DNA demethylation and hypermethylation are both required for late nodule development in Medicago

Yann Pecrix, Erika Sallet, Sandra Moreau, Olivier Bouchez, Sébastien Carrere, Jerome Gouzy, Marie-Françoise Jardinaud, Pascal Gamas

DNA demethylation and hypermethylation are both required for late nodule development in Medicago

Pecrix[1,2], Y., Sallet[1], E., Moreau[1], S., Bouchez[3], O., Carrere[1], S., Gouzy[1], J., Jardinaud[1], M-F., Gamas[1]*, P.

[1]LIPME, Université de Toulouse, INRAE, CNRS, Castanet-Tolosan, France ;

[2]CIRAD, UMR PVBMT, F-97455 Saint-Pierre, La Réunion, France

[3]INRAE, US1426, GeT-PlaGe, Genotoul, 31326, Castanet-Tolosan, France;

*Corresponding author: Pascal.Gamas@inrae.fr

Introductory paragraph

Plant epigenetic regulations are involved in transposable element (TE) silencing, developmental processes and responses to the environment[1–7]. They often involve modifications of DNA methylation, particularly through the DEMETER (DME) demethylase family and RNA-dependent DNA methylation (RdDM)[8]. Root nodules host rhizobia that can fix atmospheric nitrogen for the plant benefit in nitrogen-poor soils. The development of indeterminate nodules, as in *Medicago truncatula*, involve successive waves of gene activation[9-12], the control of which raises interesting questions. Using laser capture microdissection (LCM) coupled to RNAseq (SYMbiMICS data[11]), we previously identified 4,309 genes (termed NDD) activated in the nodule differentiation and nitrogen fixation zones, 36% of which belong to co-regulated genomic regions dubbed symbiotic islands[13]. We found *MtDME* to be upregulated in the differentiation zone and required for nodule development, and identified 474 differentially methylated regions (DMRs) hypomethylated in the nodule, by analyzing ~2% of the genome[4]. Here, we coupled LCM and whole-genome bisulfite-sequencing (WGBS) for a comprehensive view of DNA methylation, integrated with gene expression at the tissue level. Furthermore, by CRISPR-Cas9 mutagenesis of *MtDRM2,* we showed the importance of RdDM for CHH hypermethylation and nodule development. We thus proposed a model for DNA methylation dynamics during nodule development.

Main text

We first performed WGBS using *M. truncatula* nodules at 6 days post inoculation with *Sinorhizobium meliloti* (stage of maximal *MtDME* expression[4]) and nitrogen-starved non-inoculated root tips (RT) (two replicates; sequencing coverage of ~16 to 25x, supplemental Table 1). Indeterminate nodules consist of an apical meristem (or zone I), an infection and early differentiation zone (distal and proximal zone II), a late differentiation zone (interzone II-III), and a nitrogen fixation zone (zone III). To distinguish different developmental stages, we analyzed three laser-dissected nodule zones,

namely the meristematic (M), differentiation (Diff), and nitrogen-fixation (Fix) zones (Fig. 1a; three replicates; average BsSeq coverage of 5.7, 10.5 and 15.8x respectively; supplemental Table 1). BsSeq data showed a good sample reproducibility, and allowed two groups to be clearly distinguished, one corresponding to the differentiated nodule tissues (Diff and Fix) and the other to the whole organs and M zone (Fig. 1b). While little difference was observed between whole organs, an increase of CHH methylation was found in the Diff and Fix zones (Fig. 1c, 1d), on all chromosome regions (Fig. 1e), with an average methylation level of 12.0% and 15.1% respectively, vs. 6.8% in RT. Changes in DNA methylation in all three contexts CG, CHG, and CHH were then revealed by DMR analyses, when comparing Diff and Fix vs RT (Fig. 1f). The CHH DMRs (Fix vs RT) encompassed 22.8 Mb vs. only 0.82 Mb for the CG-CHG DMRs (5.3% and 0.2% of the nuclear genome, with an average size of 355 and 233 nt, respectively). The DMRs (shown with methylome data in the M.t 5.0 genome browser[13] https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/) were consistent with previously identified DMRs[4] and highly reproducible, whether comparing biological replicates, whole organ vs LCM samples, or Diff vs Fix zones. In conclusion, a strong dynamics of DNA methylation during the late stages of nodule development was uncovered.

At the gene level, in the CG context, average methylation profiles showed gene body methylation with a strong decrease at the transcription start (TSS) and termination (TTS) sites (Fig. 2a), as commonly observed in angiosperms[14]. While all samples were very similar when considering the average of all genes, gene body hypomethylation was observed in nodule samples when focusing on the 4,309 NDD genes. Such hypomethylation was not seen when considering the 4,309 genes most expressed in roots, or TE-related repeats (hereafter called TEs for simplification) (Fig. 2a). By contrast, in the CHH context, and the CHG context to a lower extent, a strong hypermethylation was observed for all genes and TEs in the Diff and even more the Fix samples, with a peak of hypermethylation in the 1kb region upstream the gene TSS (Fig. 2b and 2c).

CHG DMRs appeared more similar to CG DMRs than to CHH DMRs, both in terms of localization and pattern. Thus, 98.8% of CG DMRs, 85.5% of CHG DMRs and 99.5% CG-CHG DMRs were hypo-methylated in Fix vs RT, whereas all CHH DMRs were hyper-methylated (>99.9% in Diff and Fix vs RT), in all cases with a large difference of methylation level (mean absolute difference of 0.53 and 0.42 in the CG-CHG and CHH context, respectively) (supplemental Table 2). Many DMRs were close to genes, with 71.9% of CG-CHG DMRs and 47.9% of CHH DMRs overlapping with genes extended by 1 kb on each side. Moreover, whereas only 14% of CG-CHG DMRs overlapped with CHH DMRs, 75.9% of CG-CHG DMRs overlapped with 1kb-extended CHH DMRs, indicating their proximity (see a representative example in Extended Data Fig. 1). However, the CG-CHG DMRs preferentially targeted genes (Fig. 2d) whereas the CHH DMRs mostly targeted TEs (66.1%, 88.5% and 96.6% overlap with Tephra-annotated[13], EDTA[15]-annotated and TASR[16]-annotated TEs, respectively). Finally, CG-CHG DMRs were strongly enriched in NDD islands compared to their flanking regions, unlike in apical (NDA) and non-spatially regulated (NDN) islands (Fig. 2e). Thus, in summary, late nodule development is accompanied by hypomethylated CG-CHG DMRs centered on a limited number of genes, and hypermethylated CHH DMRs targeting TEs, with nearly half of them close to CG-CHG DMRs.

To further integrate gene expression (LCM-RNAseq) and methylation (LCM-BsSeq), we analyzed the DMR distribution amongst the 16 expression patterns previously defined within the nodule[13] (Fig. 3a-b). CG-CHG DMRs tightly positively correlated with genes expressed in the Diff and Fix zones (Patterns 5-11 and 14-16), maximally in Pattern 6 to 9 (47.5% of 1,218 genes with DMRs) (Fig. 3c; Extended Data Fig. 2). These hypo-DMRs are associated with key or potential actors in nodule development and activity, notably genes encoding 380 NCR peptides[9], leghemoglobins[17], transporters, calmodulin-like proteins, symbiotic immune response regulators (MtDNF2[18], MtSymCRK[19], MtNAD1[20], MtRSD[21]), redox control proteins and 508 long non-coding RNAs (supplemental Table 3). When comparing the regulation of NDD genes with and without CG-CHG DMRs (1,723 and 2,586 genes respectively), we found that gene induction in nodules was much

stronger for NDD genes with CG-CHG DMRs [median fold change (FC) vs root expression of 280.6 and 12.7, respectively; supplemental Table 4]. This was mostly due to a lower expression in roots (median root expression with and without DMR of 0.01 and 0.09 counts per million, respectively), correlating with a higher DNA methylation in roots (Extended Data Fig. 3). Finally, as previously observed[13], many NDD genes (particularly in expression patterns 5 to 10) presented a highly differential distribution of histone marks, with repressive H3K27me3 marks in roots and active H3K9ac marks in nodules (Fig. 3d), supporting complex epigenetic regulations in symbiotic islands.

In contrast to the CG-CHG DMRs, the abundance and distribution of CHH DMRs strongly differed between the Diff and Fix zones (Fig. 3c). In the Diff zone, CHH DMRs were located next to 11.7% of genes, preferentially those showing CG-CHG DMRs (Pearson correlation of 0.81; P-value< 0.001), whereas, in the Fix zone, they were next to 45.9% of all genes, whatever their expression profile. An attractive hypothesis is therefore that hyper CHH DMRs are generated first next to hypo CG-CHG DMRs in the Diff zone and then spread genome-wide. Two pathways have been reported in Arabidopsis for CHH methylation, involving RdDM and the CMT2 chromomethylase (in a DDM1-dependent process), respectively[8]. *CMT2* and *DDM1* genes are weakly expressed in the Diff and Fix zones, in contrast to RdDM genes (supplemental Table 4), making RdDM the best candidate pathway for CHH hypermethylation. We tested this assumption using mutants of DRM2, the main DNA methylase in RdDM. We identified three *M. truncatula* homologues of AtDRM2 with a good conservation of its reported active sites[22], termed MtDRM2, MtDRM2L1 (DRM2-like1) and MtDRM2L2 (Extended Data Fig. 4). We then performed multi-guide CRISPR-Cas9 mutagenesis of both *MtDRM2* and *MtDRM2L2* (the second best expressed family member in nodules), using hairy root transformation. WGBS was carried out with pooled DNA from nodules showing various mutations in *Mtdrm2*/*Mtdrm2l2* (hereafter termed *drm;* Extended Data Fig. 5), using nodules poorly affected in size, thereby decreasing confounding effects due to developmental problems. Control samples consisted of nodules transformed with a CRISPR construct targeting the *GUS* gene. CG methylation was mostly unaffected in *Mtdrm* mutant nodules (17 DMRs hyper methylated in CRISPR-*GUS* vs *drm*

nodules). By contrast, CHH (Fig. 3e) and to a lower extent CHG hypermethylation were strongly decreased (respectively 9,951 and 391 DMRs hypermethylated in both replicates of CRISPR-*GUS* vs *Mtdrm* nodules with a difference in absolute methylation level >0.3, vs. no hypermethylated DMR in *Mtdrm* nodules) (supplemental Table 5). This implied that RdDM is involved in CHH and part of CHG hypermethylation in nodules, which was supported by 87.4% of CHH DMRs overlapping with 24nt siRNA clusters (71.9%, 91.4% and 97.7% of which co-localize with Tephra-, EDTA- and TASR-annotated TEs, respectively). We set up a complementation experiment where the CRISPR-*DRM* construct was co-transformed with a non-editable synthetic *MtDRM2* cDNA (termed *DRM2-R*) expressed from a pAt*UBI* promoter. WGBS showed that *DRM2-R* allowed CHH hypermethylation of genes and TEs to be recovered in nodules (Fig. 3e; only 186 DMRs hypermethylated in CRISPR-*GUS* vs *Mtdrm/DRM2-R*).

To assess the impact of RdDM on nodule development, we examined the phenotype of nodules induced by *S.meliloti nifH:GFP* (a marker of late rhizobium differentiation) (Fig. 3f-g). While the nodule number was not significantly modified (Extended Data Fig. 6a), *nifH:GFP* expression was affected in 50% of CRISPR-*DRM* nodules (n=114), vs 4.5% of CRISPR-*GUS* (n=89) and 3.2% of CRISPR-*DRM pUBI:DRM2-R* (n=31) nodules. In the most severe cases (about 25% of independently transformed roots), small, round-shaped and white nodules were formed with very low *nifH:GFP* expression and nitrogenase activity (based on an acetylene reduction assay) (Fig. 3g; Extended Data Fig. 6b; Extended Data Fig. 7). As for WGBS, we analyzed the transcriptome of *Mtdrm* nodules slightly affected in size. Similar number of genes were up-and down-regulated vs CRISPR-*GUS* nodules (1,011 and 913 genes respectively, FC>2, FDR<0.05), while more TEs were up- than downregulated in *drm* nodules (282 and 91 TEs respectively, FC>2, FDR<0.05), consistent with an increased transcription of less methylated TEs. Down-regulated genes, including 83 *NCR* genes, were predominantly found in patterns 8 to 11, as strongly upregulated genes (FC>4) (supplemental Table 6). This indicated a particular impact of non-CG methylation on transcriptional regulation in the late differentiation zone. Of note, *MtROS1* and *MtCMT2* (DNA demethylase and CHH methylase genes, respectively) were

down- and up-regulated in the mutant respectively, possibly as compensation for the loss of DRM2[8,23,24]. Most Pattern 8 to 11 genes that were down and upregulated in *drm* nodules had a CHH DMR (Fix vs RT) within 1kb (85.1% and 67.6% respectively, supplemental Table 7), suggesting that the methylation status of TEs may either positively or negatively impact the expression of certain Diff/Fix genes.

Two non-exclusive hypotheses could explain an increase in RdDM during late nodule development: (i) an increased expression of TEs, leading to increased siRNA production; (ii) a more conducive environment due to chromatin relaxation[25]. A re-analysis of SYMbiMICS data (see Methods) revealed 1,923 nodule-expressed TEs, with 1,321 of them expressed in the Diff/Fix zones, enriched in type I TEs (43.2% vs 34% for all chromosomal TEs) (supplemental Table 7). These Diff/Fix TEs showed a high induction level (up to a median 47-fold increase vs zone 1 in Pattern 9) and were often close to NDD genes (48.1% within 1 kb). Visual inspection of RNAseq data on the Mt5.0 genome browser suggested that TEs could be expressed either autonomously or by read-through transcription from strongly expressed genes (see an example in Extended Data Fig. 8). A higher fraction of Diff/Fix TEs overlapped with CHH DMRs compared to the total TE population (44.7% vs. 23.3%, respectively, with a mean overlap length of 761 vs. 228 nt). Thus, an increased TE expression is a likely cause of genome-wide hypermethylation, but probably not the only one. To assess our second hypothesis, i.e. a possible decondensation of heterochromatin, we examined DAPI-stained nodule sections by confocal microscopy. We observed enlarged nuclei in the Diff/Fix zones, but their chromocenters remained well visible (Extended Data Fig. 9), which did not support an extensive decondensation of pericentromeric heterochromatin. Decondensation of facultative heterochromatin in chromosome arms remains however conceivable, all the more since we found that all eight *M. truncatula H2A.W* genes encoding a histone 2 variant involved in heterochromatin compaction[26] are strongly down-regulated in the nodule interzone and zone III (supplemental Table 4).

