Supplementary Materials for

A Histone Acetyltransferase Regulates Active DNA Demethylation in *Arabidopsis*

Weiqiang Qian, Daisuke Miki, Heng Zhang, Yunhua Liu, Xi Zhang, Kai Tang, Yunchao Kan, Honggui La, Xiaojie Li, Shaofang Li, Xiaohong Zhu, Xiaobing Shi, Kangling Zhang, Olga Pontes, Xuemei Chen, Renyi Liu, Zhizhong Gong, Jian-Kang Zhu*

*To whom correspondence should be addressed. E-mail: jkzhu@purdue.edu

Published 15 June 2012, *Science* **336**, 1445 (2012)
DOI: 10.1126/science.1219416

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/336/6087/1445/DC1)

Materials and Methods

Plant materials and mutant screening by CHOP PCR

Arabidopsis homozygous T-DNA insertion lines were obtained from the Arabidopsis Biological Resource Center (http://www.arabidopsis.org). Arabidopsis seedlings were grown on Murashige-Skoog (MS) nutrient agar plates at 23ºC with 16 h of light and 8 h of darkness for 2 weeks, and 5 individual plants for each T-DNA line were harvested for genomic DNA isolation. We conducted genome-wide profiling of DNA methylation in *ros1-1* mutant plants in the C24 ecotype (*6*) and identified a number of loci likely having greater DNA methylation than wild-type C24 plants. Based on these DNA methylation data, a PCR-based marker was developed for reporting DNA methylation status at the 3' region of *At1g26400* (also referred to as *DT-77*). To screen for *idm* (*increased DNA methylation*) mutants, 1 μg of genomic DNA was digested with the methylation-sensitive enzyme *Hha*I in a 20 μl reaction mixture. After digestion, PCR was performed using 1 μl of the digested DNA as template in a 10 μl reaction mixture and using the *At1g26400/DT-77* locus-specific primers. Other CHOP PCR was performed using the same strategy but different restriction enzyme and primer sets (table S7).

Plasmid construction and mutant plant complementation

For complementation of mutants, a 7.6 kb genomic DNA fragment containing the *IDM1* gene was amplified from Col genomic DNA by PCR and cloned into the pSAT-3HA-YFP vector. The cassette of *pIDM1::IDM1:3HA:YFP* were excised at the *PI-Psp*I restriction site and cloned into the pHPT binary vector for plant transformation. All mutation sites were introduced into *pIDM1::IDM1* (pCAMBIA 1303 vector) through site-directed mutagenesis with the 'QuikChange' kit according to the manufacturer's instructions (Stratagene). *Agrobacterium tumefacines* strain GV3101 carrying various *IDM1* constructs was used to transform mutant plants via the standard floral dipping method. Primary transformants were selected on MS plates containing hygromycin. CHOP PCR results for *DT-77* and *At4g18650* promoter regions are presented for two representative transgenic lines generated from each construct.

Individual locus bisulfite sequencing

A 100 ng quantity of extracted genomic DNA was analyzed by sodium bisulfite genomic sequencing using the BisulFlash DNA Modification Kit (Epigentek; http://www.epigentek.com/catalog/index.php) following the manufacturer's protocol. A 1 μl aliquot of bisulfite-treated DNA was used for each PCR reaction. PCR was performed in a total volume of 20 μl using ExTaq (Takara; http://www.takarabiousa.com/). Sequenced fragments were amplified with primers specific for each region (table S7). Amplified PCR products were subcloned into a pCRII-TOPO vector (Invitrogen; http://www.invitrogen.com/site/us/en/home.html) following the supplier's instructions. For each region, at least 15 independent top-strand clones were sequenced from each sample.

Real-time RT-PCR

For real-time RT-PCR analysis, total RNA was extracted using the RNeasy plant mini kit (Qiagen), and contaminating DNA was removed with RNase-free DNase

(Qiagen). mRNA (2 µg) was used for the first-strand cDNA synthesis with the Super script^{TM} III First-Strand Synthesis System (Invitrogen) for RT-PCR following the manufacturer's instructions. The cDNA reaction mixture was then diluted five times, and 2 µl was used as template in a 25 µl PCR reaction with iQ SYBR Green Supermix (Bio-Rad). PCR included a preincubation at 95ºC for 5 min followed by 40 cycles of denaturation at 95ºC for 15 s, annealing at 56ºC for 30 s, and extension at 72ºC for 30 s. All the reactions were carried out on the iQ5 Multicolour Real-Time PCR Detection System (Bio-Rad). The comparative threshold cycle (Ct) method was used for determining relative transcript levels (Bulletin 5279, Real-Time PCR Applications Guide, Bio-Rad), with *TUB8* as an internal control.

DNA methylation assay by Southern hybridization

For Southern hybridization assays, 5 µg of genomic DNA was digested with *Hpa*II or *Msp*I. The digested DNA was loaded onto a 1.2% agarose gel and transferred to Hybond-N+ membranes. The 180-bp DNA repeat, 5S rDNA repeat, and 45S rDNA repeat were labeled with (α-^{32}P) dCTP for Southern hybridization to determine their DNA methylation status.

Construction and sequencing of whole genome bisulfite sequencing libraries

Libraries were prepared according to a published protocol (*9*). Briefly, five µg of genomic DNA was sonicated and then end-repaired with the End Repair Enzyme Mix (New England BioLabs, Ipswich, MA). The blunt-ended DNA were treated with Klenow fragment (New England BioLabs, Ipswich, MA) and dATP to yield a protruding 3- 'A' base and cytosine-methylated adaptors from Illumina (Illumina, San Diego, CA) were ligated according to manufacturer's instructions. Unmethylated cytosine residues were converted to uracils using the EpiTect bisulfite Kit (Qiagen, Valencia, CA). Bisufite treated DNA was amplified by PCR with Illumina primers for 15 cycles and fragments of 250-350 bp were excised from an agarose gel and purified. Paired end sequencing of each library was performed to around 100 cycles on the Illumina Genome Analyzer II or HiSeq 2000 according to manufacturer's protocols.

Processing and mapping of bisulfite sequencing reads

Raw reads were first trimmed to the longest contiguous segment with Phred quality score $\geq$ 20 using the DynamicTrim program from the SolexaQA software (*12*) and then trimmed for the adaptor sequence using an in-house perl script. Clean reads that were shorter than 20 nt were discarded. Using BRAT(*13*) with default parameters, clean reads were mapped to the Arabidopsis genome (TAIR10) as paired-end reads if reads from both ends were retained in the previous steps, and as singletons if read from the other end was discarded. Methylation levels of cytosines on both Watson and Crick strands were calculated by dividing number of mapped C's with total number of mapped reads at each position.

Identification of differentially methylated cytosines (DMCs)

We used an algorithm similar to that in Lister et al (*14*) to identify DMCs. For each genotype, mapping results from two biological replicates were combined. To identify DMCs in the wild type (WT) and a mutant (*idm1*, *ros1*, or *rdd*), we considered cytosine

positions that had a combined coverage from two replicates between 4 and 100 in both WT and the mutant. A cytosine was identified as DMC if the p-value from the two-tailed Fisher's Exact test was 0.01 or better.

Identification of differentially methylated regions (DMRs)

To identify DMRs that were hyper- or hypo-methylated in a mutant compared to the WT, we divided the genome into 1 kb regions and counted number of DMCs in each region. A region was selected as an anchor region if it had at least 5 DMCs. The actual boundary of each anchor region was then adjusted as the locations of first and last DMCs in the region. If the distance between two anchor regions was ≤ 1kb, they were combined into a larger region. At the end of extension process, regions that contained at least 10 DMCs were reported as DMRs.

Small RNA density in hypermethylated regions in *idm1-1*

To test whether the 1098 hypermethylated regions that were identified in *idm1-1* are enriched with small RNAs (sRNAs), we used the sRNA data that were generated in the wild type and *rdd* mutant plants by Lister et al (*9*). We downloaded the raw sequence data from the NCBI Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra) using accession numbers SRX002508 and SRX002511, respectively. After removing adaptor sequences with in-house Perl scripts, we mapped clean reads from each library to the Arabidopsis genome (TAIR10) using SOAP (*15*) and retained sRNAs that had perfect match to a unique location in the genome. For each genotype, we counted total number of sRNAs that fall into the 1098 hypermethylated regions and calculated the average sRNA density by dividing total sRNAs by the total length of these regions. To test whether the sRNA density in *idm1-1* hypermethylated regions is higher than expected by chance, we performed 1000 simulations. In each simulated run, we randomly selected 1098 regions with the same length distribution from the genome and we calculated the sRNA density in the randomly selected regions. Level of significance for sRNA enrichment (p-value) was inferred by comparing the sRNA density in the hypermethylated regions to the 99 percentile of sRNA densities from the 1000 simulations.

