed and forms the key to embryo development in Arabidopsis (Le et al., 2010; Meinke, 2020). Several studies involving transcription factors: for example, *AGAMOUS LIKE* (*AGL*), *SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE1* (*SERK1*), and *WUSCHEL* (*WUS*) in different species during embryogenesis were summarized (Mendez-Hernandez et al., 2019). Studies involving SE report correlation with auxin signaling and hypermethylation at non-CG methylation (CHH; Kawakatsu et al., 2016). Embryo maturation stage was also reported to show gradual increase in the global levels of methylation (Gehring et al., 2009; Grzybkowska et al., 2020; Hsieh et al., 2009; Santos, 2002). Similar studies in induction stage and globular stage SE in soybean reported a genome-wide increase in methylation (Ji et al., 2019).

Plant retrotransposons including *Tnt1* are known to be activated during severe stress conditions such as SE (d'Erfurth et al., 2003; Grandbastien, 1998; Melayah et al., 2001). Both biotic and abiotic stress conditions induce the expression of retrotransposons (Melayah et al., 2001). In addition, plant stresses such as wounding, fungal elicitors, viral inoculation, hormone treatment, etc. have been shown to induce the expression of retrotransposons (Grandbastien, 2015). The SE protocol in the case of *M. truncatula* uses leaf tissue as an explant. Sterile leaf segments are subject to hormone treatment and other chemicals that can generate stress (Lee et al., 2018). In *M. truncatula*, it has been shown that plant regeneration by SE, but not by organogenesis, will induce the transposition of *Tnt1* (Tadege et al., 2005). Interestingly, in *M. truncatula* the osmotic treatment of explants (leaves) was able to improve transposition frequency (Iantcheva et al., 2009).

However, the expression of retrotransposons does not always correlate with transposition. It has been reported that the expression of retrotransposons can be located in specific tissues and that this expression can be mediated by element-specific regulatory sequences where host transcriptional factors play a role (Grandbastien, 1998). On the other hand, LTR retrotransposons are frequently concentrated in hypermethylated regions as the intergenic heterochromatic clusters and there is evidence of epigenetic mechanisms of targeting transposable elements (Rigal & Mathieu, 2011). Thus, when the retrotransposons *Tto1* (Hirochika et al., 2000) or *Tnt1* (Lucas et al., 1995) were introduced in Arabidopsis, they were active but methylated and silenced as their copy number increased. With this complexity, transposition, integration and/or the degradation and silencing of transposable elements can be associated with either biotic or abiotic factors that affect the outcome of the events in plant species. Most asymmetric methylation (CHH) is facilitated by "Decreased DNA Methylation" (DDM1) and mediated by the *CHROMOMETHYLASE 2* (CMT2) separately from RdDM (Zemach et al., 2013). Heterochromatic sequences preferentially require DDM1 for DNA methylation and that this preference depends on linker histone H1 (Zemach et al., 2013). Together, DDM1 and RdDM mediate nearly all transposon methylation and collaborate to repress transposition and regulate the methylation and expression of genes (Sasaki et al., 2019; Zemach et al., 2013). RdDM on the other hand often seems to target intergenic regions and plant genes located in the euchromatin regions, which suggest Pol IV and RdDM will act at the borders between open chromatin and regions with elevated CG/CHG methylation (Li et al., 2015; Zemach et al., 2013). It is possible that there might be synergistic effects taking place with both the independent pathways in action.

In our study, we observed the correlation between the *Tnt1* insertion sites and the presence of DMRs at AC and BC stages in comparison to control. DMRs identified between the tissue culture stages AC versus CON, BC versus CON, and BC versus AC, were used for correlation studies with the *Tnt1* insertion study and that included both the hypermethylated and hypomethylated DMRs (Table S1; Figure 5). A recent study in rice suggests the inverse relation of loss of CG methylation to a burst of transposable elements in the cell cultures (Hu et al., 2020). We predict that the transposition of *Tnt1* probably happens during the stages between callus initiation to SE, given the higher rate of CHH methylation and lower rates of CHG and CG methylation in AC stage of tissue culture (Figure 5). Sun et al. (2019) reported the correlation between *Tnt1* insertions in *M. truncatula* R108 ecotype and methylation frequency in the leaves of A17 ecotype, and concluded that it was not certain that gene methylation actually affects *Tnt1* insertion. Probably the increase in methylation is to reduce the ongoing transposition activity of transposable elements. We found an indirect correlation with the DMR regions spanning genes and the log transformed gene expression values of the corresponding genes, showing that active methylation is used as a regulation mechanism to promote or suppress gene expression (Figure 6). Recent

study confirms the Bestor hypothesis that "an important function of DNA cytosine methylation was to silence the expression of transposable elements" (Zhou et al., 2020). This study was aimed at understanding why the methylation would allow for the coexistence of transposable elements and the host in a type of host–parasitic relationship, given the potentially harmful effects of the ectopic expression of such elements (Zhou et al., 2020). They conclude that though we still do not know the evolutionary significance of genome expansion, CG methylation causes increased rate of conversion from C-to-T transition mutations, that generates a strong decrease in the observed/expected (O/E) ratio of CGs in the DNA of organisms (Zhou et al., 2020). In a recent study on soybean SE in continuous cultures for over 13 years, resulted in the identification of rare failures to maintain DNA methylation and that the demethylated regions accumulate over time (Ji et al., 2019). The probable explanation that maintenance of DNA methylation pathways cannot keep up with the continuous and rapid rate of cell division without reinforcement of epigenome could also explain why there might be a need for an increased methylation during embryogenesis.