In conclusion, while we previously uncovered (i) the importance of DNA demethylation for Medicago nodule development[4] and (ii) symbiotic islands enriched with differential mCHH and histone marks in nodules versus roots[13], here we showed the extent and dynamics of DNA methylation in the nodule, by coupling LCM and WGBS. We established that CG-CHG demethylation is restricted to genes in a few genomic regions, including symbiotic islands, in contrast to CHH hypermethylation, that affects TEs first next to CG-CHG DMRs and then throughout the genome. This hypermethylation is RdDM-mediated and important for nodule development. In Arabidopsis, RdDM-dependent hyper-CHH DMRs have been described in male sexual-lineage cells[27], the embryo[28] and root meristematic cells[25] but are unusual in differentiated somatic cells. The CHH hypermethylation reported here is more reminiscent of that reported for miRNA genes in soybean nodules[29] and the so-called mCHH islands found in maize and other plant species next to highly expressed genes, and proposed to act as barriers between euchromatin and heterochromatin to prevent TE expression[14,30].

In our model for DNA methylation dynamics in *M.truncatula* nodules (Fig. 4), a subset of late symbiotic genes are found in genomic regions that are strongly repressed in the root by DNA methylation and repressive histones. A combination of CG-CHG demethylation and histone modifications would make these regions accessible to the transcriptional machinery in the nodule differentiation zone. This would lead to transcriptional induction of Diff/Fix genes and nearby TEs, either from their own promoters or by read-through transcription from strongly expressed upstream genes. This would trigger an increased siRNA production and RdDM, as a plant protection against TE activation, progressively targeting all related TEs. In our model, TE hypermethylation would thus be primarily a consequence of TE transcription, as well as possibly relaxation of facultative heterochromatin. Since the methylomes of the nodule meristem and Diff/Fix zones are very different (Fig. 1b-c), the previously discussed hypothesis[4] that siRNAs produced in the Diff/Fix zone might migrate to the meristem to protect it against TE activation seems unlikely. In any case, RdDM is important for optimal expression of late symbiotic genes, possibly to decrease interference between

TE and gene expression. Taken together, our data provide new insights on the complex epigenetic landscape regulating gene expression during nodule development and nitrogen fixation.

Methods

Plant growth and inoculation

Non-transformed and *Agrobacterium rhizogenes*-transformed *M. truncatula* cv Jemalong A17 plants were grown in aeroponic caissons as described[31], with the following chamber conditions: temperature: 22°C; 75% hygrometry; light intensity: 200 $\mu E.m^{-2}.s^{-1}$; light-dark photoperiod: 16h-8h. Plants were grown for about seven days in caisson growth medium supplemented with 5 mM NH_4NO_3, then nitrogen-starved for three days, just before rhizobium inoculation. Plants were inoculated with 10 mL of a suspension (OD_{600nm} =1) of *Sinorhizobium meliloti* 2011 (for 10 L of plant growth medium), containing the pXLGD4 *hemA:lacZ* plasmid (GMI6526)[32], grown at 28°C on TY solid medium (Bacto Tryptone 5 $g.L^{-1}$ (Becton Dickinson), Yeast Extract 3 $g.L^{-1}$ (Duchefa Biochemie), Agar-Agar 15 $g.L^{-1}$, pH 7) supplemented with 10 $\mu g.mL^{-1}$ of tetracycline and 6 $mmol.L^{-1}$ calcium chloride. Following *S. meliloti* inoculation, plants were grown in nitrogen-free medium. Under these conditions, nodules became pink and nitrogen-fixing at about 7 dpi (non-transformed plants) or 10 dpi (composite plants transformed by *A. rhizogenes*).

Root transformation

Root transformation was carried out using *Agrobacterium rhizogenes* ARqua1 as described[33], except that plants were kept at 20°C for three weeks after transformation. Transformed roots were selected by kanamycin (25 mg/L) and systematically checked for the expression of a *DsRed* reporter gene present on the T-DNA construct. The nodulation phenotype of composite plants was analyzed following aeroponic growth in caissons.

RNA and DNA extraction

For BsSeq or RNAseq analyses of transformed roots/nodules, at least six independently transformed root systems were used per replicate, and at least 15 nodules per root system. For RNAseq analyses

(three biological replicates per sample) RNA was extracted using the Qiagen RNeasy Plant Mini Kit (Qiagen) with RNAse-free DNAseI (Qiagen) treatment following the manufacturer's procedure. RNA was quantified using a Nanodrop Spectrophotometer ND-100 (NanoDrop Technologies, Wilmington, DE, USA) and analyzed with a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA).

For bisulfite sequencing, DNA was prepared using the Qiagen DNA easy kit or as described for high molecular weight plant genomic DNA[34]. It was quantified using a Qubit fluorometer (Invitrogen) and analyzed with a Fragment Analyzer (Agilent).

Laser micro-dissection of nodule zones

Laser microdissection of 15 dpi nodules was carried out as described[11,35], using 24 µm nodule sections and four biological replicates. Rep 1 was only used to set up conditions for the whole procedure. DNA was extracted with the QIAmp micro DNA kit (QIAGEN), following the recommended procedure (« Isolation of Genomic DNA from laser-microdissected tissues »). DNA was quantified using a Qubit with the Qubit dsDNA HS Assay, and then pooled from 20 (rep 2 and 4) or 21 (rep 3) nodules before library production. The amount of pooled DNA was: 81, 87, 99 ng for the dissected meristematic zone (rep 2, 3, 4 respectively); 203, 218, 171 ng for the differentiation zone; 342; 361, 360 ng for the fixation zone (including *S. meliloti* DNA).

Bisulfite sequencing and methylome data analyses

WGBS of RT, N6, M, Diff and Fix samples was performed at the GeT-PlaGe core facility, INRAE Toulouse (http://www.get.genotoul.fr). WGBS libraries were prepared according to Swift Biosciences's protocol using the Accel-NGS Methyl-Seq DNA Library Kit for Illumina Sequencing. Briefly, DNA was fragmented by sonication and bisulfite conversion was performed using the EZ DNA Methylation-Lightning Kit (ZYMO Research) following manufacturer's recommendations. Sample purifications were performed using SPRI select magnetic beads. Then, adaptors were ligated and sequencing tags were added by PCR (10 PCR cycles). Library quality was assessed using a Fragment Analyzer (Advanced Analytical Technologies, Inc.) and libraries were quantified by Q-PCR using the Kapa Library Quantification Kit. Sequencing was performed using an Illumina HiSeq3000 (paired end

150 bp) and the Illumina HiSeq3000 Reagent Kits. WGBS of transformed root tips, CRISPR-*GUS,* CRISPR-*DRM* and *DRM2-R* samples was performed by BGI Hong-Kong, using the EZ DNA Methylation-Gold kit (ZYMO Research) and a HiSeq Xten (Illumina) platform (paired end 150 bp).

For each library, the raw BSseq reads were processed with Methylpy v1.4.1 using bowtie2 (v.2.3.4.3) as aligner with Methylpy embedded default parameters "-X 1000 -k 2 --no-mixed --no-discordant –sensitive", and MtrunA17 R5.0[13] genome for mapping. PCR duplicates were removed with MarkDuplicates (Picard suite v2.18.1). The methylation level of each cytosine was then computed with Methylpy v1.4.1[36]. The non-conversion rate of each sample was obtained by analyzing the rate of methyl cytosines in the chloroplast genome (MtrunA17CP). The clustering analysis of Fig. 1b was performed with methylKit[37] (R package, version 1.10), with a minimal coverage of ten for cytosines, correlation-based distances and Ward's method. Identification of differentially methylated sites (DMS) and differentially methylated region (DMR) calling were performed with Methylpy, independently for CG, CHG, CG-CHG and CHH contexts, only considering cytosines covered by at least four reads. DMS at a distance of less than 250 nt from each other were collapsed into DMRs. We also tested distances of 50 nt, 150 nt and 350 nt, with qualitatively similar results. Were retained the DMRs with FDR<0.05, containing at least 4 DMS, found in at least two replicates of each sample, and with a minimal difference of absolute methylation level of 0.1 for CHH, 0.2 for CHG, 0.3 for CG-CHG and 0.4 for CG[38].

The relative position of DMRs with other DMRs, genes or TEs were computed using Bedtools v2.30.0 (command "intersect" and "closest"). The relative proportion of methylated cytosines (Fig. 1c) was determined using a betabinomial test and considering cytosines with a minimal sequencing coverage of four. Representations of cytosine methylation level and DMR density were obtained using Circos[39] (http://circos.ca) (Fig. 1e), ViewBS[40] (v 0.1.9, using default parameters; Fig. 2a-c, Fig. 3e), and DANPOS2[41] (bin size set to 200 bp; Fig. 2d-e, Fig. 3c-d, Extended Data Fig. 3).

Despite the fact that very little LCM DNA remained and that whole organs dilute the signals, five CHH-DMRs could be independently validated by Chop-PCR[42], using DNA from LCM-BsSeq analyses for

one of them, and newly prepared genomic DNA from root tips and 15 dpi nodules for the others (Supplemental Table 8). For those, DNA was prepared by CTAB extraction from three biological replicates, consisting of pools of 50 nitrogen-starved root tips from six plants and pools of 303, 434 and 357 nodules from five, ten and eight plants, respectively. One µg of DNA was digested with either NlaIII, AluI or DdeI for one hour at 37°C according to manufacturer's procedure (ThermoFisher). The control DNA consisted of one µg of DNA treated under the same conditions without restriction enzyme. Quantitative PCR reactions were performed on a Roche Light Cycler 480 using the Light Cycler Fast Start DNA Master SYBR Green I kit according to manufacturer's instructions (Roche). Cycling conditions were as follows: 95°C for 10 min, 50 cycles at 95°C for 5 sec, 60°C for 5 sec, and 72°C for 15 sec.

Transposable element analyses

A two-step protocol based on the mapping of siRNA reads was used to complement the published structural annotation of TEs (Tephra pipeline[13]). In a first step the TASR pipeline[16] was run on the genome using 24nt siRNA reads extracted from sRNA datasets previously listed[13] (excluding A17 mutant datasets) and from a Stem and Leaves dataset downloaded from NCBI (GEO Accession GSE13761). siRNAs of 24nt with at least 10 reads among all datasets were selected.

The filtered siRNA dataset used as input of TASR contains 1,652,379 unique siRNA sequences with a median number of sequenced reads of 18 (mean 38). TASR 1.1 pipeline was run with parameters: -nsirna 4 -win 100 -minlen 80 -maxlen 20000 -idclust 0.9 -overlapclust 0.8 -cnumber 2 -idenblast 90 -evalue 1e-200 -usearchv usearch8.0.1623_i86linux64. The perl code was marginally modified to set –K bowtie parameter to 10000 and to disable centroid selection (execution failure in some families). Then, bedtools v2.27.1 was run to merge close repeat regions (merge –d 100). The TASR-based annotation identified 13,309 regions spanning 69,918,880 nucleotides (minimum, maximum and mean region lengths are 397, 204337, 5253 nucleotides, respectively).

The second step of the protocol aimed at increasing the sensitivity of the TE annotation. The bank previously built by TASR was used as input repeat library of RepeatMasker 4.1.0, run with

parameters: -gff -s -no_is -norna -nolow -e rmblast. Then, bedtools v2.27.1 (merge –d 100) was run to merge close regions. The final dataset (labeled TASR10-round2-RepeatMarker in the genome browser) identified 257,443 regions spanning a total of 218,292,089 nucleotides (minimum, maximum and mean region lengths are 25, 318,396, 847 nucleotides, respectively).

The two annotations substantially overlapped, with 93.2% Tephra-annotated repeats intersecting with TASR -annotated repeats. 86.9% of Tephra-annotated repeats were also found to overlap with TEs annotated with the EDTA pipeline[15]. All bioinformatic studies were performed with the published Tephra-annotated TEs[13] but TASR- and EDTA-annotated TEs can be found in the Mt5.0 genome portal (downloads section) and browser (https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/).

RNAseq data analysis

To evaluate TE expression in nodules, SYMbiMICS data[11] were analyzed with the pipeline used for the MtExpress gene expression atlas[43] and *M. truncatula* annotation version 5.1.8. nf-core/rnaseq pipeline version 3.0 (doi:10.5281/zenodo.4323183) was used with the following parameters " --skip_alignment --pseudo_aligner salmon ", followed by transcript assignation and quantification with salmon (version 1.4.0). The chosen expression threshold for retained genes and TEs was one CPM (count per million reads) for the sum of 15 libraries counts (i.e. three replicates of five nodule zones). Normalization was performed using trimmed mean of *M* values method[44]. Differentially expressed genes and TEs were detected with EdgeR Bioconductor package[45] version 3.34.0, using the GLM (Fitted generalized linear models) likelihood ratio test, with an FDR adjusted p-value[46]. Analyses of GO term enrichment were performed using the topGO package version 2.44.0[47].

Expression patterns (termed RG-Patterns for repeat and gene patterns), including 17,406 and 1,923 nodule-expressed genes and TEs, were defined for genes and TEs differentially regulated between at least two nodule zones (FDR<0.01 and LFC >1; supplemental Table 7), similarly to the 16 expression patterns previously defined from genes only[13]. The relative position of expressed TEs and genes was analyzed using Bedtools (v2.30.0).

siRNA distribution was analyzed using Shortstack v. 3.8.5[48], with siRNA clusters independently defined for different siRNA sizes (namely, 21, 22, 21-22 and 24 nt), using siRNAs from N0, N4, N6, N10 siRNA libraries previously generated[13]. The position of siRNA clusters vs DMRs, genes and TEs was analyzed using Bedtools v2.30.0.

CRISPR-Cas9 mutagenesis, genotyping and complementation assay

The guide RNAs (gRNA; listed in supplemental Table 9) were designed with CRISPOR (version 4.8, http://crispor.tefor.net/) program, with the INRA A17r5.0 r1.6 *M. truncatula* genome release[13] and the "20bp-NGG-Sp Cas9, SpCas9-HF1, eSpCas9 1.1" option (appropriate for the use of *Streptococcus pyogenes* Cas9). T-DNAs contained the *S. pyogenes Cas9* coding DNA sequence, to which a SV40 NLS sequence was added at the C-terminus, as well as three gRNAs, interspaced by tRNAs as described[49], and preassembled as a synthetic polycistronic gene. The guides were expressed under the control of *M. truncatula U6.1* (MtrunA17_Chr3g0136831) and *U6.6* (MtrunA17_Chr7g0251721) RNA Pol III-controlled promoters. Retained guides did not contain a TTTT stretch and did not show a perfect match with potential off target genes in the 12 nt following the NGG Protospacer Adjacent Motif (PAM). The backbones plasmids required for the assembly of the binary vectors were provided by the ENSA project (Engineering Nitrogen Symbiosis for Africa; https://www.ensa.ac.uk) and the cloning strategy was based on the Golden Gate cloning technology[50]. T-DNAs included a kanamycin resistance module (p*35S:KanR*:T*Nos*) and a DsRed fluorescent reporter module (p*AtUbi10*:*DsRed*:T*Ocs*), located respectively close to the right and left borders in order to counter-select partial insertions of T-DNAs.

Genotyping was done on individual transformed root systems, by extracting DNA from nodules or a root segment adjacent to nodules. Nested PCR was performed, followed by agarose gel electrophoresis analysis and systematic sequencing of PCR products.