Calculation of H3K4 methylation at expressed genes with high mCG and at DT loci

Whole genome mono-, di-, and tri-methylation of histone H3 lysine 4 (H3K4me1, H3K4me2 and H3K4me3, respectively) in Arabidopsis was determined by Zhang et al. (2009) at 35 bp resolution (*10*). Their results showed that all three types of H3K4me were almost exclusively genic and that H3K4me2 and H3K4me3 accumulated predominantly in promoters and 5' genic regions and H3K4me1 was distributed within transcribed regions. To test whether the DT loci are preferentially associated with low H3K4 methylation, we compared the H3K4 methylation at DT loci (table S1) to that in highly methylated and expressed genes.

Using the whole genome methylation data in WT, we calculated total number methylated C in CG context (mCG) and percentage of mCG out of all CG positions (%mCG) in each annotated gene. We chose a subset of 3417 genes that contained at least 10 mCG and %mCG ≥ 40% as highly methylated genes. We also profiled gene expression in WT Arabidopsis using the Affymetrix GeneChip ATH1 genome array (data not shown) and chose a subset of 4472 genes as expressed genes that had an average log2

expression values ≥ 7 in two biological replicates. A set of 360 genes was chosen as highly methylated and expressed genes because they satisfied both criteria. Twenty-three of these genes were removed because they overlapped with the DT loci, resulting in a set of 337 highly methylated and expressed genes. Each gene was evenly divided into 10 regions ordered from the 5' to 3' end and H3K4me1, H3K4me2 and H3K4me3 values in each region were calculated as the average of all tiling array probes within the region. The overall H3K4 methylation over each region was the average of the 337 genes. The calculation of H3K4 methylation at DT loci was the same as that at the highly methylated and expressed genes, except that target regions were not ordered according to direction of transcription.

Peptide microarray and peptide pull-down assays

The PHD domain (720–776 aa) of IDM1 was cloned into the bacterial expression vector pGEX4T-1. All mutation sites were introduced by site-direct mutagenesis with the 'QuikChange' kit according to manufacturer's instructions (Stratagene). All of the constructs were transformed into *E. coli* BL21 (DE+) for protein expression. Recombinant proteins were purified using the Glutathione Sepharose^{TM} 4B column (GE Healthcare). Peptide microarray and peptide pull-down assays were performed as described previously (*16*). Briefly, biotinylated histone peptides were printed in triplicates onto a streptavidin-coated slide (Arrayit Corp.) using a VersArray compact microarrayer (Bio-Rad). After a short blocking with biotin (Sigma), the slides were incubated overnight at 4ºC with the GST-fused PHD finger in binding buffer (50 mM Tris-HCl, pH 7.5; 200 mM NaCl; and 0.1% Nonidet P-40). After they were washed with the same buffer, the slides were probed with anti-GST antibody and fluorescein-conjugated secondary antibody and visualized using a GenePix 4000 scanner (Molecular Devices). The microarray data was quantified by software GenePix Pro6.0 (Molecular Devices). For peptide pull-down assays, 1 μg of biotinylated histone peptides with different modifications was incubated overnight with 1 μg of GST-fused PHD fingers in binding buffer. Streptavidin beads (Amersham Biosciences) were added to the mixture and incubated for 1 h with rotation. The beads were then washed three times and analyzed by SDS-PAGE and Western blotting.

Protein extraction and detection by Western blotting

Two grams of unopened flower buds from wild-type or transgenic plants were harvested and ground to fine power in liquid N_2. Total protein was extracted by protein extraction buffer (20 mM Tris-HCl [pH8.0], 150 mM NaCl, 1 mM EDTA, 10% glycerol, 5 mM DTT, 1 mM PMSF and 1x protease inhibitor cocktail (Roche)). The samples were centrifuged at 12000g at 4ºC for 30 min. To enrich the IDM1 protein for detection, immunoprecipitation was then performed by incubating approximately 10 mg total protein of the supernatants with an excess amount of anti-IDM1 antibody (anti- IDM1 peptide antibody custom made by YenZym Antibodies) at 4ºC for 3 hours with rotation, followed by adding 50 μl protein A agarose beads (Millipore) for additional 2 hours. The beads were washed 3 three times with extraction buffer. The immunoprecipitates were analyzed by Western blotting with anti-IDM1 peptide antibody.

Electrophoretic mobility shift assays (EMSA)

The N-terminal region (1–400 aa) of IDM1 (IDM1-N), containing the MBD domain, was cloned into the pET28a vector (Invitrogen) and transformed into *E. coli* BL21 (DE3+) for protein expression. Recombinant protein was purified using Ni columns (Qiagen). Unless specified otherwise, methylated and unmethylated DNA oligonucleotides for the EMSA were from Integrated DNA Technologies and were the same as used in Woo et al (*17*). For binding assays, double-stranded DNA oligonucleotides were end-labeled using T4 kinase (NEB) and radioactive γ^{32}P ATP. Labeled DNA oligonucleotides (0.6 pmol) were incubated with recombinant protein IDM1-N (4 μg) in the presence or absence of cold competitors (3 pmol unless stated otherwise) in a 20 μl binding reaction. Binding reactions were carried out in 25 mM HEPES (pH 7.6), 50 mM KCl, 0.1 mM EDTA (pH 8.0), 12.5 mM $MgCl_2$, 1 mM dithiothreitol, and 5% (w/v) glycerol; 1.5 mg of poly-dIdC was added as a non-specific competitor. After incubation for 45 min at room temperature, the reaction mixture was subjected to 8% PAGE with 0.5x TBE running buffer. Methylated double-stranded DNA oligonucleotides are shown in fig. S12B.

Purification of IDM1 protein from Sf9 insect cells

Sf9 cells were infected with baculovirus harboring recombinant IDM1-C1-FLAG or IDM1-C2-FLAG gene and grown at 27ºC for 48 hours before harvest. Cells were collected by centrifugation at 500x g at room temperature for 5 min and washed once with room temperature PBS (phosphate buffered saline). The following steps were performed on ice or in a cold room. Per 10e7 cells, add 1 ml of cold Protein Extraction Buffer (50 mM Hepes-KOH pH 7.5, 500 mM NaCl, 1% Triton X-100, 10% glycerol, 0.4 mM EDTA, 2 mM DTT, 1 mM PMSF, 1x protease inhibitor cocktail (Roche 05 892 791 001)) to resuspend the cells. After 10 min incubation on ice, centrifuge at 4ºC for 15 min and transfer the supernatant to a new tube. To the supernatant add an equal volume of cold Dilution Buffer (50 mM Hepes-KOH, pH 7.5, 10% glycerol, 1 mM PMSF, 1x protease inhibitor cocktail) and 50 μl of anti-FLAG M2 magnetic beads (Sigma M8823; pre-equilibrated with Wash Buffer). Incubate at 4ºC for 3 hours with rotation. Then the magnetic beads were washed with 1 ml of Wash Buffer (50 mM Hepes-KOH pH 7.5, 250 mM NaCl, 0.5% Triton X-100, 10% glycerol, 0.2 mM EDTA, 1 mM DTT, 1 mM PMSF) for 10 min. Repeat wash 3 more times. The FLAG-tagged proteins were then eluted by incubation with 250-μl Elution Buffer for 30 min (50 mM Hepes-KOH pH 7.5, 150 mM NaCl, 0.2 mM EDTA, 1 mM DTT, 25% glycerol, 0.15 mg/ml 3xFLAG peptide (Sigma F4799). The elution was repeated and eluates combined.

Histone acetyltransferase assay

HAT assays were performed at 30ºC in 30-μl reactions. In general, each reaction contains 0.5 μg of PCAF (ab56275, Abcam) or similar amount of purified IDM1-C1-FLAG or IDM1-C2-FLAG protein, 2 μg of recombinant histone H3 (14-411, Millipore) or 8 μg of chicken core histones (13-107, Millipore) and 1 μCi of ^{3}H-acetyl-CoA (NET290L250UC, Perkin-Elmer Life Science) in assay buffer (50 mM Tris-Cl pH8.0, 100 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 5% glycerol). Reactions were incubated at 30ºC for 1-2 hours, and then 10 μl 4x SDS-PAGE sample buffer was added and samples heated at 95ºC for 5 min. Half of the reaction was loaded onto a SDS-PAGE minigel. Proteins were separated and transferred to a PVDF membrane using a semi-dry blotter

(TE70, GE Life Sciences). Proteins on the membrane were then visualized by Coomassie blue staining. After drying down in a chemical hood, image of the stained membrane was recorded with a scanner. ^{3}H signal was detected with a Kodak BioMax Transcreen Intensifying Screen LE and X-ray films.

LC-MS/MS analysis

IDM1 and PCAF treated samples were digested with trypsin in 25 mM ammonium bicarbonate buffer. After SpeedVac dried, the resultant peptides were dissolved in 20 µl of 2% formic acid and then subjected to LC-MS/MS analysis. A peptide CapTrap (Michrom Biosciences) was brought inline with an analytical column (Thermo Scientific EasyNano column, 10 cm, 75 µm i.d., 3 µm particle size) and a 85 minute gradient (solvent A, 0.1% formic acid in water; solvent B, 0.1% formic acid in acetonitrile) from 2% to 85% solvent B with three steps (2-30% solvent B, 30-55% solvent B, and 55-85 solvent B%) was used for separating the peptides. The Thermo Scientific LTQ Orbitrap Velos mass spectrometer coupled with a Thermo Scientific Easy-nLC II HPLC was set to acquire data at a resolution of 60,000 FWHM for the parent full-scan mass spectrum followed by data-dependent HCD MS/MS spectra for the top 10 most abundant ions acquired at a resolution of 7,500 FWHM.