In this study we observed the increased methylation in all contexts of cytosine methylation after the initiation of callus and post callus formation and this increased methylation correlated with the *Tnt1* insertion frequency particularly in the hypermethylated DMRs. Hypermethylated regions (DMRs) have indirect correlation to the corresponding gene expression in the AC and BC samples compared to the CON samples.

MATERIALS AND METHODS

Plant materials and growth condition

Medicago truncatula ecotype R108 was used for all the experiments described in this study. Seed materials were prepped as per the methods described before (Lee et al., 2018). Briefly, scarified *M. truncatula* seeds were germinated overnight on moist Petri dishes and placed at 4°C for 1 week. Plants were grown at 25°C day/23°C night temperature, 16 h day/8 h night photoperiod, 60–70% relative humidity. Germinated seedlings were transplanted into soil trays. After 4 weeks of growth, the leaf explants were collected from wild type R108 *M. truncatula* plants. Explants were surface sterilized with 70% ethanol and 30% clorox. The sterilized explants (10–15) were placed on modified SH (Schenk & Hildebrandt, 1972) medium SHMab medium and incubated in a growth chamber at 24°C. Following a week's growth on the SHMab medium, explants began to lose chlorophyll and showed signs of callus induction. The explants at this stage that show a sign of callus formation was designated as BC stage. The explants were then transferred to a new SHMab medium every 2 weeks and checked for callus induction. After the third transfer, well grown callus turns into yellowish color and was designated as AC stage. Callus tissue from the AC stage was transferred to SHMb medium and incubated under light in a growth chamber at 24°C for embryogenesis. Callus is maintained on SHMb medium for 2 weeks. At the end of 2 weeks, callus starts to turn green, suggesting embryogenesis is occurring and was designated as SE stage. Samples were collected from at least three replications.

DNA isolation and library preparation

Leaf and callus tissue were flash frozen and finely ground to powder using a mortar and pestle. DNA extractions were performed on all samples using Qiagen DNeasy Plant Mini Kit (Qiagen, Germantown, MD, USA). Genomic DNA was fragmented by sonication to 200–300 bp with a Covaris S220 sonicator, followed by end repair and adenylation. Lambda DNA was used as an unmethylated control for calculating the bisulfite conversion rate. Then, the DNA fragments were treated twice with bisulfite using the EZ DNA Methylation-Gold™ Kit (Zymo Research, Irvine, CA, USA). BS-seq library constructions and sequencing utilized three biological replicates per sample. Libraries were sequenced on the Illumina Hiseq 2500 platform (Novogene, Beijing, China) using paired end sequencing.

High-throughput sequencing

Illumina sequencing was performed on an Illumina HiSeq 2500 instrument at Novogene. For MethylC-seq, raw reads were trimmed for adapters and preprocessed to remove low-quality reads using cutadapt 1.9 dev1 (Martin, 2011). All the raw data were deposited at NCBI under the bioproject "PRJNA757187" with biosample accessions SAMN20947071.

Reads mapping to reference genome

Bismark software (version 0.12.5) (Krueger & Andrews, 2011) was used to perform alignments of bisulfite-treated reads to a reference genome, *M. truncatula* R108 v1.0, with the default parameters. Briefly, low-quality reads were removed from raw data by Trimmomatic (Bolger et al., 2014). The reference genome was first transformed into bisulfite-converted version (C-to-T and G-to-A converted) and then indexed using bowtie2 (Langmead & Salzberg, 2012). Sequence reads were also transformed into fully bisulfite-converted versions (C-to-T and G-to-A converted) before they were aligned to similarly converted versions of the genome in a directional manner. Sequence reads that produce a unique best alignment from the two alignment processes (original top and bottom strand) were then compared to the normal genomic sequence and the methylation state of all cytosine positions in the read was inferred. The same reads that aligned to the same regions of genome were regarded as duplicated ones. The sequencing depth and coverage were summarized using deduplicated reads. The results of methylation extractor were transformed into bigWig format for visualization using Integrated Genome Viewer (IGV) browser. The sodium bisulfite nonconversion rate was calculated as the percentage of cytosines sequenced at cytosine reference positions in the lambda genome.

DMR and DMP analysis

Differentially methylated regions were identified by the "bsseq" package (http://www.bioconductor.org/packages/release/bioc/html/bsseq.html) with read coverage ≥5, the difference of methylation level (ML) ≥0.1 or fold change ≥2, and with corrected *P*-value <0.01, using a sliding-window approach. Bsseq is based on BSmooth algorithm (Hansen et al., 2012) that uses a sliding-window method in a specified region with a false discovery rate (FDR) <0.5. Differentially methylated sites were merged, and filtered to get the final DMR using cutoff: (a) at least 0.1 difference in ML; (b) at least three cytosine sites in the DMR region; and (c) the two adjacent cytosine sites cannot go beyond 300 bp. The

window was set to 1000 bp and the step length was 100 bp. The Fisher test was implemented to detect the DMRs. Genes overlapping with significant DMRs for at least 1 bp in the functional region were defined as DMR-associated genes (DMGs). The DMRs were identified by comparison of the MLs of 1-Kb windows/regions throughout the genome between any two genotypes. For the identification of DMGs, either the ML of the gene body or 2-Kb upstream flanks (promoter) of the gene has a difference (calculated the same as DMCs). DMP is identified by testing for each cytosine in each context (CG, CHG, and CHH) of the promoter region using Fisher's exact test and then FDR is employed to correct the *P*-value. DMP was filtered to only include promoters of which FDR value was less than 0.05 and the absolute difference of the MLs between different samples should be greater than 0.2.