For the *drm* complementation assay, a synthetic *MtDRM2* cDNA (ProteoGenix, Schiltigheim, France) was designed, in which all guide recognition sequences were mutated (with conservation of encoded amino acids). The synthetic construct (*MtDRM2-R*, 3533 nt; supplemental Note 1) included the *AtUBI*

promoter and the *35S* terminator, flanking the cDNA and sequences required for Golden Gate cloning. The whole module was inserted into the plasmid used for CRISPR-Cas9 *DRM2*/*DRM2L2* mutagenesis. Samples used for phenotyping and WGBS were checked to have bi-allelic edits of *MtDRM2* and *MtDRM2L2*, and no edits of the *MtDRM2-R* cDNA.

Gene and protein sequence analyses

The phylogenetic tree was generated using Phylogeny.fr [51] (PhyML/OneClick). Multiple alignment analyses were performed using Multalin[52] version 5.4.1. Correspondences between gene names and Mt5.0 or Mt4.0 gene identifiers were obtained using Legoo[53] and the Downloads section of Mt5.0 genome browser, as of 20211025.

Root nodule analysis and phenotyping

To analyze wild type nodule nuclei and chromocenters, 15 dpi nodules were fixed in 0.2% paraformaldehyde, then embedded in 8% low-melting agarose (NuSieve™ GTG™ Agarose Lonza); 80 µm sections (microtome LEICA VT 1000S) were stained with 0.5µg/ml DAPI (4', 6-diamidino-2-phenylindole) and observed using a confocal microscope Leica SP8 (excitation 415 nm; emission 411-478 nm).

For mutant vs control nodule phenotyping, root segments nodulated with *S.meliloti* 2011 *nifH:GFP* were harvested at 15-18 dpi from four independent caissons (in two independent experiments). Nodules were visually scored for their size, shape and color. Acetylene reduction assays were then performed, with a 4 hour incubation followed by quantification of ethylene production using an Agilent 7820A gas chromatograph. All nodulated roots were then observed with bright field and fluorescence imaging, using a stereo microscope Leica DFC7000T. They were given a GFP signal score from 0 to 4, corresponding to 0%, ≤25%, 25 to 50%, 50 to 75% and >75% GFP+ nodules, respectively. A fraction of the nodules was then used for DNA and RNA extraction, while another fraction was fixed in 1.5% glutaraldehyde, and embedded in 5% low-melting agarose. Nodule sections (60 µm thick) were prepared from 41 transformed roots (in two independent experiments), with ~10 to 20

nodules per root, and observed using a Zeiss Axiophot light/fluorescence microscope, following staining with 0.5µg/ml DAPI.

Data availability:

Raw reads from BsSeq and RNAseq experiments have been deposited at the Sequence Read Archive (SRA) (project accession numbers: SRP355902 and SRP349933). Data related to gene annotation, TE annotation (Tephra-based, TASR10_round2_RepeatMasker–based and EDTA-based), methylome and DMR analyses are available at the *M. truncatula* genome portal and browser https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/

Correspondence and requests for material should be addressed to PG.

Acknowledgements

We thank the ENSA project for providing plasmid backbones for Golden gate cloning and *M. truncatula* root transformation. This work was supported by the ANR grants EPISYM (ANR-15-CE20-0002) (PG) and Laboratoire d'Excellence (LABEX) TULIP (ANR-10-LABX-41) (PG, JG), and made use of data previously generated in the ANR SYMbiMICS (ANR-08-GENO-106) (PG) and the INRA SPE "EPINOD" projects (PG). The sequencing platform was supported by France Génomique National infrastructure (ANR-10-INBS-09) (OB). We are grateful to the Genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) for providing computing and storage resources.

Author contributions

Y.P., M-F.J. and P.G. conceived the research plans; Y.P., S.M., and M-F.J. performed most of the experiments; O.B. performed bisulfite sequencing; E.S., S.C., and J.G. performed bioinformatics analyses; Y.P., M-F.J. and P.G. analyzed the data; P.G. conceived the project and wrote the article with contributions of J.G., Y.P. and M-F.J.

Competing financial interests

The authors declare no competing financial interests

Figure Legends

Fig. 1. Coupling laser capture microdissection (LCM) to whole-genome bisulphite sequencing (WGBS) strongly increases the sensitivity of detection of differentially methylated DNA regions (DMRs).

a, representative image (out of 61 microdissected nodules, in three independent replicates) of a nodule following LCM of meristematic (M), differentiation (Diff) and nitrogen fixation (Fix) nodule zones; bar= 100 µm. **b**, methylome-based clustering analysis of WGBS libraries, with three biological replicates of laser-dissected M, Diff and Fix zones, as well as two biological replicates of N-starved root tips (RT) and whole nodules six days-post-inoculation with *Sinorhizobium meliloti* (N6). **c**, mean methylation level in the CG, CHG and CHH contexts of genomic DNA from RT, N6, M, Diff and Fix samples, with a level of 1 corresponding to 100% of methylated cytosines. **d**, relative proportion of cytosine methylation levels in mCG, mCHG and mCHH contexts (minimal coverage of 10 reads) in RT and Fix. **e**, distribution on the eight *M. truncatula* chromosomes of methylated cytosines and DMRs (Fix vs RT), as well as genes, transposable elements (TEs) and symbiosis related islands (SRIs) expressed in the nodule differentiation zone. Centromers and SRIs are depicted by thick black lines and grey lines, respectively. **f**, number of DMRs detected in the CG, CHG and CHH contexts, with pairwise comparisons of whole organs (N6 and RT) or laser-dissected samples (M, Diff, Fix); indicated DMRs were found in at least 2 biological replicates of each comparison, with a minimal difference of absolute methylation level of 0.4 for CG, 0.2 for CHG and 0.1 for CHH contexts. H= A, C or T nucleotide.

Fig. 2. Average methylation patterns of genes and transposable elements (TEs), in whole organs and laser-dissected nodule differentiation and nitrogen-fixation zones, and DMR localization.
a, **b**, **c**, CG, CHG and CHH contexts, respectively. From left to right, average methylated cytosine (mC) profile of all annotated genes (50,773 genes), of NDD genes (4,309 genes upregulated in the nodule differentiation and fixation zone), of the top 4,309 genes expressed in roots, and of small TEs (215,171 TEs <1 kb; 89% of total Tephra-annotated TEs). Two biological replicates of root tips (RT), 6 dpi nodules (N6) and laser-dissected nodule differentiation (Diff) and fixation (Fix) zones. Flanking regions of 2 kb for genes and 1 kb for TEs. Profiles correspond to Cs shared between all samples, with a minimal coverage of four reads.
d, average distribution of differentially methylated regions (fixation zone vs root tips) (DMRs) on NDD genes.
e, average CG-CHG DMR density in the 211 symbiotic differentiation islands (NDD; 1,558 genes) vs. 49 apical (NDA; 242 genes) and 57 non-spatially regulated (NDN; 275 genes) control islands[13] (island underlined, with length normalized to 50 kb and 50 kb flanking regions)
TSS= transcription start site. TTS= transcription termination site.

Fig. 3. Localization of DMRs and histone marks in relation to gene expression profiles, and importance of *MtDRM2* for nodule CHH hypermethylation and late nodule development.
a, zonation of the *M. truncatula* nodule (ZI: zone I, meristematic region; ZIId: distal zone II, (pre)infection; ZIIp: proximal zone II, early differentiation; IZ: interzone II-III, late differentiation; ZIII: zone III, nitrogen-fixation) and *pDME:GUS* expression (blue signal around IZ) (n=30 in three replicates). **b**, the 16 expression patterns previously defined[3], with strong and gradual differences between nodule zones for patterns 1-11 and 12-16, respectively; differentiation zone: patterns 5-10 and 14-15; fixation zone: patterns 11 and 16. **c**, distribution of CG-CHG and CHH DMRs at the transcription start site (TSS) and 2 kb flanking regions of all 16 pattern genes and non-spatially regulated (NDN) genes, observed for the laser-dissected differentiation (Diff) and nitrogen-fixation (Fix) zones. **d**, distribution of active H3K9ac and repressive H3K27me3 histone marks on gene bodies (normalized to 2kb) and 2 kb flanking regions, in root tip and whole nodule samples[13]. In c and d,

genes were ranked first by expression pattern and then by decreasing expression level. **e**, Average CHH methylation of all genes and small transposable elements (215,171 TEs<1 kb), in *drm* mutant nodules (CRISPR-Cas9 mutagenesis of *MtDRM2* and *MtDRM2L*) compared to control nodules transformed with a *GUS*-targeting CRISPR-Cas9 construct and *drm* mutant nodules complemented with a synthetic non-editable *MtDRM2* cDNA (*MtDRM2-R*). Two biological replicates from whole-genome bisulfite sequencing are shown. **f**, **g**, **h**, sections of *S. meliloti nifH:GFP*-induced nodules, transformed with a *GUS*-targeting CRISPR-Cas9 construct (f), or *MtDRM2* and *MtDRM2L* genes (g, h), with in addition *MtDRM2-R* cDNA in h; observed using light and fluorescence microscopy (left and right panels, respectively). f, g, h: 150, 200 and 40 sectioned nodules from 9, 9 and 4 independently transformed roots, respectively. Note the strong GFP signal in the ZIII of the control nodule and the mutant nodule complemented with *MtDRM2-R*, whereas in *drm* mutant nodules the GFP signal is either absent (g top: round-shaped nodule) or present in only a few cells (g bottom: elongated nodules). Bars = 100 µm.

Fig. 4. Model for the dynamics of DNA methylation during nodule development.
a, In the root, genes expressed in the nodule differentiation zone and transposable elements (TEs) are silenced, with cytosine methylation in all contexts (mCG, mCHG, mCHH: violet, blue and yellow solid circles, respectively), and repressive histone marks (H3K9me2 and H3K27me1 for heterochromatic TEs; H3K27me3 for genes)[5,13]. The histone H2A.W variant contributes to heterochromatin condensation[26].
b, In the nodule differentiation zone, mCs are removed in all contexts, notably by DME. Histone repressive marks are replaced by activating histone marks (such as H3K9ac), particularly in symbiosis-related islands involved in nodule differentiation (SRIs-NDD)[13] while the eight *MtH2A.W* genes are down-regulated. This leads to local chromatin opening and enables the transcription machinery including transcription factors (TFs) to access the promoters and activate gene expression. Transcriptional activation of some TEs also occurs, either from TE promoters or by transcriptional read-through from upstream highly expressed genes.
c, In the nodule differentiation and nitrogen fixation zones, TE transcription leads to the production of siRNAs and increased RNA-directed DNA methylation (RdDM), targeting siRNA-related TEs first *in cis* and then *in trans,* genome-wide. Increased RdDM might also result from a better access of the RdDM machinery to the relaxed facultative heterochromatin (e.g. due to *MtH2A.W.1-8* downregulation). In the fixation zone, no DNA re-methylation takes place, and differentiation zone genes remain expressed or not depending on the presence or absence of required TFs.

References

1 Chow, H. T., Chakraborty, T. & Mosher, R. A. RNA-directed DNA Methylation and sexual reproduction: expanding beyond the seed. *Curr Opin Plant Biol* **54**, 11-17, doi:10.1016/j.pbi.2019.11.006 (2019).

2 Liu, R. *et al.* A DEMETER-like DNA demethylase governs tomato fruit ripening. *Proc Natl Acad Sci U S A* **112**, 10804-10809, doi:10.1073/pnas.1503362112 (2015).

3 Whittaker, C. & Dean, C. The FLC Locus: A Platform for Discoveries in Epigenetics and Adaptation. *Annu Rev Cell Dev Biol* **33**, 555-575, doi:10.1146/annurev-cellbio-100616-060546 (2017).

4 Satgé, C. *et al.* Reprogramming of DNA methylation is critical for nodule development in *Medicago truncatula*. *Nat Plants* **2**, 16166, doi:10.1038/nplants.2016.166 (2016).

5 Nagymihaly, M. *et al.* Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression. *Proc Natl Acad Sci U S A* **114**, 4543-4548, doi:10.1073/pnas.1704211114 (2017).

6 Deleris, A., Halter, T. & Navarro, L. DNA Methylation and Demethylation in Plant Immunity. *Annu Rev Phytopathol* **54**, 579-603, doi:10.1146/annurev-phyto-080615-100308 (2016).

7 Niyikiza, D. *et al.* Interactions of gene expression, alternative splicing, and DNA methylation in determining nodule identity. *Plant J*, **103**, 1744-1766, doi:10.1111/tpj.14861 (2020).

8 Zhang, H., Lang, Z. & Zhu, J. K. Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Biol* **19**, 489-506, doi:10.1038/s41580-018-0016-z (2018).

9 Mergaert, P., Kereszt, A. & Kondorosi, E. Gene Expression in Nitrogen-Fixing Symbiotic Nodule Cells in Medicago truncatula and Other Nodulating Plants. *Plant Cell* **32**, 42-68, doi:10.1105/tpc.19.00494 (2020).

10 Maunoury, N. *et al.* Differentiation of Symbiotic Cells and Endosymbionts in Medicago truncatula Nodulation Are Coupled to Two Transcriptome-Switches. *PLoS One* **5**, e9519 (2010).

11 Roux, B. *et al.* An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J* **77**, 817-837, doi:10.1111/tpj.12442 (2014).

12 Limpens, E. *et al.* Cell- and Tissue-Specific Transcriptome Analyses of *Medicago truncatula* Root Nodules. *PLoS One* **8**, e64377, doi:10.1371/journal.pone.0064377 (2013).

13 Pecrix, Y. *et al.* Whole-genome landscape of Medicago truncatula symbiotic genes. *Nat Plants* **4**, 1017-1025, doi:10.1038/s41477-018-0286-7 (2018).

14 Niederhuth, C. E. *et al.* Widespread natural variation of DNA methylation within angiosperms. *Genome Biol* **17**, 194, doi:10.1186/s13059-016-1059-0 (2016).

15 Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol* **20**, 275, doi:10.1186/s13059-019-1905-y (2019).

16 El Baidouri, M. *et al.* A new approach for annotation of transposable elements using small RNA mapping. *Nucleic Acids Res* **43**, e84, doi:10.1093/nar/gkv257 (2015).

17 Wang, L. *et al.* CRISPR/Cas9 knockout of leghemoglobin genes in Lotus japonicus uncovers their synergistic roles in symbiotic nitrogen fixation. *New Phytol* **224**, 818-832, doi:10.1111/nph.16077 (2019).

18 Bourcy, M. *et al.* Medicago truncatula DNF2 is a PI-PLC-XD-containing protein required for bacteroid persistence and prevention of nodule early senescence and defense-like reactions. *New Phytol* **197**, 1250-1261, doi:10.1111/nph.12091 (2013).

19 Berrabah, F. *et al.* A nonRD receptor-like kinase prevents nodule early senescence and defense-like reactions during symbiosis. *New Phytol* **203**, 1305-1314, doi:10.1111/nph.12881 (2014).