ChIP assays

Chromatin immunoprecipitation (ChIP) assays were performed according to a published protocol (*18*). The following antibodies were used for ChIP assays: anti-H3K4me2 (07-030, Millipore), anit-H3K18ac (ab1191, Abcam), anti-H3K23ac (07-355, Millipore), anti-HA (H9658, Sigma), and anti-MYC (M4439, Sigma). ChIP products were eluted into 50 µl of TE buffer, and a 2-µl aliquot was used for each qPCR reaction. Two biological replicates were performed, and very similar results were obtained.

Immunolocalization

Immunostaining was performed as described by He et al (*19*). Briefly, nuclei were isolated from WT, *idm1-1* and *MYC-ROS1* (*20*) plant leaves. Nuclei preparations were incubated overnight at 4ºC with primary antibodies for anti-IDM1 (1:100) and anti-cMyc (1:200, Millipore). Then the nuclei samples were incubated with anti-mouse Alexa488 and anti-rabbit Alexa-594 secondary antibodies (1:200, Molecular Probes) for 3 hours at 37ºC. DNA was counterstained using DAPI in Prolong Gold (Invitrogen). Nuclei were examined using a Nikon Eclipse E800i epifluorescence microscope equipped with a Photometrics Coolsnap ES.

References

1. M. Tariq, J. Paszkowski, DNA and histone methylation in plants. *Trends Genet.* **20**, 244 (2004). doi:10.1016/j.tig.2004.04.005 Medline

2. J. A. Law, S. E. Jacobsen, Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204 (2010). doi:10.1038/nrg2719 Medline

3. J. K. Zhu, Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* **43**, 143 (2009). doi:10.1146/annurev-genet-102108-134205 Medline

4. Materials and methods are available as supplementary materials on *Science* Online.

5. J. Zhu, A. Kapoor, V. V. Sridhar, F. Agius, J. K. Zhu, The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in *Arabidopsis*. *Curr. Biol.* **17**, 54 (2007). doi:10.1016/j.cub.2006.10.059 Medline

6. Z. Gong *et al*., ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. *Cell* **111**, 803 (2002). doi:10.1016/S0092-8674(02)01133-9 Medline

7. Y. Onodera *et al*., Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**, 613 (2005). doi:10.1016/j.cell.2005.02.007 Medline

8. J. Penterman *et al*., DNA demethylation in the *Arabidopsis* genome. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6752 (2007). doi:10.1073/pnas.0701861104 Medline

9. R. Lister *et al*., Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523 (2008). doi:10.1016/j.cell.2008.03.029 Medline

10. X. Zhang, Y. V. Bernatavichute, S. Cokus, M. Pellegrini, S. E. Jacobsen, Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol.* **10**, R62 (2009). doi:10.1186/gb-2009-10-6-r62 Medline

11. A. Kirmizis *et al*., Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature* **449**, 928 (2007). doi:10.1038/nature06160 Medline

12. M. P. Cox, D. A. Peterson, P. J. Biggs, SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **11**, 485 (2010). doi:10.1186/1471-2105-11-485 Medline

13. E. Y. Harris, N. Ponts, A. Levchuk, K. L. Roch, S. Lonardi, BRAT: Bisulfite-treated reads analysis tool. *Bioinformatics* **26**, 572 (2010). doi:10.1093/bioinformatics/btp706 Medline

14. R. Lister *et al*., Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315 (2009). doi:10.1038/nature08514 Medline

15. R. Li *et al*., SOAP2: An improved ultrafast tool for short read alignment. *Bioinformatics* **25**, 1966 (2009). doi:10.1093/bioinformatics/btp336 Medline

16. H. Wen *et al*., Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. *J. Biol. Chem.* **285**, 9322 (2010). doi:10.1074/jbc.C109.097667 Medline

17. H. R. Woo, O. Pontes, C. S. Pikaard, E. J. Richards, VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. *Genes Dev.* **21**, 267 (2007). doi:10.1101/gad.1512007 Medline

18. A. Saleh, R. Alvarez-Venegas, Z. Avramova, An efficient chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in *Arabidopsis* plants. *Nat. Protoc.* **3**, 1018 (2008). doi:10.1038/nprot.2008.66 Medline

19. X. J. He *et al.*, NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and V and is required for RNA-directed DNA methylation. *Genes Dev.* **23**, 318 (2009). doi:10.1101/gad.1765209 Medline

20. X. Zheng *et al.*, ROS3 is an RNA-binding protein required for DNA demethylation in *Arabidopsis*. *Nature* **455**, 1259 (2008). doi:10.1038/nature07305 Medline

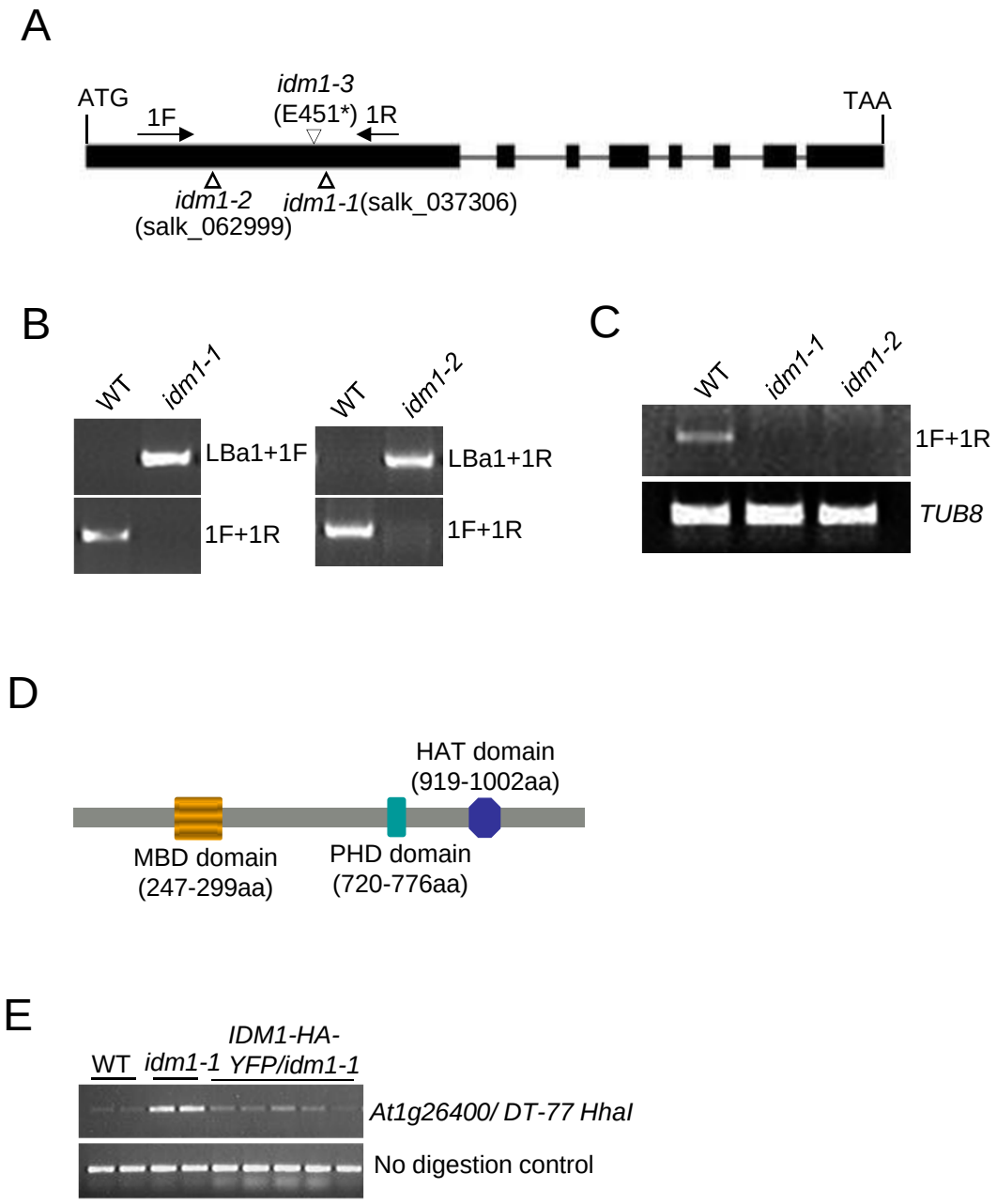


fig. S1. Characterization of *idm1* mutants. (**A**) Schematic diagram showing the positions of the T-DNA insertions or nucleotide change at the *IDM1* locus. Black rectangles represent exons. Arrows indicate PCR primers used in genotyping and RT-PCR analysis of T-DNA alleles. (**B**) Genotyping of *idm1* T-DNA alleles. (**C**) RT-PCR analysis of *IDM1* transcript levels in *idm1* T-DNA mutants. *TUB8* served as a positive control. (**D**) Domain structure for IDM1 protein. (**E**) Expression of the genomic DNA containing *IDM1* with 3xHA and YFP tags complemented the *idm1-1* mutant phenotype. CHOP PCR results for *DT-77* are shown for four representative transgenic lines.