Correlation analysis

To test the reliability of the experiment and the rationality of the sample selection, correlation between the MLs of the sample replicates was conducted in R. The closer the correlation coefficient is to 1, the higher the similarity of methylation patterns between the samples.

5-Azacytidine treatment

Callus transfers to SHM medium were done either in the presence or absence of 50 μM 5-Azacytidine (5-azaC) for embryo development. Cytidine analogue, 5-azaC, is used as an inhibitor of DNA methyl transferases (Santos, 2002). Callus samples were collected in at least three replications. Differential methylation data processed was analyzed using Bismark software and the sample treatments that were more than 70% methylated than the 5-azaC-treated samples were chosen for further analysis. Among the samples chosen with greater than 70% methylated than 5-azaC samples were further assessed to analyze the genes that had greater than methylated cytosines between 5-azaC and non 5-azaC samples.

Correlation of methylation to the *Tnt1* insertion analysis in SE

The data were averaged for all three biological samples either among the azacytidine or non-azacytidine treatments. The azacytidine values either in the promoter regions or genic regions were subtracted from the non-azacytidine values to arrive at the MLs and for calculation of methylation percentages for each of the transcript/gene in *M. truncatula*. Corresponding *M. truncatula* gene transcript names were mapped and the *Tnt1* insertions were plotted against the transcripts. The *Tnt1* insertion frequencies in *M. truncatula* were plotted against the number of genes that were either hyper or hypomethylated. For hypermethylation, the number of genes that fall in the category of 70–100% methylation range were summed up to get the number of genes that were hypermethylated for the specific *Tnt1* frequency. In case of hypomethylation, the number of genes that fall in the category of 1–20% methylation range were summed up to get the number of genes that were hypomethylated for the specific *Tnt1* frequency.

Gene ontologies and KEGG pathway analysis

Gene ontology (GO) enrichment (Ashburner et al., 2000) analysis of genes related to DMRs was implemented by the GO-seq R package (Young et al., 2010), in which gene length bias was corrected. GO terms with corrected *P*-value less than 0.05 were considered significantly enriched by DMR-related genes. All the DMGs were annotated by BGI Web Gene Ontology Annotation Plotting (http://wego.genomics.org.cn/). KEGG (Kanehisa et al., 2016; Masoudi-Nejad et al., 2007; Okuda et al., 2008; Tanabe & Kanehisa, 2012) is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism, and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high throughput experimental technologies. KOBAS software (Mao et al., 2005) was used to test the statistical enrichment of DMR-related genes in KEGG pathways.

Expression analysis

For *M. truncatula* embryogenic callus gene expression profiles, total RNA was isolated from the same samples used for methylome analysis and RNA-Seq data was generated from total RNA. Clean tags filtered by Trimmomatic (version 0.32) were aligned to *M. truncatula* R108 genome (v1.0) by Tophat (version 2.0.13). The suite of Cufflink software (version 2.2.1) was used for splicing transcripts and analyzing expression scores (FPKM, reads per kilobase of exon model per million mapped reads) of each gene. Genes with at least 2 log-fold expression change and significant statistical difference (FDR cutoff 0.05) were identified as DEGs using DEseq2 in R (Love et al., 2014). All the heatmaps and volcano plots to visualize DEGs were generated in R (R Core Team, 2020).

FST sequence analysis

To accurately identify *Tnt1* insertion sites in the *M. truncatula* genome, all FST sequences with length shorter than 50 bp or without the *Tnt1* signature sequence ("CCCAACA", "CATCATCA," or "TGATGATGTCC") or the *Tnt1* signature sequence not within 28 bp from the beginning or at the end of FST sequences were discarded. The preprocessed reliable FST sequences were aligned to the *M. truncatula* R108 v1.0 reference genome (Kaur et al., 2021) using BLASTN with an e-value threshold ≤1.00E-5. The FST sequences with the best hit from BLAST analyses were further processed to filter incorrect alignments if the similarity score was less than 90% (Altschul et al., 1997; Johnson et al., 2008). DMR regions that are presented in the Supporting Information were plotted in the visualization.

Visualization and statistics of *Tnt1* insertion frequency

GC content was calculated as following: GC content = (G + C)/(A + T + G + C) × 100%. Circos was used to visualize the GC content and *Tnt1* insertion frequencies (Krzywinski et al., 2009; Zhang et al., 2013).

AUTHOR CONTRIBUTIONS

RSN and KSM contributed to the conception and design of the experiment. RSN contributed to the acquisition of the data. H-KL, SO contributed to the sample treatments and DNA extractions, RSN, NK and KSM contributed to data analysis and interpretation. RSN and KSM drafted the manuscript, and all the authors critically revised and approved the final version of the manuscript for publication.

ACKNOWLEDGMENTS

We thank Janie Gallaway and Cindy Green for their excellent work in plant maintenance. We also thank the genomics core facility at the Noble Research Institute for their help with Illumina library

generation and sequencing. We thank Novogene for generating bisulfite libraries, sequencing and for help with data analysis. This work was supported by the Noble Research Institute, LLC and a National Science Foundation grant (IOS-1733470) awarded to KSM. This research used resources provided by the SCINet project of the USDA Agricultural Research Service, ARS project number 0201-88888-002-000D, 0201-88888-003-000D, 0500-00093-001-00D and ARS project number 3060-21000-046-00D. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All the raw data are deposited at NCBI under the accession/bio project "PRJNA757187" with bio sample accessions (SAMN20947071).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Schematic to generate the embryogenic calli from the leaf explants. Schema to regenerate calli from leaf explants is shown and the embryogenic calli samples grown from the leaf explants are shown in the upper panel. Embryogenic calli samples are shown as a representative picture in the inset.