20 Domonkos, A. *et al.* NAD1 Controls Defense-Like Responses in Medicago truncatula Symbiotic Nitrogen Fixing Nodules Following Rhizobial Colonization in a BacA-Independent Manner. *Genes (Basel)* **8**, 387, doi:10.3390/genes8120387 (2017).

21 Sinharoy, S. *et al.* The C2H2 transcription factor regulator of symbiosome differentiation represses transcription of the secretory pathway gene VAMP721a and promotes symbiosome development in Medicago truncatula. *Plant Cell* **25**, 3584-3601, doi:10.1105/tpc.113.114017 (2013).

22 Henderson, I. R. *et al.* The de novo cytosine methyltransferase DRM2 requires intact UBA domains and a catalytically mutated paralog DRM3 during RNA-directed DNA methylation in Arabidopsis thaliana. *PLoS Genet* **6**, e1001182, doi:10.1371/journal.pgen.1001182 (2010).

23 Lei, M. *et al.* Regulatory link between DNA methylation and active demethylation in Arabidopsis. *Proc Natl Acad Sci U S A* **112**, 3553-3557, doi:10.1073/pnas.1502279112 (2015).

24 He, L. *et al.* Pathway conversion enables a double-lock mechanism to maintain DNA methylation and genome stability. *Proc Natl Acad Sci U S A* **118**, e2107320118, doi:10.1073/pnas.2107320118 (2021).

25 Kawakatsu, T. *et al.* Unique Cell-type-specific patterns of DNA methylation in the root meristem. *Nature Plants* **2**, 16058 (2016).

26 Bourguet, P. *et al.* The histone variant H2A.W and linker histone H1 co-regulate heterochromatin accessibility and DNA methylation. *Nat Commun* **12**, 2683, doi:10.1038/s41467-021-22993-5 (2021).

27 Walker, J. *et al.* Sexual-lineage-specific DNA methylation regulates meiosis in Arabidopsis. *Nat Genet* **50**, 130-137, doi:10.1038/s41588-017-0008-5 (2018).

28 Bouyer, D. *et al.* DNA methylation dynamics during early plant life. *Genome Biol* **18**, 179, doi:10.1186/s13059-017-1313-0 (2017).

29 Piya, S. *et al.* Hypermethylation of miRNA Genes During Nodule Development. *Front Mol Biosci* **8**, 616623, doi:10.3389/fmolb.2021.616623 (2021).

30 Li, Q. *et al.* RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. *Proc Natl Acad Sci U S A* **112**, 14728-14733, doi:10.1073/pnas.1514680112 (2015).

31 Barker, D. G. *et al. Growing Medicago truncatula: choice of substrates and growth conditions*. (ISBN 0-9754303-1-9. http://www.noble.org/MedicagoHandbook/, 2007).

32 Ardourel, M. *et al.* Rhizobium meliloti lipooligosaccharide nodulation factors: different structural requirements for bacterial entry into target root hair cells and induction of plant symbiotic developmental responses. *Plant Cell* **6**, 1357-1374 (1994).

33 Boisson-Dernier, A. *et al.* Agrobacterium rhizogenes-transformed roots of Medicago truncatula for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. *Mol Plant Microbe Interact* **14**, 695-700 (2001).

34 Mayjonade, B. *et al.* Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *Biotechniques* **61**, 203-205, doi:10.2144/000114460 (2016).

35 Roux, B., Rodde, N., Moreau, S., Jardinaud, M. F. & Gamas, P. Laser Capture Micro-Dissection Coupled to RNA Sequencing: A Powerful Approach Applied to the Model Legume Medicago truncatula in Interaction with Sinorhizobium meliloti. *Methods Mol Biol* **1830**, 191-224, doi:10.1007/978-1-4939-8657-6_12 (2018).

36 Schultz, M. D. *et al.* Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212-216, doi:10.1038/nature14465 (2015).

37 Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* **13**, R87, doi:10.1186/gb-2012-13-10-r87 (2012).

38 Zhou, M., Palanca, A. M. S. & Law, J. A. Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family. *Nat Genet* **50**, 865-873, doi:10.1038/s41588-018-0115-y (2018).

39 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-1645, doi:10.1101/gr.092759.109 (2009).

40 Huang, X. *et al.* ViewBS: a powerful toolkit for visualization of high-throughput bisulfite sequencing data. *Bioinformatics* **34**, 708-709, doi:10.1093/bioinformatics/btx633 (2018).

41 Chen, K. *et al.* DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res* **23**, 341-351, doi:10.1101/gr.142067.112 (2013).

42 Zhang, H. *et al.* Protocol: a beginner's guide to the analysis of RNA-directed DNA methylation in plants. *Plant Methods* **10**, 18, doi:10.1186/1746-4811-10-18 (2014).

43 Carrere, S., Verdier, J. & Gamas, P. MtExpress, a Comprehensive and Curated RNAseq-based Gene Expression Atlas for the Model Legume Medicago truncatula. *Plant Cell Physiol* **62**, 1494-1500, doi:10.1093/pcp/pcab110 (2021)

44 Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25, doi:10.1186/gb-2010-11-3-r25 (2010).

45 Robinson, M. D. & Smyth, G. K. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9**, 321-332, doi:10.1093/biostatistics/kxm030 (2008).
46 Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* **29**, 1165–1188 (2001).
47 Alexa, A., Rahnenfuhrer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600-1607, doi:10.1093/bioinformatics/btl140 (2006).
48 Johnson, N. R., Yeoh, J. M., Coruh, C. & Axtell, M. J. Improved Placement of Multi-mapping Small RNAs. *G3 (Bethesda)* **6**, 2103-2111, doi:10.1534/g3.116.030452 (2016).
49 Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc Natl Acad Sci U S A* **112**, 3570-3575, doi:10.1073/pnas.1420294112 (2015).
50 Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6**, e16765
51 Lemoine, F. et al. NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res* **47**, W260-W265, doi:10.1093/nar/gkz303 (2019).
52 Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881-10890 (1988).
53 Carrère, S., Verdenaud, M., Gough, C., Gouzy, J. & Gamas, P. LeGOO: An Expertized Knowledge Database for the Model Legume Medicago truncatula. *Plant Cell Physiol.* **61**, 203–211, doi:https://doi.org/10.1093/pcp/pcz177 (2019)

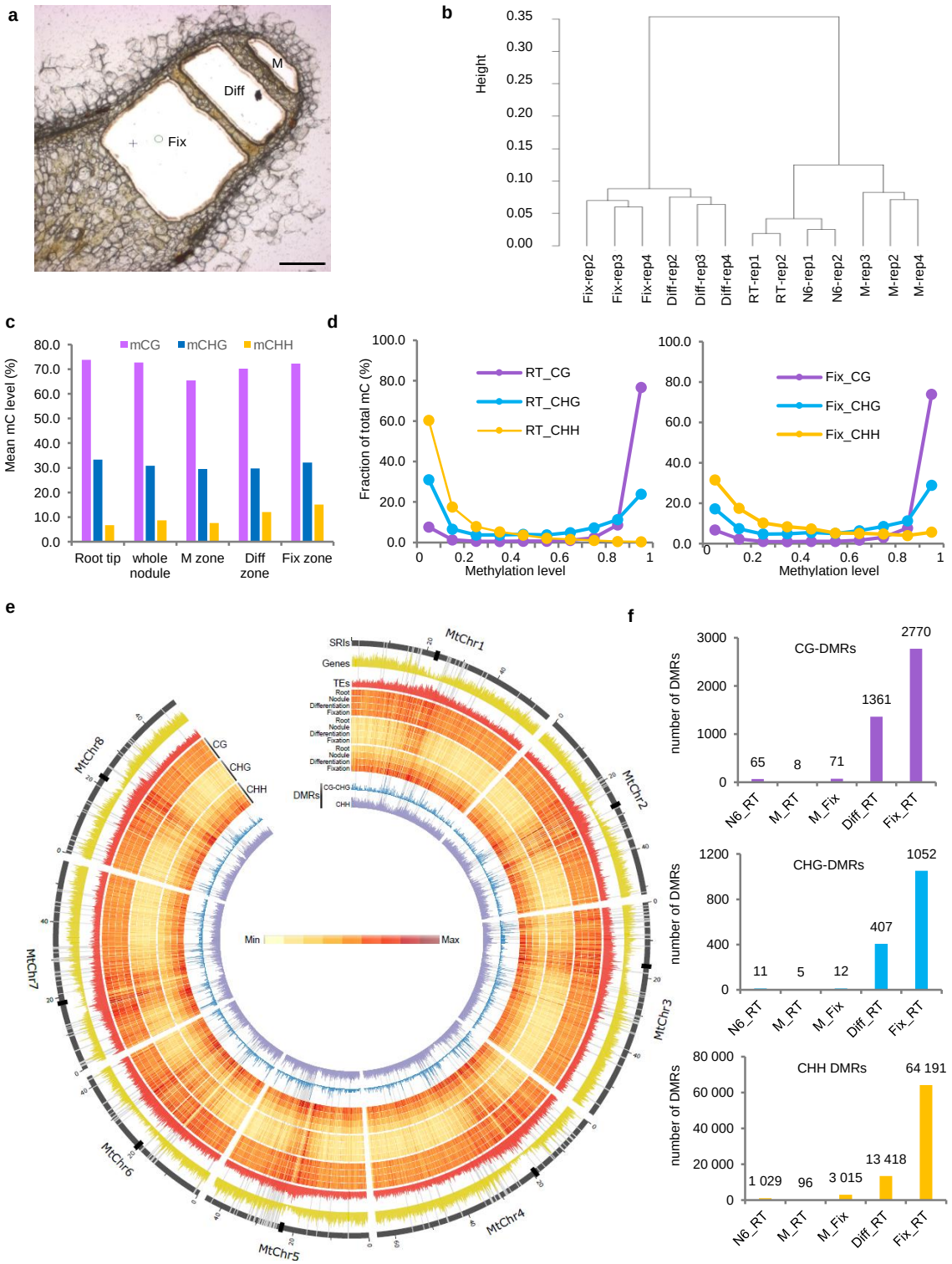

a
M
Diff
Fix
b
Height
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Fix-rep2
Fix-rep3
Fix-rep4
Diff-rep2
Diff-rep3
Diff-rep4
RT-rep1
RT-rep2
N6-rep1
N6-rep2
M-rep3
M-rep2
M-rep4
c
mCG
mCHG
mCHH
Mean mC level (%)
Root tip
whole nodule
M zone
Diff zone
Fix zone
d
Fraction of total mC (%)
RT_CG
RT_CHG
RT_CHH
Methylation level
Fix_CG
Fix_CHG
Fix_CHH
e
SRIs
Genes
TEs
DMRs
CG
CHG
CHH
Min
Max
MtChr1
MtChr2
MtChr3
MtChr4
MtChr5
MtChr6
MtChr7
MtChr8
f
CG-DMRs
number of DMRs
65
8
71
1361
2770
N6_RT
M_RT
M_Fix
Diff_RT
Fix_RT
CHG-DMRs
11
5
12
407
1052
CHH DMRs
1 029
96
3 015
13 418
64 191

Figure 1

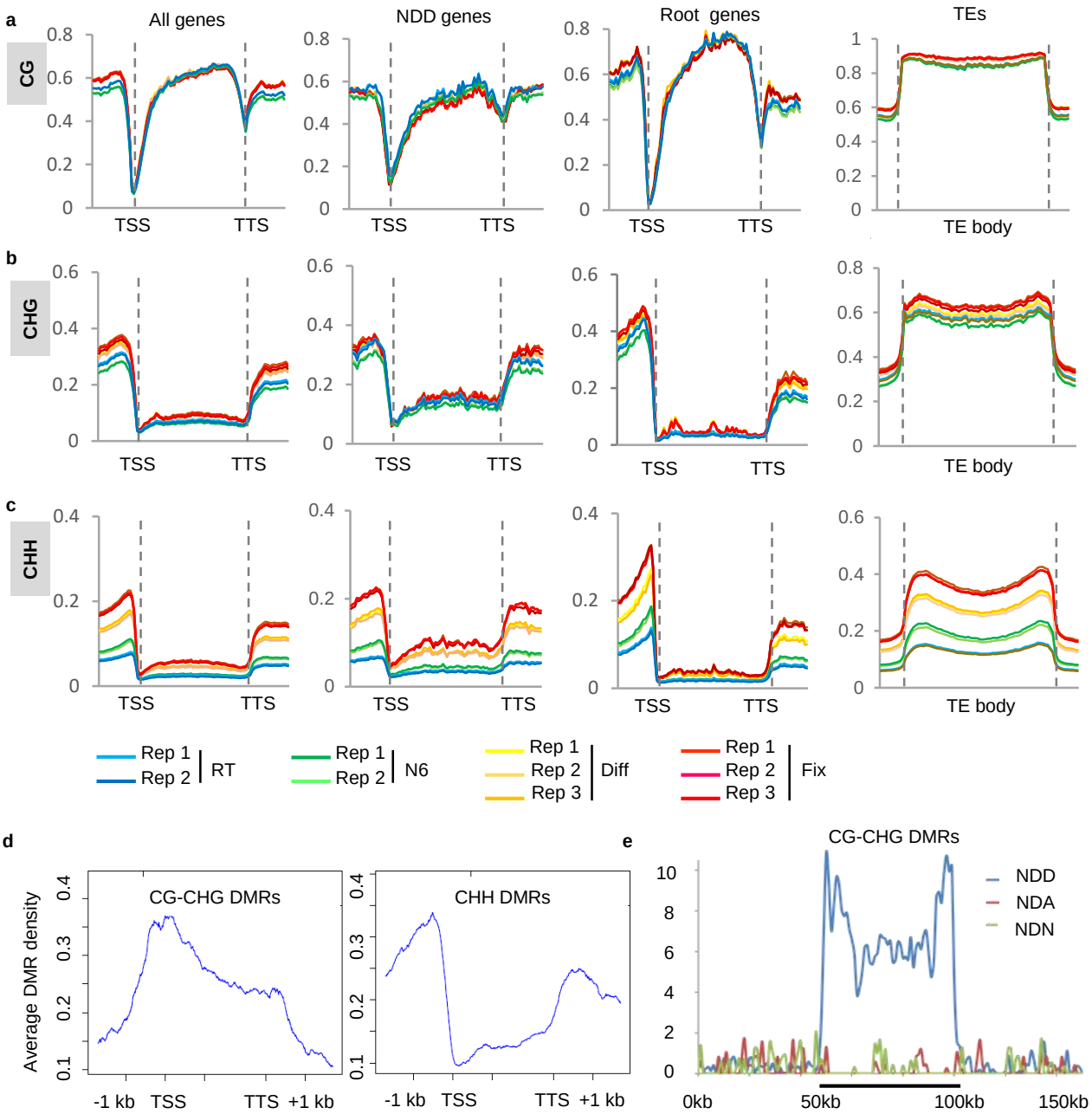

a
All genes
NDD genes
Root genes
TEs
CG
CHG
CHH
TSS
TTS
TE body
b
c
Rep 1
Rep 2
RT
N6
Rep 3
Diff
Fix
d
CG-CHG DMRs
CHH DMRs
Average DMR density
-1 kb
+1 kb
e
NDD
NDA
NDN
0kb
50kb
100kb
150kb