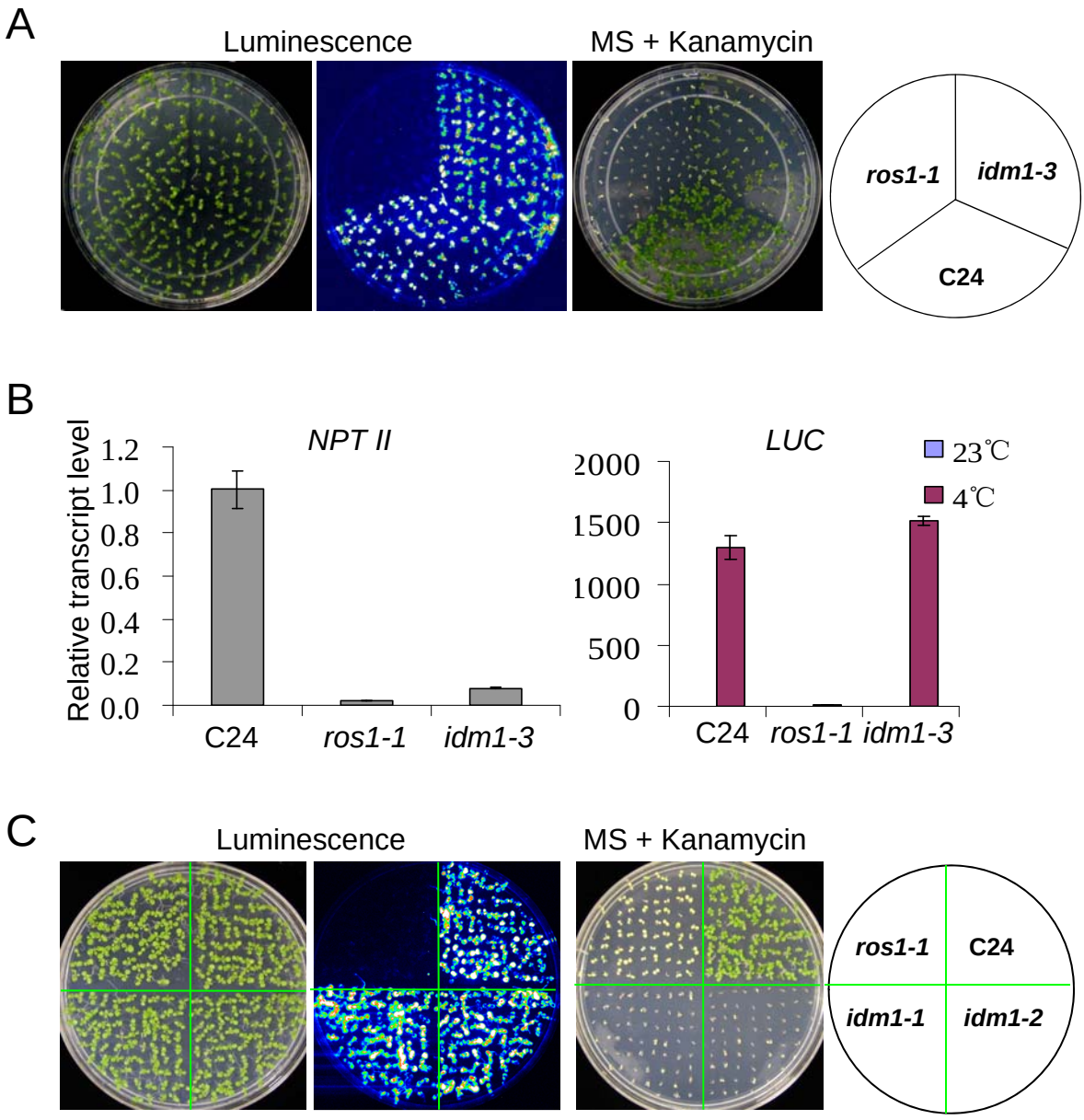


fig. S2. Effects of *idm1* on *RD29A-LUC* and *35S-NPTII* reporter genes as indicated by luciferase activity and kanamycin resistance of *idm1* mutants harboring the reporters. (**A**) The *idm1-3* mutation causes the silencing of *35S-NPTII* but not the *RD29A-LUC* transgene. Seedlings grown in MS plate were imaged after cold treatment at 4ºC for 24 hours. For kanamycin resistance test, the seeds were planted on MS medium supplemented with 50 mg/L kanamycin and incubated for 2 weeks before being photographed. (**B**) Real-time PCR analysis of the expression level of *LUC* and *NPTII* reporter genes in the different genotypes. *TUB8* was used as an internal control. Error bars represent standard error (n=3). (**C**) Effect of *idm1-1* and *idm1-2* on *RD29A-LUC* and *35S-NPTII* reporter genes. The reporter genes were introduced to *idm1* mutant plants by crossing.

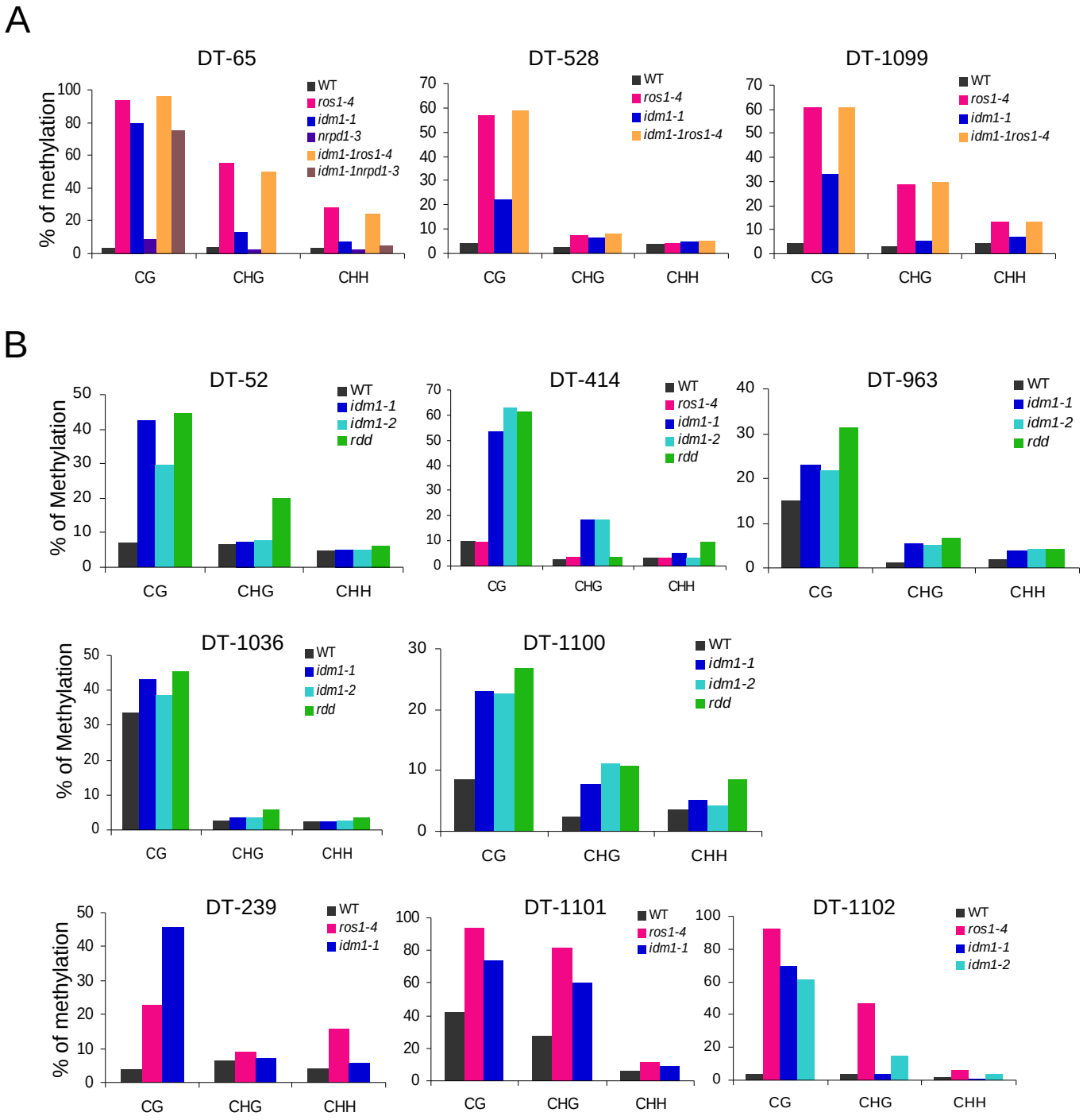


fig. S3. DNA methylation levels assayed by individual locus bisulfite sequencing in wild-type, *idm1*, and other mutant plants. (**A**) Bisulfite sequencing data showing the effects of *idm1-1* mutation on the DNA methylation of three loci and genetic interactions with *ros1* or *nrpd1*. (**B**) Bisulfite sequencing results in wild type, *idm1*, and *ros1-4* or *rdd* mutant plants.