Figure S2. (a) Correlation analysis between the methylation levels of the sample replicates. In different stages of tissue culture (AC, BC, and CON), presented, correlation between the sample replicates is calculated. Each of the scatter plots in the top panel represent the correlation analysis performed between any two replicates in the same stage (AC1 versus AC2, 0.975; BC1 versus BC2, 0.974; CON1 versus CON2, 0.972). (b) In different stages of tissue culture (AC, BC, and CON), presented, correlation between the sample replicates is calculated. Correlation ratios of all the sample replicates within the tissue stages is shown as a Pearson correlation table. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture.

Figure S3. DMRs in the sequence context between tissue culture stages. DMRs are separated into sequence context (CG, CHG, and CHH) methylation between different stages of tissue culture (AC, BC, and CON). DMR numbers are represented on *Y*-axis and the stages and types of methylation are represented on *X*-axis. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture. Hypermethylation is represented by green colored bars and hypomethylation is represented by red colored bars.

Figure S4. DMP cluster analysis during various tissue culture stages in *M. truncatula* R108 genome in the context of CG, CHH, and CHG methylation. In each of the panels, cluster analysis is presented as a heatmap in CG, CHH, and CHG context. The bar chart in the inset showing 0–1 (blue to red progression) shows the methylation changes with blue being lowest and red being highest.

Figure S5. GO enrichment of DMR-related genes during various tissue culture stages in *M. truncatula* R108 genome. In different stages of tissue culture (AC and CON), presented, DMRs are analyzed for GO-related gene enrichment, and the genes representing various biological and metabolical processes are plotted on *X*-axis with numbers of the genes along with their percentages on the *Y*-axis. AC, after callus; CON, controls represent different stages of tissue culture. Significantly enriched GO terms are calculated based on *P*-values in hypergeometric tests with *P*-values less than 0.05.

Figure S6. GO enrichment of DMP-related genes during various tissue culture stages in *M. truncatula* R108 genome. (a) In different stages of tissue culture (AC and CON) presented, DMPs are analyzed for GO-related gene enrichment, and the genes representing various biological, cellular, and molecular function processes are plotted on *Y*-axis with numbers of the genes on the *X*-axis. (b) In different stages of tissue culture (AC and CON) presented, DMPs that are in hypermethylated regions are analyzed for GO-related gene enrichment, and the genes representing various biological, cellular, and molecular function processes are plotted on *Y*-axis with numbers of the genes on the *X*-axis. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture. Enrichment *P*-values were calculated using the hypergeometric test.

Figure S7. GO enrichment of DMR-related genes during various tissue culture stages in *M. truncatula* R108 genome. (a) In different stages of tissue culture (AC and CON) presented, DMRs are analyzed for GO-related gene enrichment, and the genes representing various biological, cellular, and molecular function processes are plotted on *Y*-axis and number of genes on the *X*-axis. (b) In different stages of tissue culture (AC and CON) presented, DMRs that are in hypermethylated regions are analyzed for GO-related gene enrichment, and the genes representing various biological, and molecular function processes are plotted on *Y*-axis and the number of genes on the *X*-axis. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture. The term with an asterisk is significantly enriched.

Figure S8. KEGG enrichment of DMR-related genes during various stages of tissue culture in *M. truncatula* R108 genome. (a) In different stages of tissue culture (AC and CON) presented, DMRs are analyzed for KEGG-related gene enrichment, and the genes representing various biological, cellular, and molecular function processes are plotted on *Y*-axis with richness factor of the genes on the *X*-axis. (b) In different stages of tissue culture (AC and BC) presented, DMRs are analyzed for KEGG-related gene enrichment, and the genes representing various biological, cellular and molecular function processes are plotted on *Y*-axis with richness factor of the genes on the *X*-axis. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture. Enrichment *P*-values were calculated using the hypergeometric test and the FDR correction method applied was Benjamini and Hochberg.

Figure S9. Correlation between DMRs and *Tnt1* insertion frequency in *M. truncatula* R108 genome between various stages of tissue culture. In each of the panels, DMRs are mapped to the corresponding regions where *Tnt1* FSTs are present. (a) AC versus CON. (b) BC versus AC. (c) BC versus CON. *Tnt1* insertion frequency is plotted on the *X*-axis and DMRs are plotted on the *Y*-axis. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture. *P*-value is based on hypergeometric tests with 95% confidence interval with odds ratio greater than 1 for total number of DMRs with genes of interest and *Tnt1* insertions.

Figure S10. Differential methylation at *Tnt1* insertion sites in *M. truncatula* R108 genome between various stages of tissue culture. In each of the panels, differential methylation sites are mapped to the corresponding *Tnt1* insertion sites and 50-bp flanking regions.

(a) AC versus CON. (b) BC versus AC. (c) BC versus CON. (d) Absolute methylation values in R108 corresponding to the *Tnt1* sites are shown as a boxplot. Differential methylated sites per stages in terms of sequencing context CG, CHG, and CHH is plotted on the *X*-axis and *Tnt1* insertion site numbers are plotted on the *Y*-axis. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture; Cyt, cytosine.