Figure 2

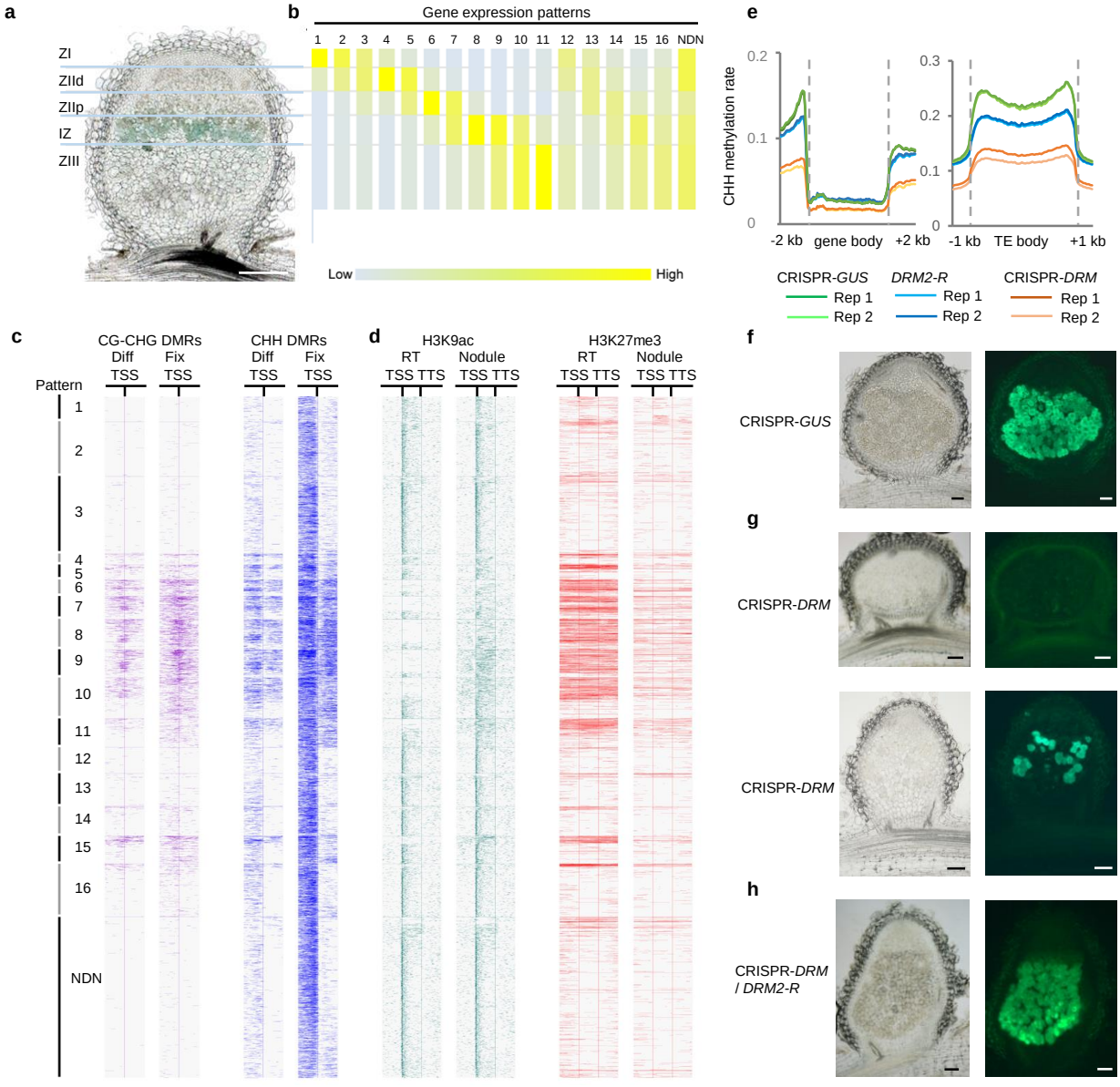

a
ZI
ZIId
ZIIp
IZ
ZIII
b
Gene expression patterns
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 NDN
Low
High
e
CHH methylation rate
0.2
0.1
0
0.3
0.2
0.1
0
-2 kb gene body +2 kb
-1 kb TE body +1 kb
CRISPR-GUS
Rep 1
Rep 2
DRM2-R
Rep 1
Rep 2
CRISPR-DRM
Rep 1
Rep 2
c
CG-CHG DMRs
Diff TSS
Fix TSS
CHH DMRs
Diff TSS
Fix TSS
Pattern
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
NDN
d
H3K9ac
RT TSS TTS
Nodule TSS TTS
H3K27me3
RT TSS TTS
Nodule TSS TTS
f
CRISPR-GUS
g
CRISPR-DRM
CRISPR-DRM
h
CRISPR-DRM / DRM2-R

Figure 3

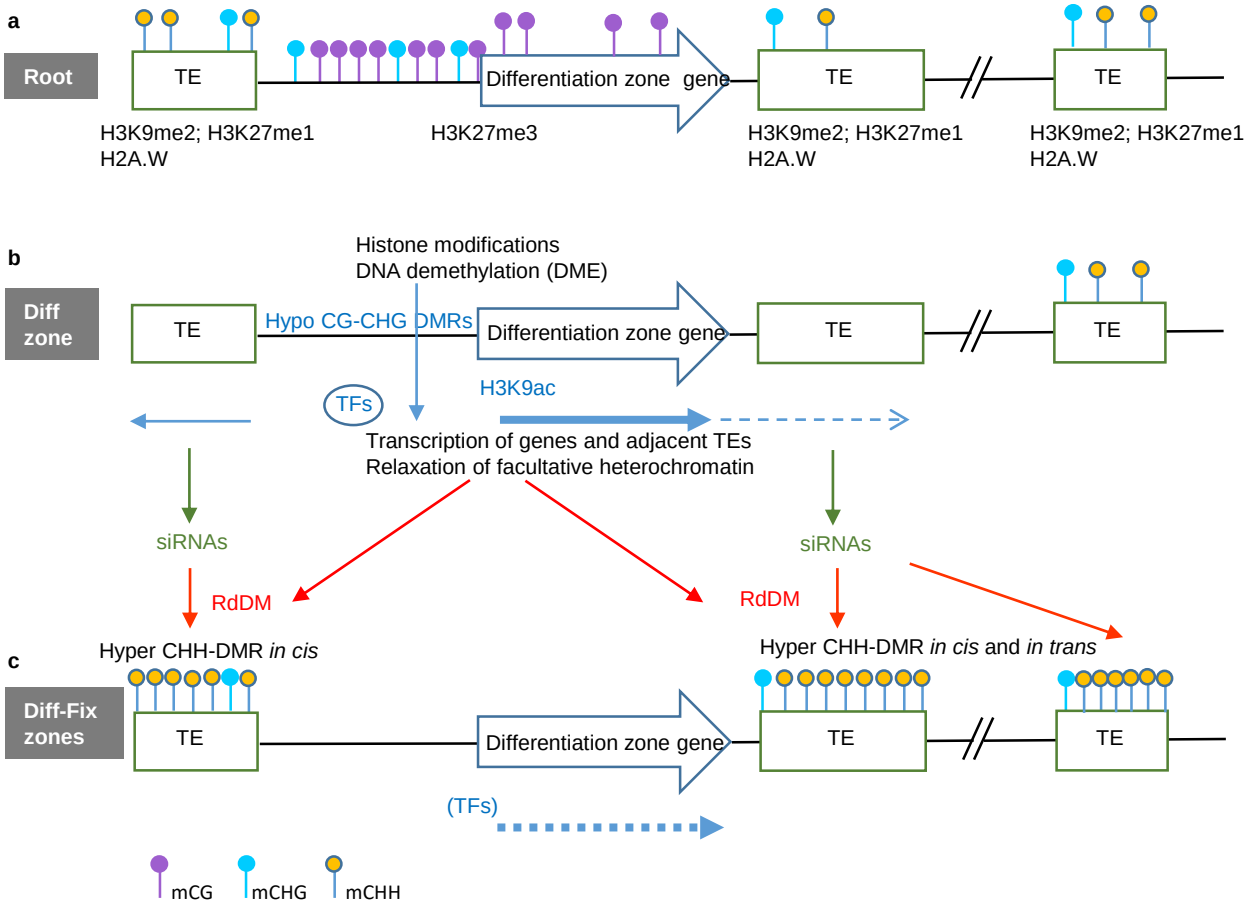
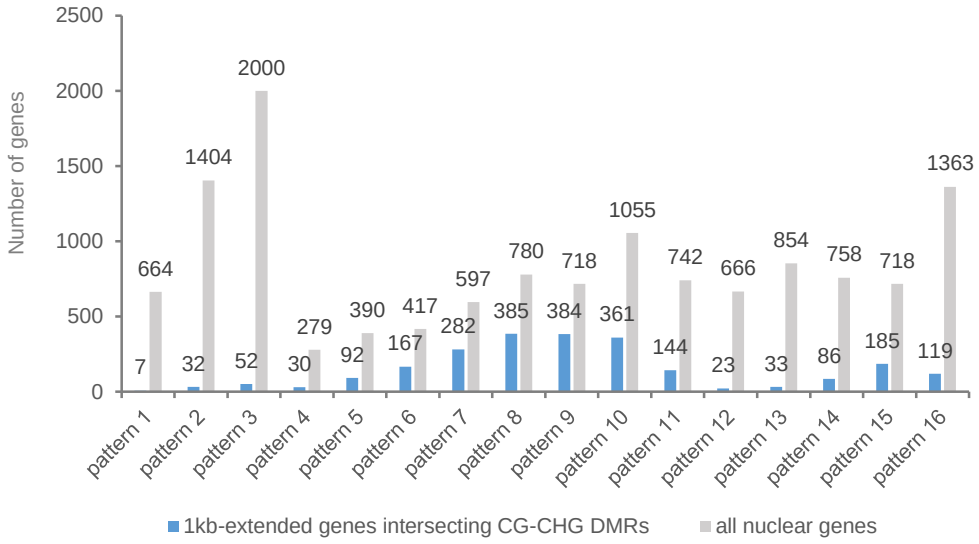
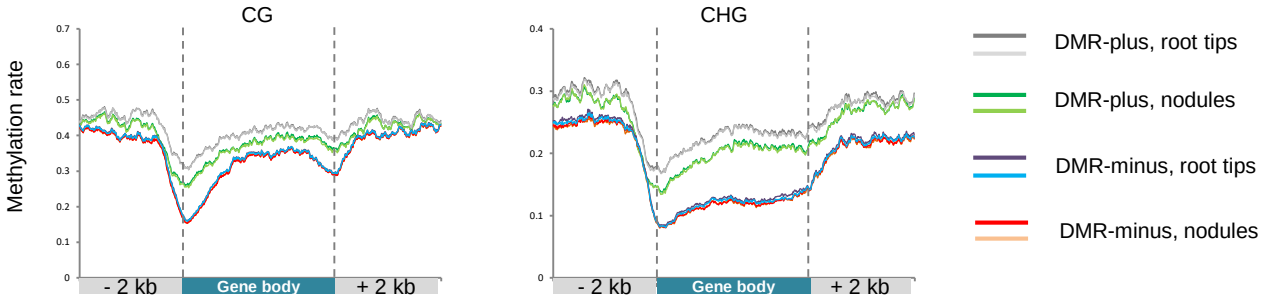

a
Root
TE
Differentiation zone gene
TE
TE
H3K9me2; H3K27me1
H2A.W
H3K27me3
H3K9me2; H3K27me1
H2A.W
H3K9me2; H3K27me1
H2A.W
b
Diff zone
Histone modifications
DNA demethylation (DME)
TE
Hypo CG-CHG DMRs
Differentiation zone gene
TE
TE
TFs
H3K9ac
Transcription of genes and adjacent TEs
Relaxation of facultative heterochromatin
siRNAs
RdDM
siRNAs
RdDM
c
Diff-Fix zones
Hyper CHH-DMR in cis
Hyper CHH-DMR in cis and in trans
TE
Differentiation zone gene
TE
TE
(TFs)
mCG
mCHG
mCHH

Figure 4

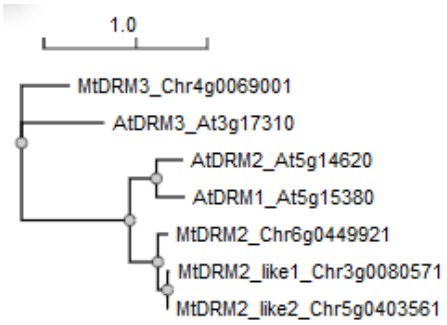
Extended Figures 1 to 9


Medicago truncatula A17 v5
File
View
Help
0
5,000,000
10,000,000
15,000,000
20,000,000
25,000,000
30,000,000
35,000,000
MtrunA17Chr1
MtrunA17Chr1:21950386..2196250
Go
21,952,500
21,955,000
21,957,500
1.e. Complete annotation (EuGene/Repeats/Rescued), release 1.8
MtrunA17_Chr1R0114350
MtrunA17_Chr1R0114360
MtrunA17_Chr1R0114370
MtNCR060, MtrunA17_Chr1g0172681
Nodule-specific Cysteine-Rich peptide
MtrunA17_Chr1R0114380
MtrunA17_Chr1g1001759
Root polyA
+2σ
+1σ
mean
2 000
1 000
0
Nodule polyA
-1 000
24nt-siRNAs
methylation level root tips (RT) rep2
1
0
-1
Methylation level Fix rep2
[CG-CHG DMR] fix-root tips
[CHH DMR] fix-root tips