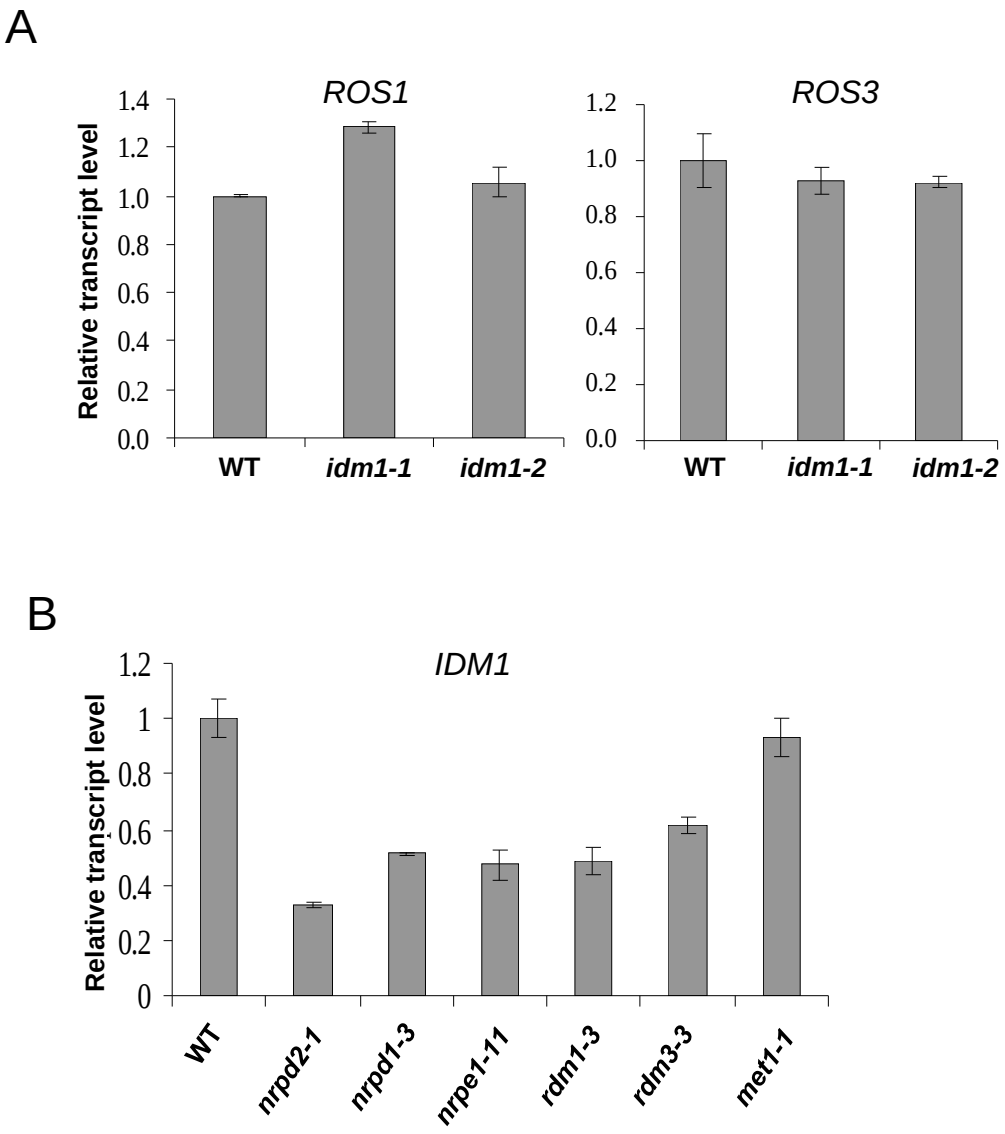


fig. S4. Real-time PCR analysis of transcript levels. (**A**) *ROS1 and ROS3* mRNA levels in wild-type and *idm1* mutant plants. Three biological replicates were performed, and very similar results were obtained. Standard deviations were calculated from three technical repeats. The transcript levels were normalized using *TUB8* as an internal standard. (**B**) *IDM1* transcript levels in various RdDM pathway mutants and the *met1-1* mutant.

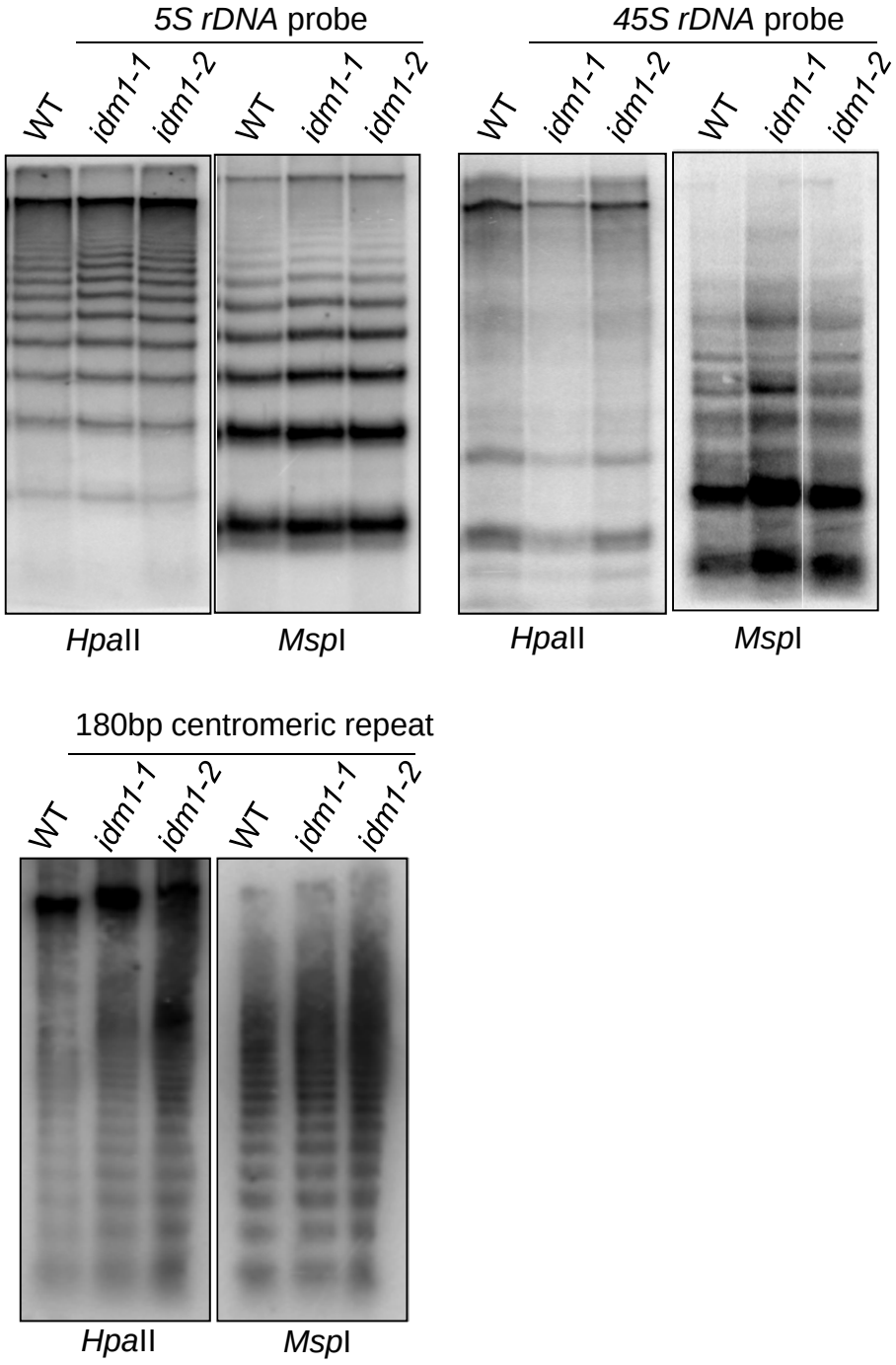


fig. S5. Methylation status of rDNA and centromeric DNA in wild-type and *idm1* mutant plants. Genomic DNA from wild-type and *idm1* mutant plants was digested with the methylation sensitive enzyme *Hpa*II (for CG & CHG methylation) or *Msp*I (for CHG methylation), and hybridized with 5S rDNA, 45S rDNA, or180-bp repeat probes.