Figure S11. Methylation patterns in *M. truncatula* R108 genome between various stages of tissue culture. In each of the panels, methylation sites are mapped to the corresponding genomic positions. We have randomly chosen a gene to depict the differing methylation at different stages of tissue culture. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture; Cyt, cytosine.

Figure S12. Differential gene expression presented as volcano plots. Genes that are differentially expressed are presented as volcano plots with log-transformed values on *Y*-axis. All the values are log transformed and the legends represent the range of their expression. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture.

Figure S13. Correlation plots between gene expression and methylation. Genes that are differentially expressed are presented on *X*-axis and log-transformed methylation values are plotted on *Y*-axis. (a) Correlation plots for overall methylation versus gene expression between AC and CON stages. (b) Correlation plots for gene expression ranging from −10- to −25-fold against the methylation changes. (c) Correlation plots for gene expression ranging from 10- to 25-fold against the methylation changes. (d) Correlation plots for gene expression ranging from 2- to −2-fold against the methylation changes. AC, after callus; CON, controls represent different stages of tissue culture. Correlation plots calculated using *R*-statistics.

Figure S14. Cytosine methylation in the context of CG, CHH, and CHG methylation from the somatic embryogenesis samples. Cytosine methylation in the terms of three different contexts CG, CHH, and CHG are plotted on *X*-axis, while the % of methylation is plotted on the *Y*-axis.

Figure S15. Number of genes expressed as percent of methylation in all contexts of cytosine methylation in *M. truncatula* R108 genome during somatic embryogenesis (SE). In each of the panels, percent methylation is presented on *X*-axis and the number of genes with corresponding methylation on *Y*-axis. (a) CG methylation at SE. (b) CHG methylation at SE. (c) CHH methylation at SE. The values on *Y*-axis are transformed logarithmically to base 10.

Table S1. A list of all DMRs identified between tissue culture stages in *M. truncatula*. (a) DMR regions between tissue culture stages in *M. truncatula* R108 genome—AC versus CON. (b) DMR regions between tissue culture stages in *M. truncatula* R108 genome—BCs versus AC. (c) DMR regions between tissue culture stages in *M. truncatula* R108 genome—BC versus CON. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture.

Table S2. A list of all DMRs mapped to *M. truncatula* R108 HiC genome. (a) DMR regions mapped to R108 HiC assembly between AC versus CON. (b) DMR regions mapped to R108 HiC assembly between AC versus BC. (c) DMR regions mapped to R108 HiC assembly between BC versus CON. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture.

Table S3. A list of all DMPs mapped to *M. truncatula* R108 genome. (a) DMPs between stages AC versus CON in *M. truncatula* R108 genome. (b) DMPs between stages AC versus. BC in *M. truncatula* R108 genome. (c) DMPs between stages BC versus CON in *M. truncatula* R108 genome. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture.

Table S4. Enriched GO classification of gene counts in DMRs between tissue culture stages in *M. truncatula*. (a) Enriched GO classification gene count in DMRs between *M. truncatula* AC versus CON. (b) Enriched GO classification gene count in DMRs between *M. truncatula* BC versus CON. (c) Enriched GO classification gene count in DMRs between *M. truncatula* BCs versus AC. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture.

Table S5. Enriched GO classification of gene counts in between tissue culture stages in *M. truncatula*. (a) Enriched GO classification gene count in DMPs between *M. truncatula* AC versus CON. (b) Enriched GO classification gene count in DMPs between *M. truncatula* BC versus CON. (c) Enriched GO classification gene count in DMPs between *M. truncatula* BCs versus AC. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture.

Table S6. Enriched KEGG pathway classification of gene counts in DMRs between tissue culture stages in *M. truncatula*. (a) Enriched KEGG classification gene count in DMRs between *M. truncatula* AC versus CON. (b) Enriched KEGG classification gene count in DMRs between M. truncatula BC versus CON. (c) Enriched KEGG classification gene count in DMRs between *M. truncatula* BCs versus AC. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture.

Table S7. Enriched KEGG pathway classification of gene counts in DMPs between *M. truncatula* pre-embryogenesis callus formation stages. (a) Enriched KEGG classification gene count in DMPs between *M. truncatula* AC versus CON. (b) Enriched KEGG classification gene count in DMPs between *M. truncatula* BC versus CON. (c) Enriched KEGG classification gene count in DMPs between *M. truncatula* BC versus AC. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture.

Table S8. A list of *Tnt1* insertions correlating with DMRs in *M. truncatula* genome. (a) A list of all *Tnt1* insertions correlating with the DMRs in *M. truncatula* genome during AC versus CON comparison. (b) Statistics of all *Tnt1* insertions correlating with the DMRs in *M. truncatula* genome during AC versus CON comparison. (c) A list of all *Tnt1* insertions correlating with the DMRs in *M. truncatula* genome during BC versus AC comparison. (d) Statistics of all *Tnt1* insertions correlating with the DMRs in *M. truncatula* genome during BC versus AC comparison. (e) A list of all *Tnt1* insertions correlating with the DMRs in *M. truncatula* genome during BC versus CON comparison. (f) Statistics of all *Tnt1* insertions correlating with the DMRs in *M. truncatula* genome during BC versus CON comparison. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture.

Table S9. Differentially expressed genes using RNAseq in various stages of tissue culture in *M. truncatula*. (a) Differential gene expression studies between tissue culture stages CON and AC. (b) Differential gene expression studies between tissue culture stages BC and AC. (c) Differential gene expression studies between tissue culture stages CON and BC. (d) Summary statistics for FPKM values of RNAseq transcript levels in *M. truncatula* in comparison between tissue culture stages CON, BC, and AC. AC, after callus; BC, before callus; CON, controls represent different stages of tissue culture.