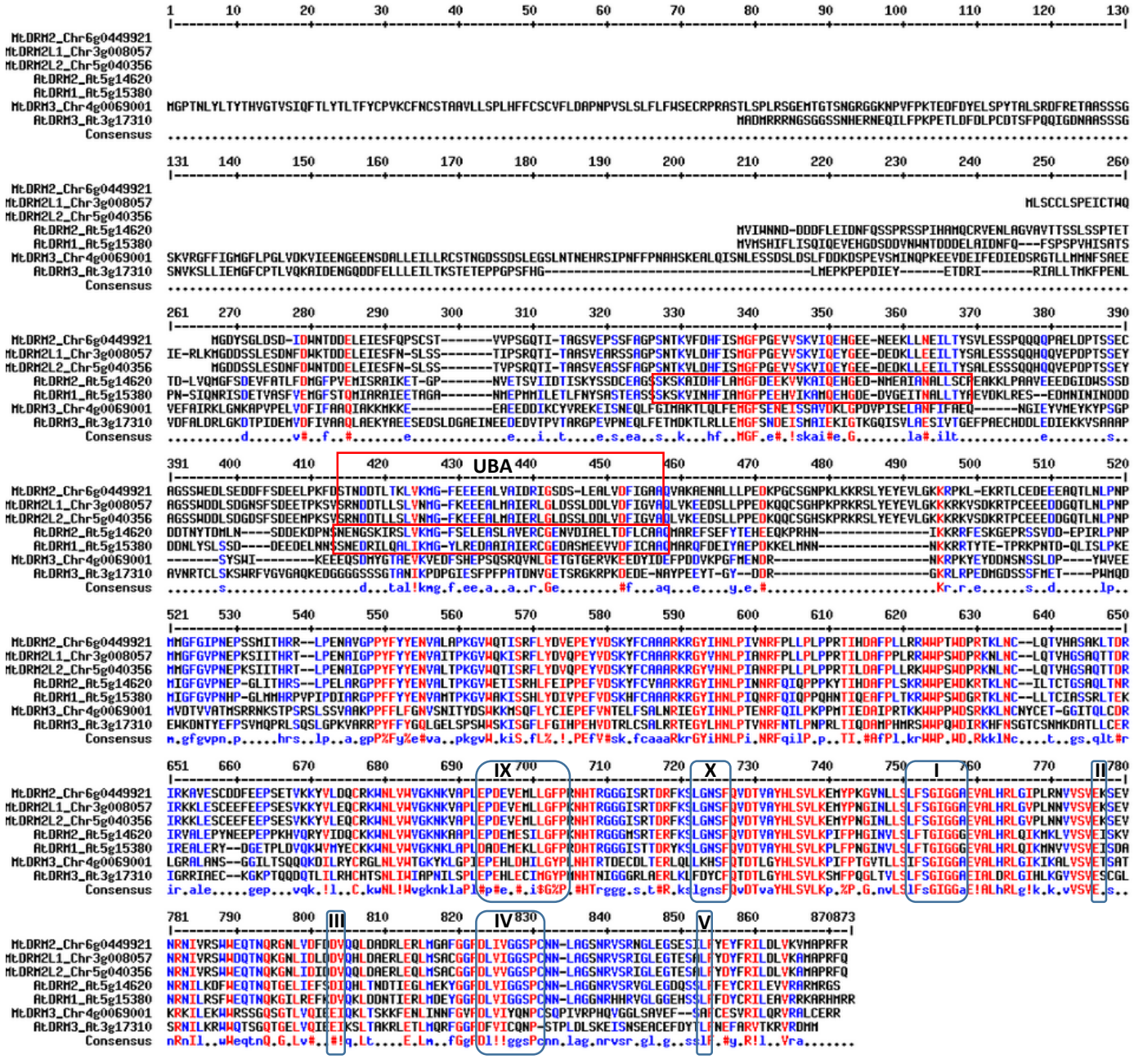

Number of genes
2500
2000
1500
1000
500
0
664
1404
2000
279
390
417
597
780
718
1055
742
666
854
758
718
1363
7
32
52
30
92
167
282
385
384
361
144
23
33
86
185
119
pattern 1
pattern 2
pattern 3
pattern 4
pattern 5
pattern 6
pattern 7
pattern 8
pattern 9
pattern 10
pattern 11
pattern 12
pattern 13
pattern 14
pattern 15
pattern 16
1kb-extended genes intersecting CG-CHG DMRs
all nuclear genes


CG
CHG
Methylation rate
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.4
0.3
0.2
0.1
0
- 2 kb
Gene body
+ 2 kb
- 2 kb
Gene body
+ 2 kb
DMR-plus, root tips
DMR-plus, nodules
DMR-minus, root tips
DMR-minus, nodules

a


1.0
MtDRM3_Chr4g0069001
AtDRM3_At3g17310
AtDRM2_At5g14620
AtDRM1_At5g15380
MtDRM2_Chr6g0449921
MtDRM2_like1_Chr3g0080571
MtDRM2_like2_Chr5g0403561

b


UBA
IX
X
I
II
III
IV
V

a

DRM2 EXON8-INTRON-EXON9

CCGGGATGCTCCGGGAATCCAAAGCTGAAAAAGCGCAGTCTCTATGAATATGAGGTCCTTGGGAAGAAAAGACCGAAGTTAGAGAAGAGAACACTATGTGAAGACGAGGAGGAGGCGCAAACGCTTAAT

Guide a PAM

CTGCCTAATCCTATGATGGGGTTTGGTATTCCTAATGAGCCCAGTTCTATGATTACACACAGGAGACTGCCTGAAAATGCCGTTGGCCCGCCCTATTTCTACTACGAGAATGTGGCGTTGGCACCGAAA

PAM Guide b

GGTGTCTGGCAAACAATCTCAAGGTTCTTATATGATGTGGAACCCGAGTACGTCGACTCAAAATACTTTTGTGCTGCTGCGCGTAAACGTGGATATATTCACAATCTCCCAATTGTGAATAGATTCCCT

CTTTTACCTCTTCCACCGCGCACCATCCATGATGCATTCCCCTTTACTAAGGAGGTGGTGGCCGACATGGGACCCCAGGACAAAGCTTAATTGTTTGCAGACCGTACATGCTAGTGCGAAACTTACCGAC

AGAATCCGGAAAGCGGTAGAAAGTTGTGATGATTTTGAGGAACCTTCTGAAACGGTAAAAAAATATGTGTTGGATCAGTGTCGGAAATGGAACCTGGTATGGGTGGGTAAGAATAAAGTTGCCCCCCTA

GAGCCCGATGAAGTTGAGATGCTGTTGGGATTCCCTAGGAATCATACAAGGGGTGGTGGAATCAGTAGGACGGACAGATTCAAGTCACTTGGAAATTCTTTCCAG

Guide c PAM

ATTGGTGGTGCAGAGGTAGCTCTGCATAGACTTGGCATCCCTCTAAGGAATGTAGTGTCAGTCGAGAAATCTGAAGTGAACAGGAACATTGTTAGGAGTTGGTGGGAGCAAACCAACCAAAGGGGTAAT

TTAGTTGATTTTGATGATGTGCAACAGCTAGATGCCGACCGTTTGGAGCGGCTGATGGGCGCATTTGGTGGCTTTGATCTAATTGTTGGTGGAAGCCCTTGCAATAATCTGGCTGGAAGCAATAGGGTT

AGTCGGAATGGACTTGAGGGCTCGGAATCTATCCTATTTTATGAATACTTTAGGATTTTAGACTTAGTGAAGGTTATGGCGCCTAGATTTCGATGA

DRM2L2 EXON8-INTRON-EXON9

CAALCAATGCAGTGGGCATTCAAAACCAAGAAAGCGTAGTCTTTATGAATATGAAGTGCTAGGAAAGAAAAAACGAAAGGTGTCCGACAAGAGAACTCCATGTGAAGAGGAGGATGATGGCCAAACACT

TAATCTGCCTAACCCAATGATGGGGTTTGGGGTTCCCAACGAGCCAAAGTCTATAATTACACACCGGACACTCCCTGAGAATGCTATTGGACCTCCCTACTTCTACTATGAAAATGTGGCACTAACACC

GAAAGGTGTTTGGCAAACAATATCTAGGTTTTTGTATGACGTGCAGCCGGAATATGTAGACTCAAAATATTTTTGTGCTGCGGCTAGGAAAAGGGGATATGTTCACAATCTCCCTATTGCCAATAGATT

Guide e PAM

CCCGCTTTTACCTCTTCCACCACGGACAATACTTGATGCATTCCCCCTATTGAGGAAATGGTGGCCATCATGGGATCCTAGAAAAAACCTGAATTGTTTGCAAACGGTACATGGCAGTGCACAAACCAC

PAM Guide f

CGACAGAATCCGGAAAAAGCTGGAAAGTTGTGAAGAATTTGAAGAACCATCGGAATCTGTCAAGAAGTATGTTTTGGAACAATGTCGGAAATGGAATTTGGTATGGGTGGGCAAGAATAAAGTTGCCCC

CCTAGAGCCCGACGAAGTGGAAATGCTGCTGGGCTTCCCTAGGAATCATACAAGGGGGGGTGGAATCAGTCGGACCGACAGATTCAAATCACTTGGAAATTCTTTCCAG

CTCTATATATATATATATATATATATATATATATTTCTGATACGTCTGTTGGTAGAAATAAATAAGGCATAATAGGCTTATTTGGTGGTAATTAGAATTGGGAGTGCATTAGATCTCTCAAATATCAAGTGAAAGGGAATCTGTGTGTTCTTTCTGTTCATTGGCAACAATACGTTCTTTCTATTCATAATCTAAAACAGTTTTTATCAGACCAATAAACTTTGAAT

AAAACGTTAAAATAGATTTTGGTGCAACAAGTCTCTACATAGAATTGCTTTCATTTACAGCGCATCAGTGGCATTTTTGTGTTTTACATTGTATTCAGCATAATTTTGTCCAAGTGCCAATTTTTTTTCTAATACTATTTATTTCGTTGTTTTCCCTGTGAAACTTTCCAATAGATTGTGATAGCAAGTTACATTGTTTTTCTCTATAAAAGAAATTTGTGGCTTATGCA

TTTAAGTTTTCCGTGTACAATTCGTTTTCCTTATGTGTTCAGTTTATTTCAGGTTGACAGTGGCGTATCACCTATCTGTTCTTAAAGAAATGTATCCTAATGGGATCAATCTTTTATCCCTCTTTTCTGGAATTGGTGGTGCAGAAGTAG

CTCTCCATCGACTCGGTGTCCCTCTAAATAATGTTGTGTCGGTGGAAAAATCCGAAGTAAATAGGAACATTGTTAGAAGTTGGTGGGAACAGACAAATCAAAAAGGTAATTTGATTGATATCGATGATG

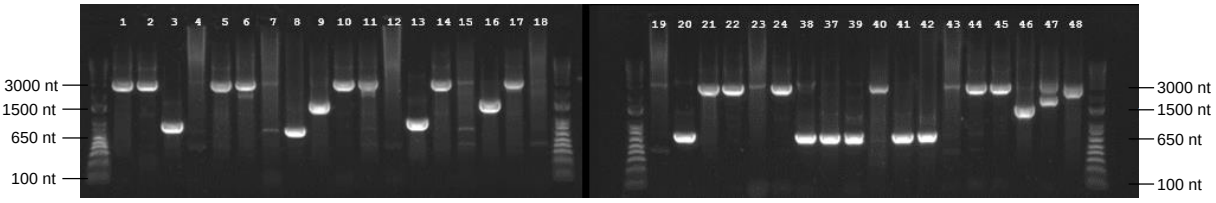
Guide d PAM

TGCAACAGCTAGATGCTGAGCGTTTGGAGCAGCTCATGAGTGCAGTTGACACAGT

b

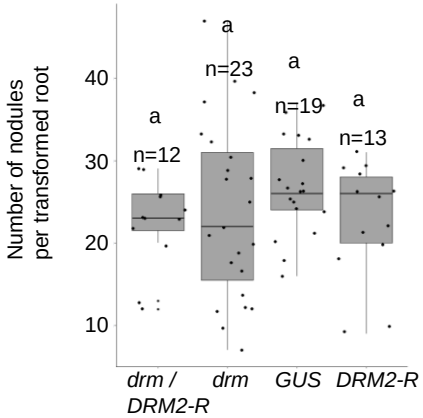
	DRM2 guides (220 plants)			*DRM2L2* guides (158 plants)		
	a	b	c	d	e	f
Mono-guide editions	5	9	69	6	4	4
Multi-guide editions	15	80	50	57	22	32

c



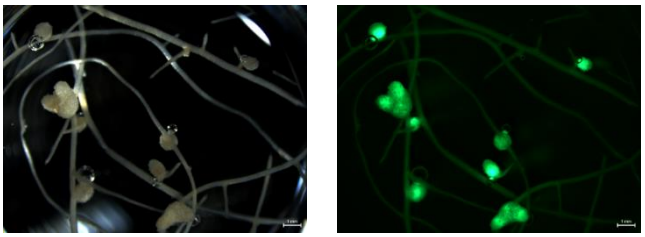
d

a



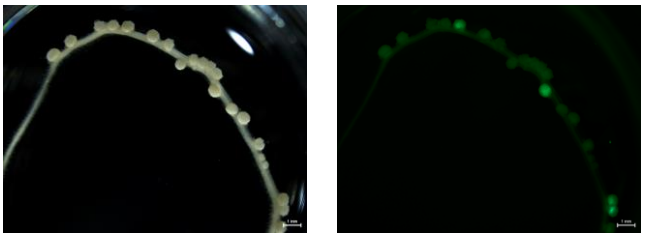
b

CRISPR-*GUS*



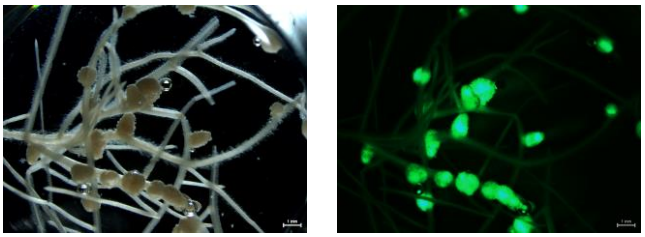
GFP score: 4; ethylene area / nodule: 8.4 pA*s