WT

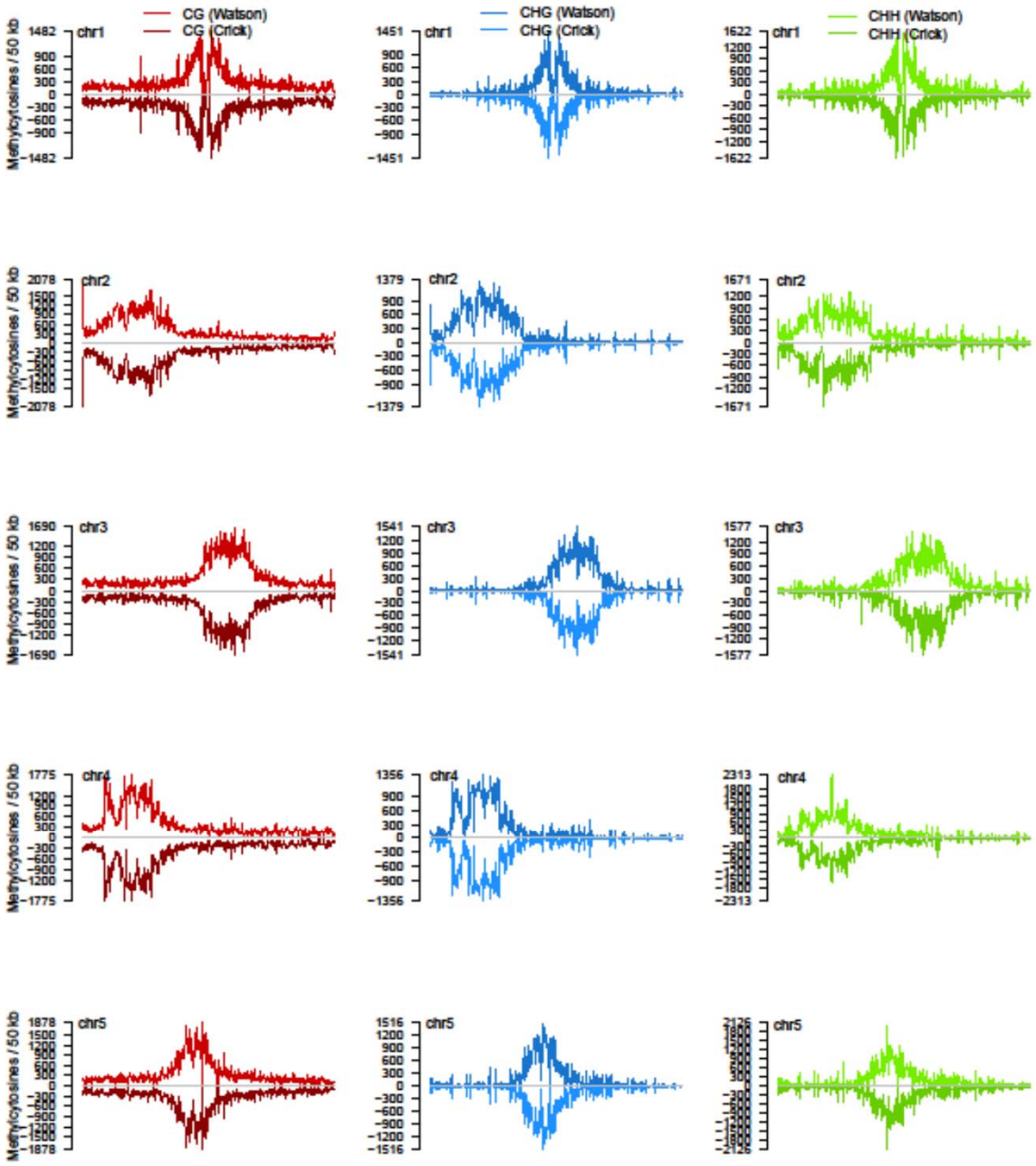


fig. S6. Density of DNA methylation in wild type (Col-0), *idm1-1*, *ros1-4* and *rdd* mutant plants. The densities of methylcytosines of each sequence context across each chromosome in 50 kb segments are presented.