Table S10. Cytosine methylation in the context of CG, CHH, and CHG methylation in *M. truncatula* somatic embryogenesis. (a) Cytosine methylation in CG context in *M. truncatula* in comparison between non-azacytidine versus azacytidine treatments during

somatic embryogenesis. (b) Summary table of cytosine methylation in CG context in *M. truncatula* in comparison between non-azacytidine versus azacytidine treatments during somatic embryogenesis. (c) Summary statistics of cytosine methylation in CG context in *M. truncatula* in comparison between non azacytidine versus azacytidine treatments during somatic embryogenesis. (d) Cytosine methylation in CHG context in *M. truncatula* in comparison between non-azacytidine versus azacytidine treatments during somatic embryogenesis. (e) Summary table of cytosine methylation in CHG context in *M. truncatula* in comparison between non-azacytidine versus azacytidine treatments during somatic embryogenesis. (f) Summary statistics of cytosine methylation in CHG context in *M. truncatula* in comparison between non-azacytidine versus azacytidine treatments during somatic embryogenesis. (g) Cytosine methylation in CHH context in *M. truncatula* in comparison between non-azacytidine versus azacytidine treatments during somatic embryogenesis. (h) Summary table of cytosine methylation in CHH context in *M. truncatula* in comparison between non-azacytidine versus azacytidine treatments during somatic embryogenesis. (i) Summary statistics of cytosine methylation in CHH context in *M. truncatula* in comparison between non-azacytidine versus azacytidine treatments during somatic embryogenesis.

REFERENCES

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, **25**, 25–29.

Barker, D.G., Bianchi, S., Blondon, F., Dattée, Y., Duc, G., Essad, S. *et al.* (1990) *Medicago truncatula*, a model plant for studying the molecular genetics of the rhizobium-legume symbiosis. *Plant Molecular Biology Reporter*, **8**, 40–49.

Baubec, T., Pecinka, A., Rozhon, W. & Mittelsten Scheid, O. (2009) Effective, homogeneous and transient interference with cytosine methylation in plant genomic DNA by zebularine. *The Plant Journal*, **57**, 542–554.

Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bravo, S., Bertin, A., Turner, A., Sepulveda, F., Jopia, P., Parra, M.J. *et al.* (2017) Differences in DNA methylation, DNA structure and embryogenesis-related gene expression between embryogenic and non embryogenic lines of *Pinus radiata* D. don. *Plant Cell, Tissue and Organ Culture*, **130**, 521–529.

Burks, D., Azad, R., Wen, J. & Dickstein, R. (2018) The *Medicago truncatula* genome: genomic data availability. *Methods in Molecular Biology*, **1822**, 39–59.

Cheng, X., Krom, N., Zhang, S., Mysore, K.S., Udvardi, M. & Wen, J. (2017) Enabling reverse genetics in *Medicago truncatula* using high-throughput sequencing for *Tnt1* flanking sequence recovery. *Methods in Molecular Biology*, **1610**, 25–37.

Cheng, X., Wang, M., Lee, H.K., Tadege, M., Ratet, P., Udvardi, M. *et al.* (2014) An efficient reverse genetics platform in the model legume *Medicago truncatula*. *The New Phytologist*, **201**, 1065–1076.

Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Research*, **14**, 1188–1190.

d'Erfurth, I., Cosson, V., Eschstruth, A., Lucas, H., Kondorosi, A. & Ratet, P. (2003) Efficient transposition of the *Tnt1* tobacco retrotransposon in the model legume *Medicago truncatula*. *The Plant Journal*, **34**, 95–106.

Eprintsev, A.T., Fedorin, D.N., Dobychina, M.A. & Igamberdiev, A.U. (2017) Expression and promoter methylation of succinate dehydrogenase and fumarase genes in maize under anoxic conditions. *Journal of Plant Physiology*, **216**, 197–201.

Eprintsev, A.T., Fedorin, D.N., Gataullina, M.O. & Igamberdiev, A.U. (2020) Two forms of NAD-malic enzyme in maize leaves are regulated by light in opposite ways via promoter methylation. *Journal of Plant Physiology*, **251**, 153193.

Garmier, M., Gentzbittel, L., Wen, J., Mysore, K.S. & Ratet, P. (2017) *Medicago truncatula*: genetic and genomic resources. *Current Protocols in Plant Biology*, **2**, 318–349.

Gehring, M., Bubb, K.L. & Henikoff, S. (2009) Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science*, **324**, 1447–1451.

Gepts, P., Beavis, W.D., Brummer, E.C., Shoemaker, R.C., Stalker, H.T., Weeden, N.F. *et al.* (2005) Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. *Plant Physiology*, **137**, 1228–1235.

Grandbastien, M.A. (1998) Activation of plant retrotransposons under stress conditions. *Trends in Plant Science*, **3**, 181–187.

Grandbastien, M.A. (2015) LTR retrotransposons, handy hitchhikers of plant regulation and stress response. *Biochimica et Biophysica Acta*, **1849**, 403–416. https://doi.org/10.1016/j.bbagrm.2014.07.017

Grandbastien, M.A., Spielmann, A. & Caboche, M. (1989) *Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature*, **337**, 376–380.

Grzybkowska, D., Nowak, K. & Gaj, M.D. (2020) Hypermethylation of auxin-responsive motifs in the promoters of the transcription factor genes accompanies the somatic embryogenesis induction in Arabidopsis. *International Journal of Molecular Sciences*, **21**, 6849.