CRISPR-*DRM2/DRM2L2*



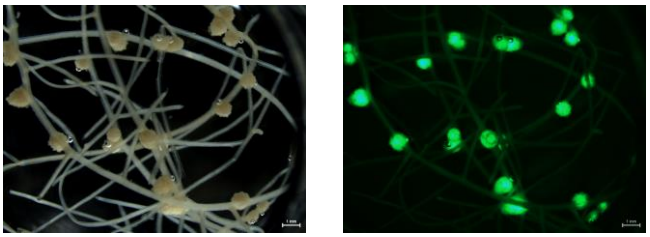
GFP score: 1; ethylene area / nodule: 0.8 pA*s

pAtUBI:DRM2-R

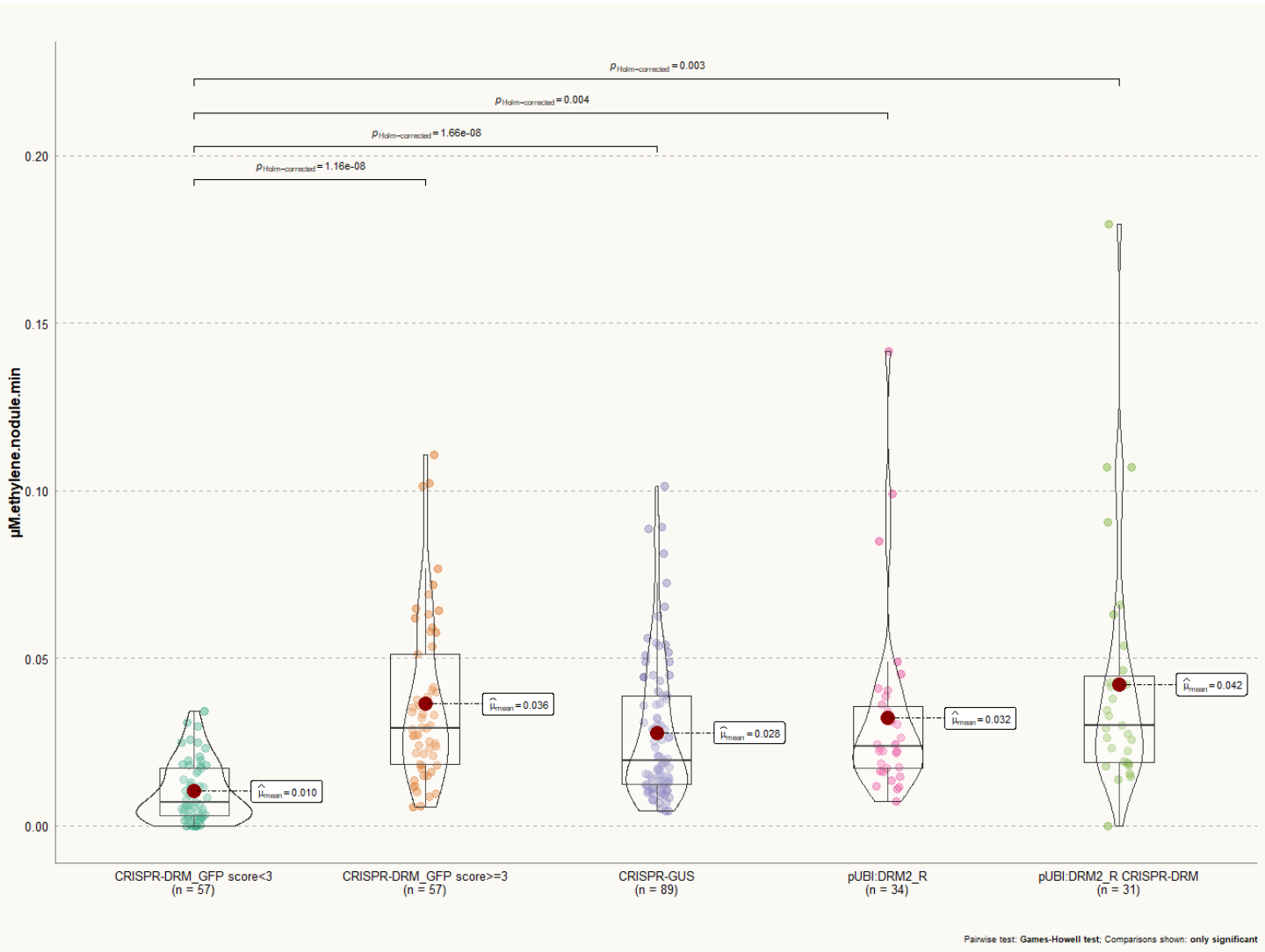


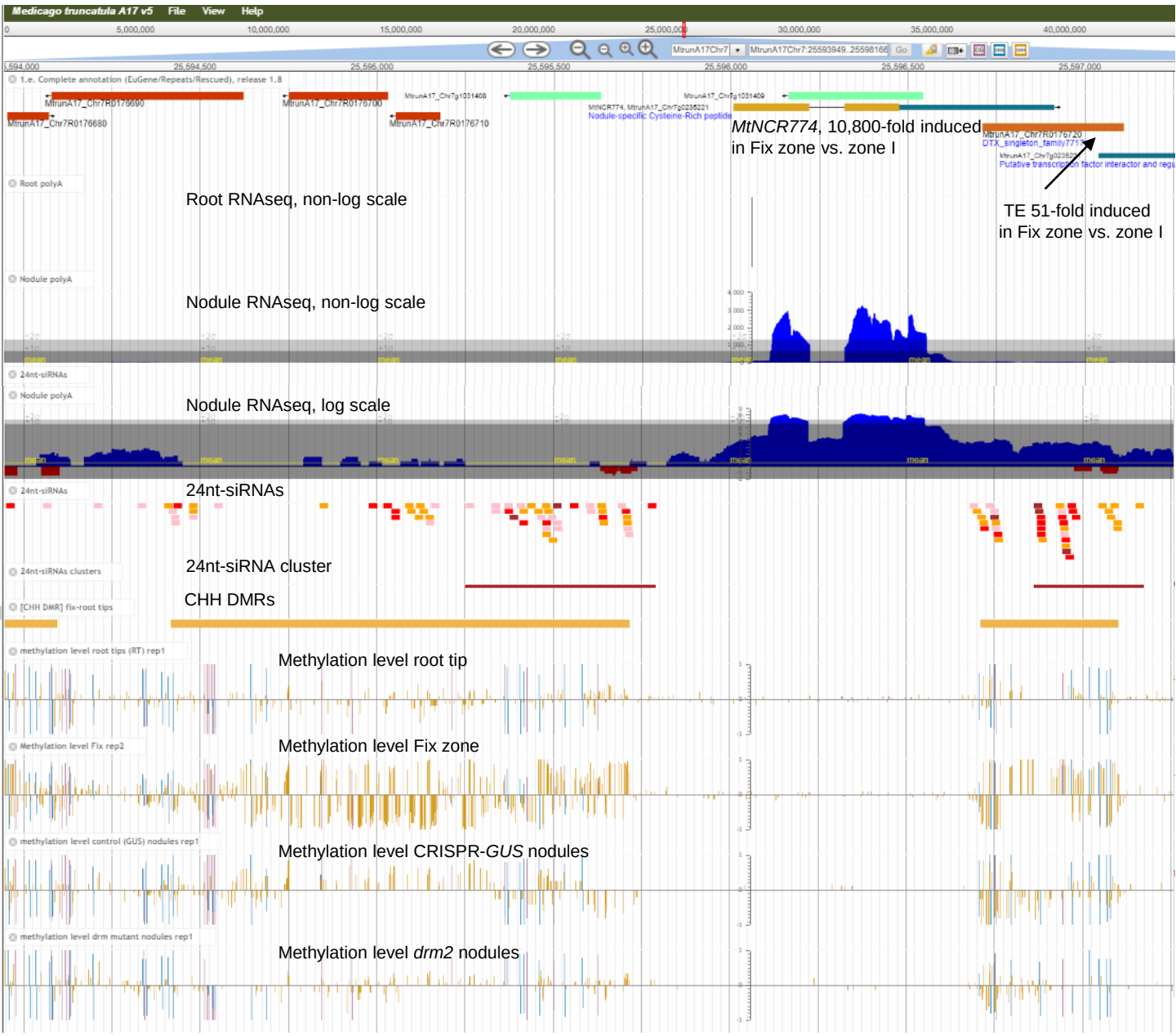
GFP score: 4; ethylene area / nodule: 3.5 pA*s

CRISPR-*DRM2/DRM2L2 pAtUBI:DRM2-R*

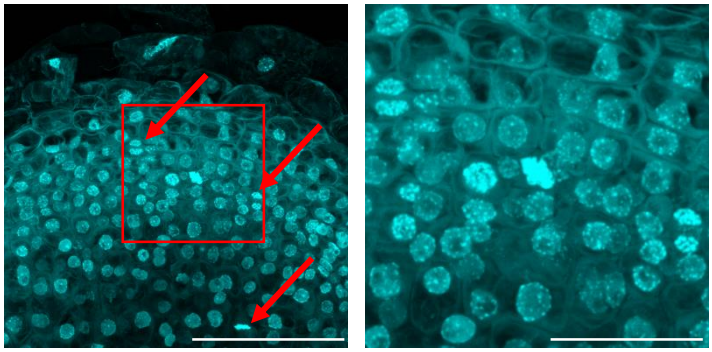


GFP score: 4; ethylene area / nodule: 6.6 pA*s

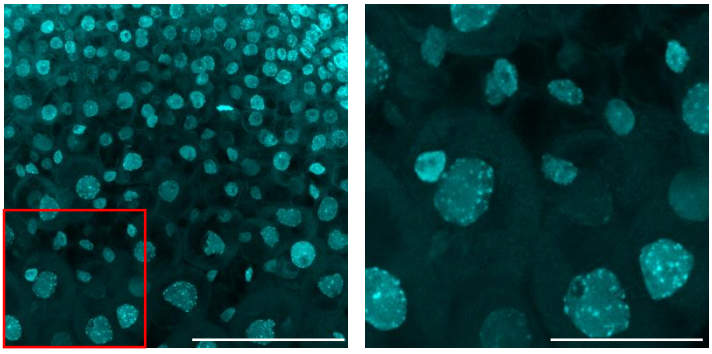

$p_{Holm-corrected}$ = 0.003
$p_{Holm-corrected}$ = 0.004
$p_{Holm-corrected}$ = 1.66e-08
$p_{Holm-corrected}$ = 1.16e-08
µM.ethylene.nodule.min
0.20
0.15
0.10
0.05
0.00
$\hat{\mu}_{mean}$ = 0.010
$\hat{\mu}_{mean}$ = 0.036
$\hat{\mu}_{mean}$ = 0.028
$\hat{\mu}_{mean}$ = 0.032
$\hat{\mu}_{mean}$ = 0.042
CRISPR-DRM_GFP score<3
(n = 57)
CRISPR-DRM_GFP score>=3
(n = 57)
CRISPR-GUS
(n = 89)
pUBI:DRM2_R
(n = 34)
pUBI:DRM2_R CRISPR-DRM
(n = 31)
Pairwise test: **Games-Howell test**; Comparisons shown: **only significant**


Medicago truncatula A17 v5
File
View
Help
1.e. Complete annotation (EuGene/Repeats/Rescued), release 1.8
MtrunA17_Chr7R0176690
MtrunA17_Chr7R0176700
MtrunA17_Chr7R0176680
MtrunA17_Chr7R0176710
Nodule-specific Cysteine-Rich peptide
MtNCR774, 10,800-fold induced in Fix zone vs. zone I
TE 51-fold induced in Fix zone vs. zone I
Root polyA
Root RNAseq, non-log scale
Nodule polyA
Nodule RNAseq, non-log scale
24nt-siRNAs
Nodule RNAseq, log scale
24nt-siRNAs
24nt-siRNA cluster
24nt-siRNAs clusters
CHH DMRs
[CHH DMR] fix-root tips
methylation level root tips (RT) rep1
Methylation level root tip
Methylation level Fix rep2
Methylation level Fix zone
methylation level control (GUS) nodules rep1
Methylation level CRISPR-GUS nodules
methylation level drm mutant nodules rep1
Methylation level drm2 nodules

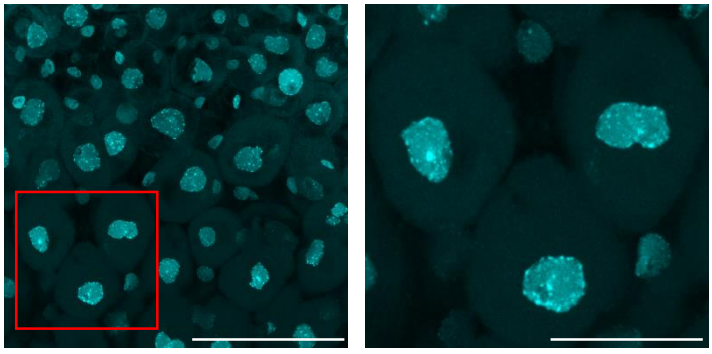
Zone I
Zone IId



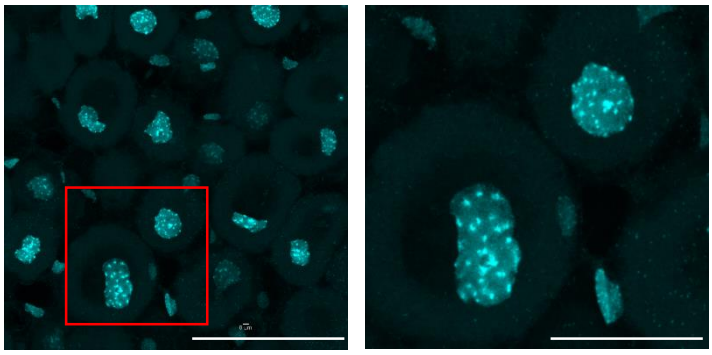
Zone IId
Zone IIp



Interzone II-III



Zone III



Supplemental Tables 1, 8 and 9
Supplemental Note 1

Supplemental Table 1. Whole-genome bisulfite sequencing libraries and BsSeq data summary.
corresponding names in the *M.truncatula* genome browser are in parentheses; estimated genome size = 430Mb

sample	sequencing platform	read length	single-end (1) / paired-end (2)	raw reads	uniquely mapping reads	uniquely mapping reads %	non-clonal reads	non-clonal reads %	useful reads %	sequencing depth	non-conversion rate %
root tips (RT) rep1	GeT Plage	150	1	198 744 374	58 646 991	29.5%	46 501 259	79.3%	23.4%	16.2	1.26
root tips (RT) rep2	GeT Plage	150	1	188 125 142	94 395 254	50.2%	72 308 131	76.6%	38.4%	25.2	1.07
6 dpi nodules (N6) rep1	GeT Plage	150	1	298 305 105	68541505	23.0%	48 103 904	70.2%	16.1%	16.8	1.09
6 dpi nodules (N6) rep2	GeT Plage	150	1	298 440 135	74343704	24.9%	54 204 340	72.9%	18.2%	18.9	1.07
laser-dissected meristematic zone (Meristem rep2)	GeT Plage	150	1	399 966 743	69 966 258	17.5%	18 684 803	26.7%	4.7%	6.5	1.65
laser-dissected meristematic zone (Meristem rep3)	GeT Plage	150	1	290 385 535	79 699 717	27.4%	15 210 897	19.1%	5.2%	5.3	2.11
laser-dissected meristematic zone (Meristem rep4)	GeT Plage	150	1	352 885 728	60 100 310	17.0%	15 065 571	25.1%	4.3%	5.3	1.72
laser-dissected differentiation zone (Diff rep2)	GeT Plage	150	1	169 696 939	66 412 042	39.1%	29 992 961	45.2%	17.7%	10.5	1.63
laser-dissected differentiation zone (Diff rep3)	GeT Plage	150	1	174 374 489	66 159 837	37.9%	30 506 850	46.1%	17.5%	10.6	1.6
laser-dissected differentiation zone (Diff rep4)	GeT Plage	150	1	204 599 807	52 958 871	25.9%	29 552 672	55.8%	14.4%	10.3	1.56
laser-dissected fixation zone (Fix rep2)	GeT Plage	150	1	209 807 770	68 238 361	32.5%	47 547 459	69.7%	22.7%	16.6	2.18
laser-dissected fixation zone (Fix rep3)	GeT Plage	150	1	190 976 920	72 515 451	38.0%	45 607 620	62.9%	23.9%	15.9	1.88
laser-dissected fixation zone (Fix rep4)	GeT Plage	150	1	177 392 752	66 952 746	37.7%	42 646 733	63.7%	24.0%	14.9	2.26
CRISPR-GUS root tips (bRT rep1)	BGI	150	2	110 472 066	46 047 157	41.7%	33 572 922	72.9%	30.4%	23.4	1.19
CRISPR-GUS root tips (bRT rep2)	BGI	150	2	101 908 938	46 078 533	45.2%	33 548 766	72.8%	32.9%	23.4	1.26
CRISPR-GUS nodules (bGUS rep1)	BGI	150	2	119 566 679	28 799 615	24.1%	20 789 431	72.2%	17.4%	14.5	1.01
CRISPR-GUS nodules (bGUS rep2)	BGI	150	2	112 338 095	29 120 814	25.9%	21 146 217	72.6%	18.8%	14.8	1.15
CRISPR-DRM nodules (bDRM rep1)	BGI	150	2	85 503 038	32 533 045	38.0%	25 256 897	77.6%	29.5%	17.6	1.19
CRISPR-DRM nodules (bDRM rep2)	BGI	150	2	116 447 133	43 630 496	37.5%	33 644 685	77.1%	28.9%	23.5	1.28
CRISPR-DRM /pAtUbi:DRM2-R nodules (DRM2-R rep1)	BGI	150	2	128 638 179	40 325 617	31.3%	30 097 018	74.6%	23.4%	21.0	1.29
CRISPR-DRM /pAtUbi:DRM2-R nodules (DRM2-R rep2)	BGI	150	2	125 208 648	32 192 586	25.7%	24 050 933	74.7%	19.2%	16.8	1.11

Supplemental Table 8. Chop-PCR validation of CHH DMRs.