idm1-1

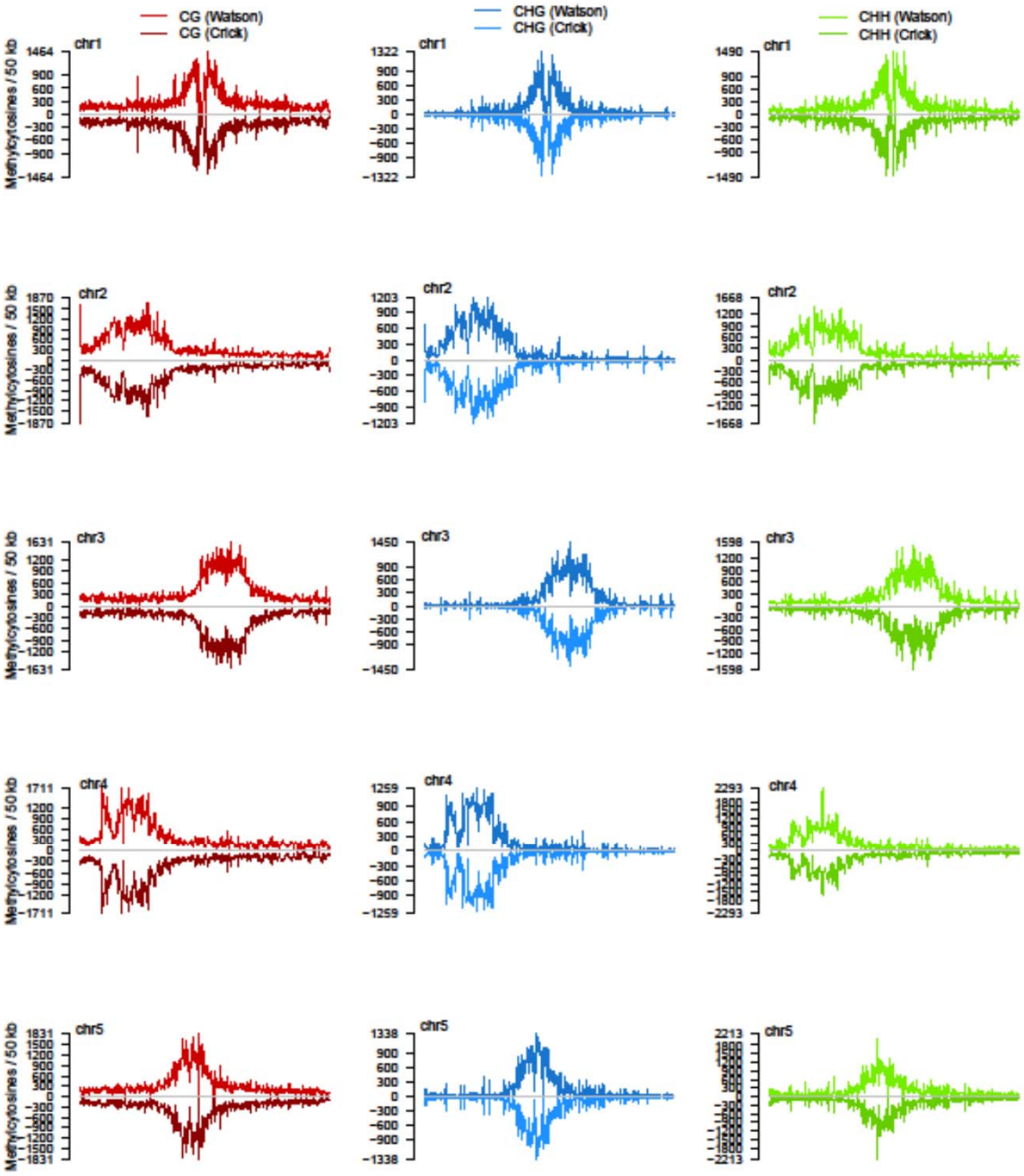


fig.S6 continued

ros1-4

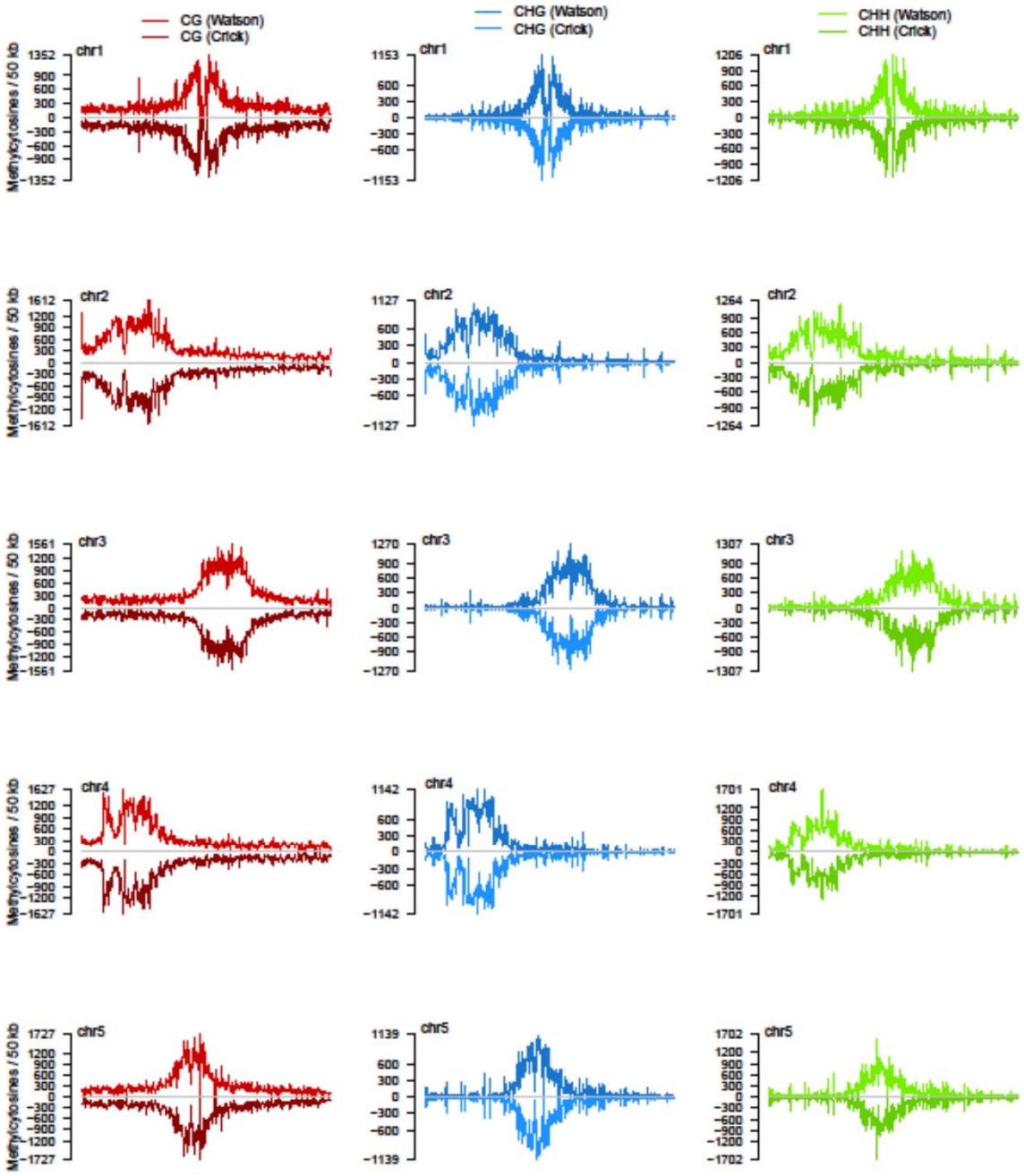


fig.S6 continued

rdd

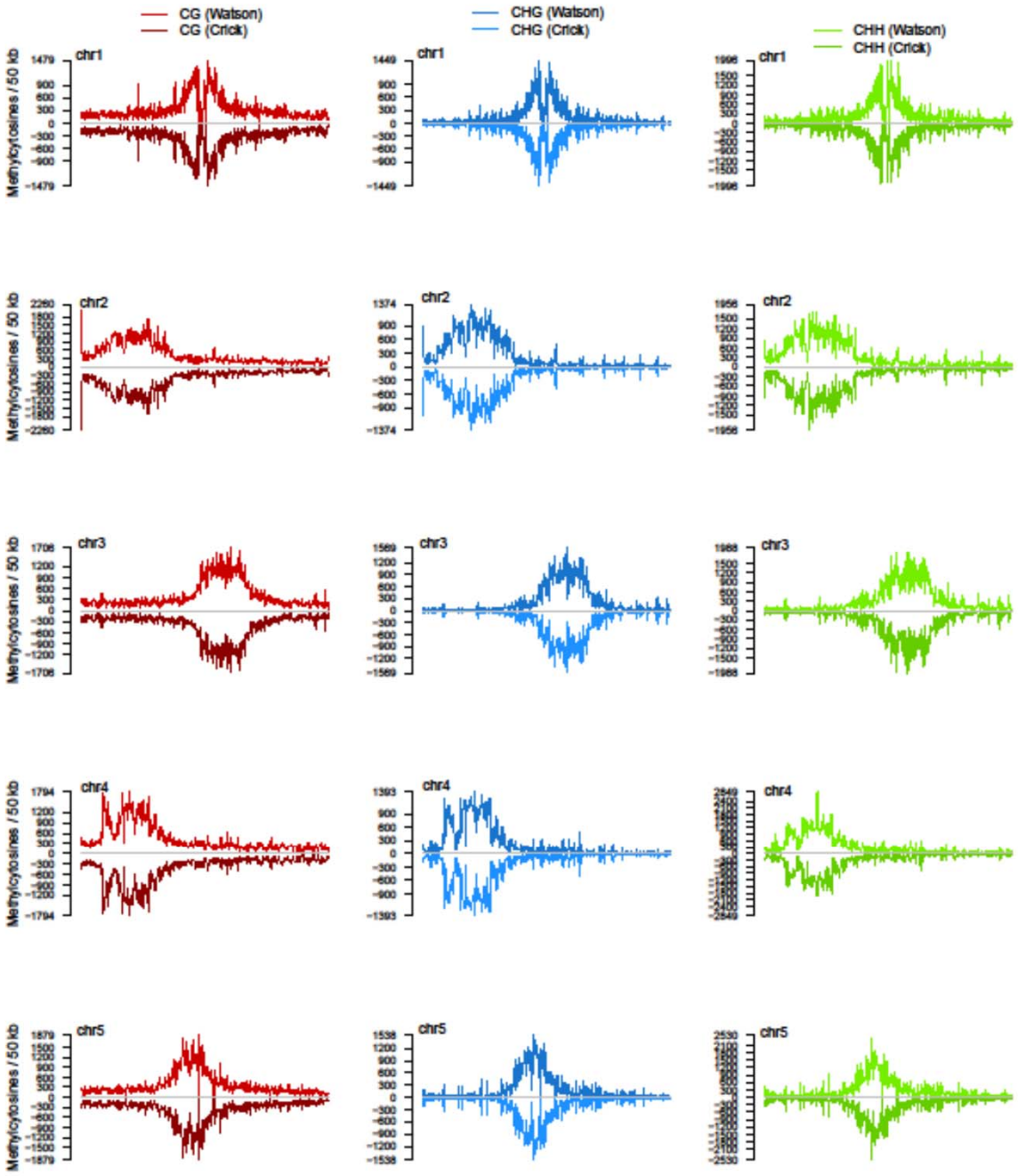
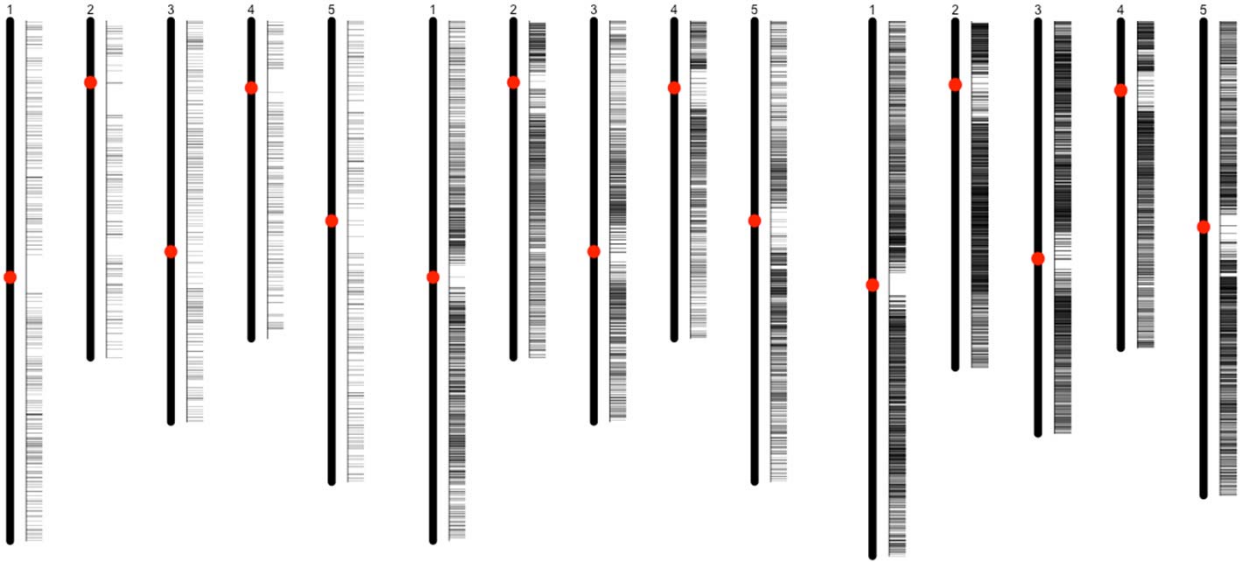