Gyulai, G., Kiss, E., Csillag, A. & Heszky, L.E. (1993) Developmental analysis of primary and secondary somatic embryogenesis in soybean tissue culture. *Acta Biologica Hungarica*, **44**, 189–196.

Hansen, K.D., Langmead, B. & Irizarry, R.A. (2012) BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biology*, **13**, R83.

Higo, A., Saihara, N., Miura, F., Higashi, Y., Yamada, M., Tamaki, S. *et al.* (2020) DNA methylation is reconfigured at the onset of reproduction in rice shoot apical meristem. *Nature Communications*, **11**, 4079.

Hirochika, H., Okamoto, H. & Kakutani, T. (2000) Silencing of retrotransposons in Arabidopsis and reactivation by the ddm1 mutation. *Plant Cell*, **12**, 357–368.

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.

Hu, L., Li, N., Zhang, Z., Meng, X., Dong, Q., Xu, C. *et al.* (2020) CG hypomethylation leads to complex changes in DNA methylation and transpositional burst of diverse transposable elements in callus cultures of rice. *The Plant Journal*, **101**, 188–203.

Iantcheva, A., Chabaud, M., Cosson, V., Barascud, M., Schutz, B., Primard-Brisset, C. *et al.* (2009) Osmotic shock improves *Tnt1* transposition frequency in *Medicago truncatula* cv. Jemalong during in vitro regeneration. *Plant Cell Reports*, **28**, 1563–1572.

Ji, L., Mathioni, S.M., Johnson, S., Tucker, D., Bewick, A.J., Do Kim, K. *et al.* (2019) Genome-wide reinforcement of DNA methylation occurs during somatic embryogenesis in soybean. *Plant Cell*, **31**, 2315–2331.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. & Madden, T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Research*, **36**, W5–W9.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, **44**, D457–D462.

Kaur, P., Lui, C., Dudchenko, O., Nandety, R.S., Hurgobin, B., Pham, M. *et al.* (2021) Delineating the *Tnt1* insertion landscape of the model legume *Medicago truncatula* cv. R108 at the hi-C resolution using a chromosome-length genome assembly. *International Journal of Molecular Sciences*, **22**, 4326.

Kawakatsu, T., Stuart, T., Valdes, M., Breakfield, N., Schmitz, R.J., Nery, J.R. *et al.* (2016) Unique cell-type-specific patterns of DNA methylation in the root meristem. *Nature Plants*, **2**, 16058.

Krueger, F. & Andrews, S.R. (2011) Bismark: a flexible aligner and methylation caller for bisulfite-Seq applications. *Bioinformatics*, **27**, 1571–1572.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D. *et al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome Research*, **19**, 1639–1645.

Langmead, B. & Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Law, J.A. & **Jacobsen, S.E.** (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews. Genetics*, **11**, 204–220.

Le, B.H., **Cheng, C.**, **Bui, A.Q.**, **Wagmaister, J.A.**, **Henry, K.F.**, **Pelletier, J.** ***et al.*** (2010) Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 8063–8070.

Lee, H.K., **Mysore, K.S.** & **Wen, J.** (2018) *Tnt1* insertional mutagenesis in *Medicago truncatula*. *Methods in Molecular Biology*, **1822**, 107–114.

Li, A., **Liu, A.**, **Wu, S.**, **Qu, K.**, **Hu, H.**, **Yang, J.** ***et al.*** (2022) Comparison of structural variants in the whole genome sequences of two *Medicago truncatula* ecotypes: Jemalong A17 and R108. *BMC Plant Biology*, **22**, 77.

Li, Q., **Gent, J.I.**, **Zynda, G.**, **Song, J.**, **Makarevitch, I.**, **Hirsch, C.D.** ***et al.*** (2015) RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 14728–14733.

Love, M.I., **Huber, W.** & **Anders, S.** (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.

Lu, X., **Wang, X.**, **Chen, X.**, **Shu, N.**, **Wang, J.**, **Wang, D.** ***et al.*** (2017) Single-base resolution methylomes of upland cotton (*Gossypium hirsutum* L.) reveal epigenome modifications in response to drought stress. *BMC Genomics*, **18**, 297.

Lucas, H., **Feuerbach, F.**, **Kunert, K.**, **Grandbastien, M.A.** & **Caboche, M.** (1995) RNA-mediated transposition of the tobacco retrotransposon *Tnt1* in *Arabidopsis thaliana*. *The EMBO Journal*, **14**, 2364–2373.

Mao, X., **Cai, T.**, **Olyarchuk, J.G.** & **Wei, L.** (2005) Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics*, **21**, 3787–3793.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **2011**(17), 200.

Masoudi-Nejad, A., **Goto, S.**, **Endo, T.R.** & **Kanehisa, M.** (2007) KEGG bioinformatics resource for plant genomics research. *Methods in Molecular Biology*, **406**, 437–458.

Matzke, M.A., **Kanno, T.** & **Matzke, A.J.** (2015) RNA-directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants. *Annual Review of Plant Biology*, **66**, 243–267.

Matzke, M.A. & **Mosher, R.A.** (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews. Genetics*, **15**, 394–408.

Meinke, D.W. (2020) Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in Arabidopsis. *The New Phytologist*, **226**, 306–325.

Melayah, D., **Bonnivard, E.**, **Chalhoub, B.**, **Audeon, C.** & **Grandbastien, M.A.** (2001) The mobility of the tobacco *Tnt1* retrotransposon correlates with its transcriptional activation by fungal factors. *The Plant Journal*, **28**, 159–168.