Whole organ analyses CHH DMR position	Methylation-sensitive restriction enzyme	mean delta CT nodule (n=3)	SE delta CT nodule	mean delta CT root tips (n=3)	SE delta CT nodule	Shapiro test*	Fisher's test*	one-sided t.test
Chr1_35839283-35840045	AluI	7.53	1.12	12.54	1.30	0.9730	0.8565	0.0324
	DdeI	9.76	3.28	14.93	3.16	0.4665	0.9270	0.1904
Chr2_32878129-32878605	AluI	9.00	1.58	12.67	3.27	0.0523	0.3803	0.2149
	DdeI	6.66	0.42	7.59	0.50	0.2860	0.3103	0.4908
Chr3_51183413-51184086	AluI	4.05	1.30	8.54	0.95	0.6230	0.6981	0.0455
	DdeI	8.13	1.56	10.99	0.79	0.4083	0.5227	0.1082
Chr4_30558968-30560144	AluI	2.03	0.96	5.38	2.13	0.6684	0.3397	0.1409
	DdeI	3.09	1.37	4.37	1.32	0.6019	0.9645	0.2960
Chr4_31795384-31796098	AluI	1.71	1.18	7.54	1.41	0.8531	0.8236	0.0256
	DdeI	4.14	1.93	9.64	1.50	0.8754	0.7566	0.0614
Chr4_31798378-31799002	AluI	5.94	1.24	14.82	0.86	0.6587	0.6447	0.0035
	DdeI	5.83	1.83	21.72	1.74	0.4678	0.9517	0.0028
Chr4_9710107-9710722	AluI	4.76	1.45	6.42	1.06	0.2362	0.7014	0.2339
	DdeI	4.94	1.81	6.96	0.92	0.2130	0.4092	0.2174

Laser dissected Fix zone vs root tip DNA	Methylation-sensitive restriction enzyme	delta CT Fix zone (n=2)**	delta CT root tips (n=2)**
Chr1_26645685-26647163	NlaIII-1	-1.04	3.36
	NlaIII-2	1.82	2.94
	DdeI-1	1.47	5.74
	DdeI-2	0.33	6.55

delta CT: compared to mock non-digested samples

n= biological replicates, using pools of 303, 434 and 357 nodules (15 dpi) from 5, 10 and 8 plants, respectively.

*Shapiro test: normality test; Fisher's test: variance equality test

** DNA remaining from LCM-BsSeq experiments; the results for each replicate are shown since no statistical test was possible (only two replicates being available)

Supplemental Table 9. Multi-guide (mg) cloning, sequencing and Chop-PCR primers

cloning primers	5' to 3' sequence
mgDRM2a-F	TAGGTCTCCTCTATGAATATGGTTTTAGAGCTAGAA
mgDRM2a-R	CGGGTCTCATAGAGACTGCGCTGCACCAGCCGGG
mgDRM2b-F	TAGGTCTCCAAACCCCATCATGTTTTAGAGCTAGAA
mgDRM2b-R	CGGGTCTCGGTTTGGTATTCCTGCACCAGCCGGG
mgDRM2c-F	TAGGTCTCATGTATCCTAAAGGTTTTAGAGCTAGAA
mgDRM2c-R	CGGGTCTCATACATCTCCTTCTGCACCAGCCGGG
mgDRM2L2d-F	TAGGTCTCCTCTCCATCGACTGTTTTAGAGCTAGAA
mgDRM2L2d-R	CGGGTCTCGGAGAGCTACTTCTGCACCAGCCGGG
mgDRM2L2e-F	TAGGTCTCGCAAACAATATCTGTTTTAGAGCTAGAA
mgDRM2L2e-R	CGGGTCTCGTTTGCCAAACACTGCACCAGCCGGG
mgDRM2L2f-F	TAGGTCTCTTTCCTCAATAGGGTTTTAGAGCTAGAA
mgDRM2L2f-R	CGGGTCTCAGGAAATGGTGGCTGCACCAGCCGGG
sequencing primers	
DRM2-F1	GGTTGTTTCATTGGCCTCACTGT
DRM2-F2	GAATGCCCTTCTGCTTCCTG
DRM2-F3	CGCTATCCGGGATTGACAGCTA
DRM2-F4	CTGTTTTTCCCAATGAAACTTTCCA
DRM2-F5	CCTTTACTAAGGAGGTGGTGGCC
DRM2-R1	AGGATGCATTCTTACCGATGAACA
DRM2-R2	ACTGACTCAACCTAACCAACCA
DRM2-R3	CGCGGTGGAAGAGGTAAAAGAG
DRM2-R4	GATTCCGAGCCCTCAAGTCCAT
DRM2-R5	TCTAGGCGCCATAACCTTCACT
DRM2L2-F1	GGCCTAGACTCCTCACTTGACG
DRM2L2-F2	GCAGTGGGCATTCAAAACCAAG
DRM2L2-R1	CACTCATGAGCTGCTCCAAACG
DRM2L2-R2	CCAACAACGAGATCAAAACCACCA
DRM2L2-R3	GGCAACTTTATTCTTGCCCACCC
Chop-PCR primers	
Chr1-26645685..26647163_F	AGCTGGTGGGGTCCATGACACA
Chr1-26645685..26647163_R	AGCAGGCTTACAACCTTCTTGGC
Chr1_35839283..35840045_F	ACCCACCGAAATCGACACTAGGGG
Chr1_35839283..35840045_R	GGTTGAGCTGAGTTGAGATTTGCA
Chr2_32878129..32878605_F	GTTGAGGATGTGGTCGTCTTAGCT
Chr2_32878129..32878605_R	TGCACAACCGCGACAAGAAGGA
Chr3_51183413..51184086_F	CCGAAAAATGACGCGACATACCCCA
Chr3_51183413..51184086_R	TCGGTATTTTCAAATTCACCGCTGT
Chr4_24118045..24119280_F	TCCCTTGACTTGGGACCCACTT
Chr4_24118045..24119280_R	CGGCATTGGTGAAAACGTGTCACG
Chr4_30558968..30560144_F	AGACGAGCCACACAATCCAGTG
Chr4_30558968..30560144_R	TTCTTCGTCTTCCCAAGGCCCA
Chr4_31798378..31799002_F	AGACACCAGGAGACAAGCCCTGA
Chr4_31798378..31799002_R	ACCGTGGAAGCTATGTCTGAACGT
Chr4_9710107..9710722_F	TGCTCAAAAAGGTGCCACGTGT
Chr4_9710107..9710722_R	GGTTCAAACCCTGGACCTGGCA

Supplemental note 1. Synthetic *DRM2-R* sequence used for cloning a non-editable Mt*DRM2* cDNA

*ACGTGCAGAAGACAAGTAAGGAG*gacgtcgttgtggttggtgctttccttacattctgagcctctttccttctaatccactcatctgcatcttcttgtgtcc
ttactaatacctcattggttccaaattccctccctttaagcaccagctcgtttctgttcttccacagcctcccaagtatccaagggactaaagcctccacattcttcag
atcaggatattcttgtttaagatgttgaactctatggaggtttgtatgaactgatgatctaggaccggataagttcccttcttcatagcgaacttattcaaagaatgt
tttgtgtatcattcttgttacattgttattaatgaaaaaatattattggtcattggactgaacacgagtgttaaatatggaccaggccccaaataagatccattgata
tatgaattaaataacaagaataaatcgagtcaccaaaccacttgccttttttaacgagacttgttcaccaacttgatacaaaagtcattatcctatgcaaatcaat
aatcatacaaaaatatccaataacactaaaaaaattaaaagaaatggataatttcacaatatgttatacgataaagaagttacttttccaagaaattcactgatt
ttataagcccacttgcattagataaatggcaaaaaaaaacaaaaaggaaaagaaataaagcacgaagaattctagaaaatacgaaatacgcttcaatgcagt
gggacccacggttcaattattgccaattttcagctccaccgtatatttaaaaaataaaacgataatgctaaaaaaatataaatcgtaacgatcgttaaatctcaac
ggctggatcttatgacgaccgttaaggaaattgtggttgtcggacgaagtccagtaataaacggcgtcaaagtggttgcagccggcacacacgagtcgtgtttat
caactcaaagcacaaatacttttcctcaacctaaaaataaggcaattagccaaaaacaactttgcgtgtaaacaacgctcaatacacgtgtcattttattattagc
tattgcttcaccgccttagctttctcgtgacctagtcgtcctcgtcttttcttcttcttcttctataaaacaatacccaaagagctcttcttcttcacaattcagatttca
atttctcaaaatcttaaaaactttctctcaattctctctaccgtgatcaaggtaaatttctgtgttccttattctctcaaaatcttcgattttgttttcgttcgatcccaat
ttcgtatatgttctttggtttagattctgttaatcttagatcgtagacgattttctgggtttgatcgttagatatcatcttaattctcgattagggtttcatagatatcatc
cgatttgttcaaataatttgagttttgtcgaataattactcttcgatttgtgatttctatctagatctggtgttagtttctagtttgtgcgatcgaatttgtcgattaatct
gagtttttctgattaacagatgcagatctta*ATGGGTGACTATTCTGGTCTGGATAGTGATATAGATTGGAACACCGATGATGAGCTT*
GAGATTGAGAGTTTTCAACCCTCCTGTTCAACTGTTGTTCCCAGTGGGCAGACTATTACTGCTGGGTCTGTAGAGCCAAGC
TCATTCGCAGGTCCATCTAACACCAAGGTGTTTGACCACTTCATCAGTATGGGATTTCCTGGTGAAGTCGTTTCAAAAGTCA
TTCAAGAGCATGAGGAGAATGAAGAGAAACTGCTTAATGAGATTCTCACATACTCAGTTCTTGAAAGTTCTCCTCAGCAGC
AACAGCCAGCTGAACTAGACCCCACCTCATCCGAGTGTGCGGGGAGTTCGTGGGAGGATTTATCGGAAGATGATTTTTTTT
CTGATGAAGAATTGCCAAAATTCGATTCTACCAATGATGATACGTTGACGAAACTTGTAAAAATGGGATTTGAAGAGGAG
GAGGCTTTAGTGGCCATCGACAGAATATCAGACTCACTTGAAGCTCTGGTCGATTTTATTGGTGCTGCTCAAGTGGCGAAA
*GCAGAGAATGCCCTTCTGCTTCCTGAAGATAAGCCGGGATGCTCCGGGAATCCAAAGCTGAAAAAGCG*aAGcCTaTAcGAg
TAcGAa*GTCCTTGGGAAGAAAAGACCGAAGTTAGAGAAGAGAACACTATGTGAAGAtGAGGAGGAGGCGCAAACGCTTA*
*ATCTGCCTAATCCTATGATGGG*tTTcGGcATc*CCTAATGAGCCCAGTTCTATGATTACACACAGGAGACTGCCTGAAAATGC*
*CGTTGGCCCGCCCTATTTCTACTACGAGAATGTGGCGTTGGCACCGAAAGGcGT*aTGGCAgACgATaTCgcGa*TTCTTATATG*
ATGTGGAACCCGAGTACGTCGACTCAAAATACTTTTGTGCTGCTGCGCGTAAACGTGGATATATTCACAATCTCCCAATTGT
*GAATAGATTCCCTCTTTTACCTCTTCCACCGCGCACCATCCATGATGCATTtCC*actAttgcgacgc*TGGTGGCCGACATGGGAC*
CCCAGGACAAAGCTTAATTGTTTGCAGACCGTACATGCTAGTGCGAAACTTACCGACAGAATCCGGAAAGCGGTAGAAAG
TTGTGATGATTTTGAGGAACCTTCTGAAACGGTAAAAAAATATGTGTTGGATCAGTGTCGGAAATGGAACCTGGTATGGG
TGGGTAAGAATAAAGTTGCCCCCCTAGAGCCCGATGAAGTTGAGATGCTGTTGGGATTCCCTAGGAATCATACAAGGGGT
GGTGGAATCAGTAGGACGGACAGATTCAAGTCACTTGGAAATTCTTTCCAGGTCGACACAGTAGCATATCATCTATCAGTT
*TT*aAAaGAaATGTAcCCaAAgGGa*GTGAATCTTCTATCTCTTTTTTCTGGAATTGGTGGTGCAGA*aGTtGCaCTtCAcAGgCTa
GGCATCCCTCTAAGGAATGTAGTGTCAGTCGAGAAATCTGAAGTGAACAGGAACATTGTTAGGAGTTGGTGGGAGCAAA
CCAACCAAAGGGGTAATTTAGTTGATTTTGATGATGTGCAACAGCTAGATGCCGACCGTTTGGAGCGGCTGATGGGCGCA
TTTGGTGGCTTTGATCTAATTGTTGGTGGAAGCCCTTGCAATAATCTGGCTGGAAGCAATAGGGTTAGTCGGAATGGACTT
GAGGGCTCGGAATCTATCCTATTTTATGAATACTTTAGGATTTTAGACTTAGTGAAGGTTATGGCGCCTAGATTTCGATGAc
tctagctagagtcgatcgacaagctcgagtttctccataataatgtgtgagtagttcccagataagggaattagggttcctatagggtttcgctcatgtgttgagca
tataagaaacccttagtatgtatttgtatttgtaaaatacttctatcaataaaatttctaattcctaaaaccaaaatccagtactaaaatccagat*CGCTGAGCT*
CGAATTCTAGTTTGTCTTCACAGA

XXXX : sequences added for Golden Gate cloning using BpiI
xxxx *: AtUBI* promoter
xxxx : 35S terminator
XXXX : *DRM2* cDNA
XXXX : mutated *DRM2* sequences (targeted by DRM2 guides)
x: mutated nucleotides