Mendez-Hernandez, H.A., **Ledezma-Rodriguez, M.**, **Avilez-Montalvo, R.N.**, **Juarez-Gomez, Y.L.**, **Skeete, A.**, **Avilez-Montalvo, J.** ***et al.*** (2019) Signaling overview of plant somatic embryogenesis. *Frontiers in Plant Science*, **10**, 77.

Moll, K.M., **Zhou, P.**, **Ramaraj, T.**, **Fajardo, D.**, **Devitt, N.P.**, **Sadowsky, M.J.** ***et al.*** (2017) Strategies for optimizing BioNano and Dovetail explored through a second reference quality assembly for the legume model, *Medicago truncatula*. *BMC Genomics*, **18**, 578.

Nic-Can, G.I., **Lopez-Torres, A.**, **Barredo-Pool, F.**, **Wrobel, K.**, **Loyola-Vargas, V.M.**, **Rojas-Herrera, R.** ***et al.*** (2013) New insights into somatic embryogenesis: LEAFY COTYLEDON1, BABY BOOM1 and WUSCHEL-RELATED HOMEOBOX4 are epigenetically regulated in *Coffea canephora*. *PLoS One*, **8**, e72160.

Okuda, S., **Yamada, T.**, **Hamajima, M.**, **Itoh, M.**, **Katayama, T.**, **Bork, P.** ***et al.*** (2008) KEGG atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Research*, **36**, W423–W426.

Pecrix, Y., **Staton, S.E.**, **Sallet, E.**, **Lelandais-Brière, C.**, **Moreau, S.**, **Carrère, S.** ***et al.*** (2018) Whole-genome landscape of *Medicago truncatula* symbiotic genes. *Nature Plants*, **4**, 1017–1025.

R Core Team. (2020) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing https://www.R-project.org/

Rigal, M. & **Mathieu, O.** (2011) A "mille-feuille" of silencing: epigenetic control of transposable elements. *Biochimica et Biophysica Acta*, **1809**, 452–458.

Santos, D.a.F., P. (2002) Loss of DNA methylation affects somatic embryogenesis in *Medicago truncatula*. *Plant Cell, Tissue and Organ Culture*, **70**, 155–162.

Sasaki, E., **Kawakatsu, T.**, **Ecker, J.R.** & **Nordborg, M.** (2019) Common alleles of CMT2 and NRPE1 are major determinants of CHH methylation variation in *Arabidopsis thaliana*. *PLoS Genetics*, **15**, e1008492.

Schenk, R.U. & **Hildebrandt, A.C.** (1972) Medium and techniques for induction and growth of monocotyledonous and dicotyledonous plant-cell cultures. *Canadian Journal of Botany*, **50**, 199.

Stroud, H., **Do, T.**, **Du, J.**, **Zhong, X.**, **Feng, S.**, **Johnson, L.** ***et al.*** (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nature Structural & Molecular Biology*, **21**, 64–72.

Sun, L., **Gill, U.S.**, **Nandety, R.S.**, **Kwon, S.**, **Mehta, P.**, **Dickstein, R.** ***et al.*** (2019) Genome-wide analysis of flanking sequences reveals that *Tnt1* insertion is positively correlated with gene methylation in *Medicago truncatula*. *The Plant Journal*, **98**, 1106–1119.

Tadege, M., **Ratet, P.** & **Mysore, K.S.** (2005) Insertional mutagenesis: a Swiss Army knife for functional genomics of *Medicago truncatula*. *Trends in Plant Science*, **10**, 229–235.

Tadege, M., **Wen, J.**, **He, J.**, **Tu, H.**, **Kwak, Y.**, **Eschstruth, A.** ***et al.*** (2008) Large-scale insertional mutagenesis using the *Tnt1* retrotransposon in the model legume *Medicago truncatula*. *The Plant Journal*, **54**, 335–347.

Tanabe, M. & **Kanehisa, M.** (2012) Using the KEGG database resource. *Current Protocols in Bioinformatics*, **1**(Unit1), 12.

Tian, R., **Paul, P.**, **Joshi, S.** & **Perry, S.E.** (2020) Genetic activity during early plant embryogenesis. *The Biochemical Journal*, **477**, 3743–3767.

Wollmann, H., **Stroud, H.**, **Yelagandula, R.**, **Tarutani, Y.**, **Jiang, D.**, **Jing, L.** ***et al.*** (2017) The histone H3 variant H3.3 regulates gene body DNA methylation in *Arabidopsis thaliana*. *Genome Biology*, **18**, 94.

Yaish, M.W., **Al-Lawati, A.**, **Al-Harrasi, I.** & **Patankar, H.V.** (2018) Genome-wide DNA methylation analysis in response to salinity in the model plant caliph medic (*Medicago truncatula*). *BMC Genomics*, **19**, 78.

Young, M.D., **Wakefield, M.J.**, **Smyth, G.K.** & **Oshlack, A.** (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, **11**, R14.

Young, N.D. & **Udvardi, M.** (2009) Translating *Medicago truncatula* genomics to crop legumes. *Current Opinion in Plant Biology*, **12**, 193–201.

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.

Zhang, H., **Meltzer, P.** & **Davis, S.** (2013) RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics*, **14**, 244.

Zhou, W., **Liang, G.**, **Molloy, P.L.** & **Jones, P.A.** (2020) DNA methylation enables transposable element-driven genome expansion. *Proceedings of the National Academy of Sciences of the United States of America*, **117**, 19359–19366.