fig.S6 continued

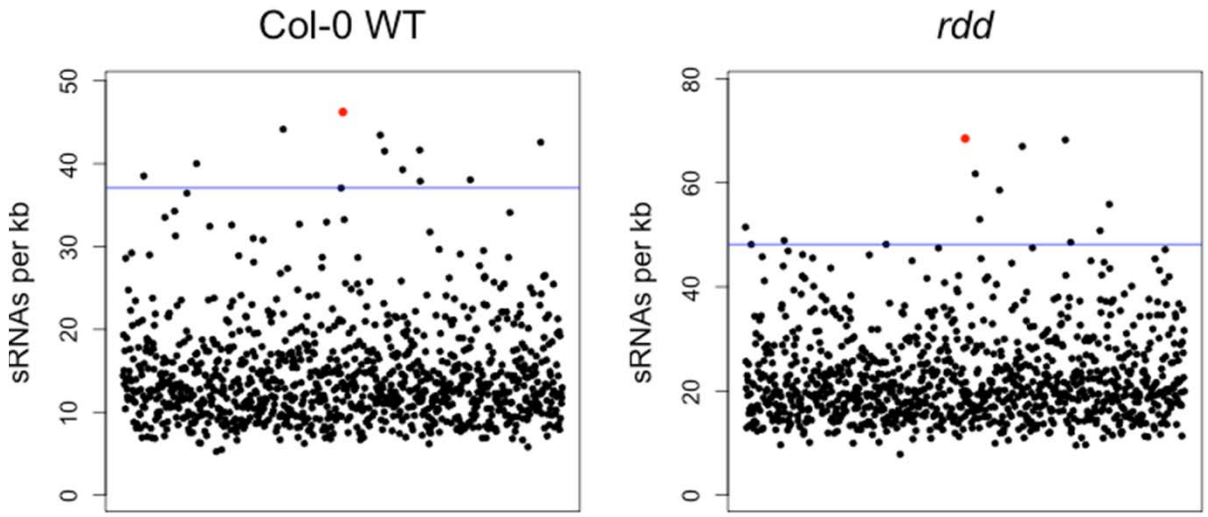
A



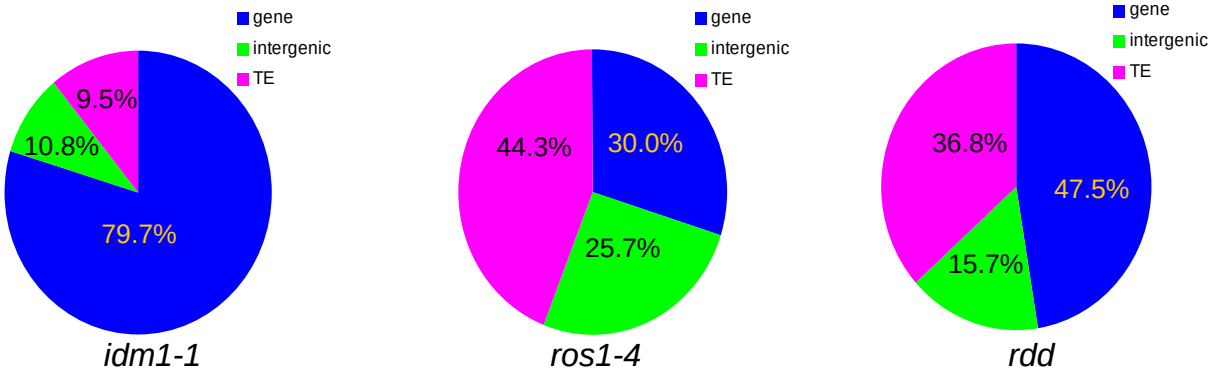
idm1-1 *ros1-4* *rdd*

fig. S7. DNA methylome analysis by whole genome bisulfite sequencing in wild-type, *idm1-1*, *ros1-4* and *rdd* mutant plants. (**A**) Distribution of hypermethylated loci on the five chromosomes in *idm1-1*, *ros1-4* and *rdd* mutants. (**B**) Enrichment of sRNAs in the 1098 hypermethylated regions identified in *idm1-1*. Small RNA reads that were generated from the wild type and *rdd* mutant plants by Lister et al (*9*). Only sRNAs that have perfect match to a unique location in the genome were used in the calculations. Black dots denoted the sRNA densities from randomly selected regions in the genome and red dots denoted the sRNA densities in the hypermethylated regions in *idm1-1*. Horizontal lines indicated the 99 percentile of sRNA density from 1000 simulated runs. (**C**) Composition of the hypermethylated loci in *idm1-1*, *ros1-4* and *rdd* mutants. The total numbers of hypermethylated loci are indicated in the parenthesis. (**D**) Numbers of hypermethylated regions that are overlapping among or unique to the *idm-1*, *ros1-4*, and *rdd* mutants. Hypermethylated regions in each mutant were identified by comparing single-base resolution cytosine methylation profile in the mutant to that in the wild type Col-0. In a few cases where one region in one mutant overlaps with two regions in another mutant, we calculated number of overlapping regions from the perspective of *idm-1*, *ros1-4*, and *rdd*, in this order.

B



C



D

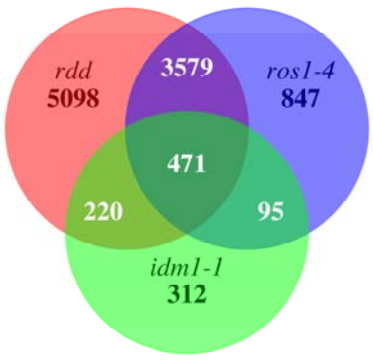
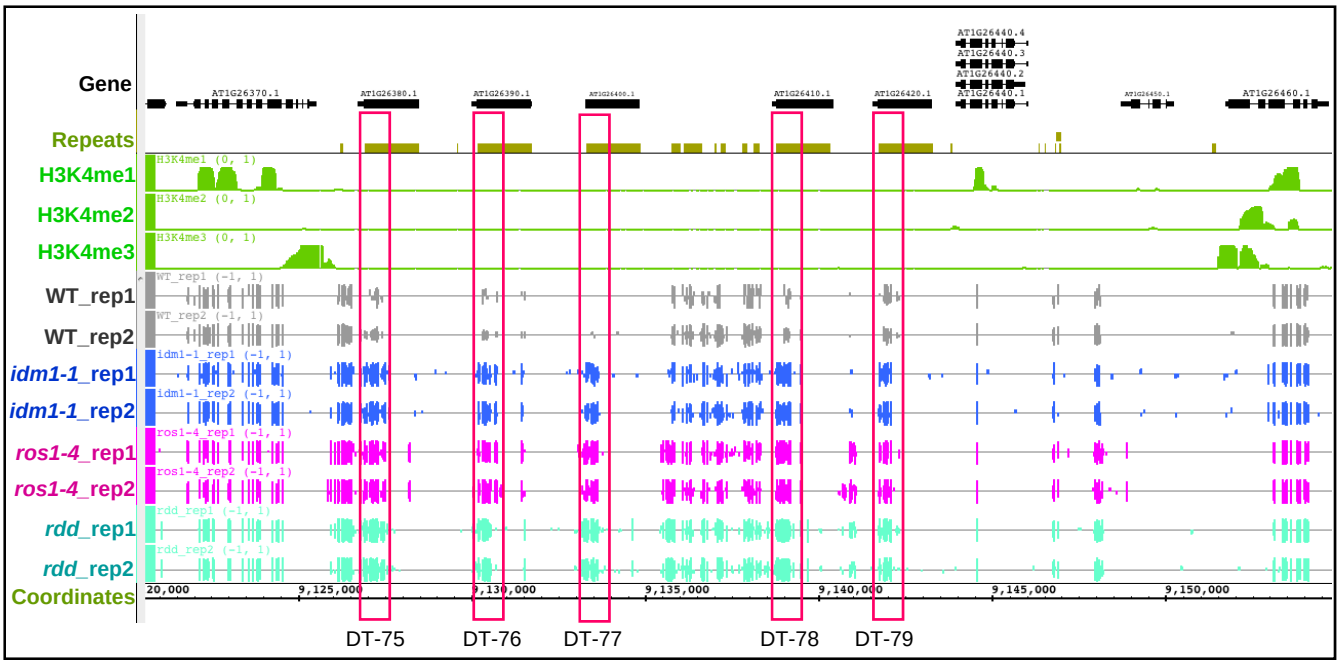


fig.S7 continued

A



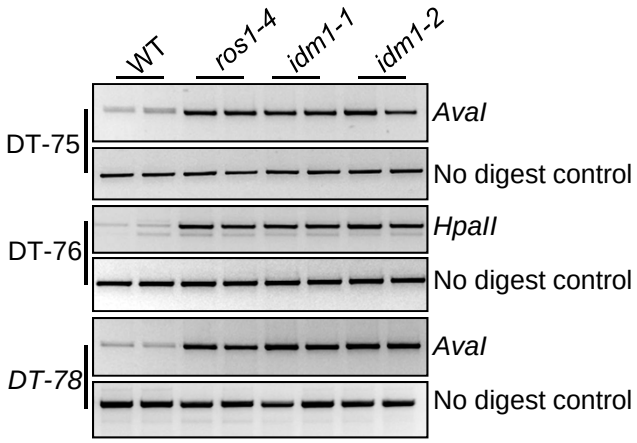
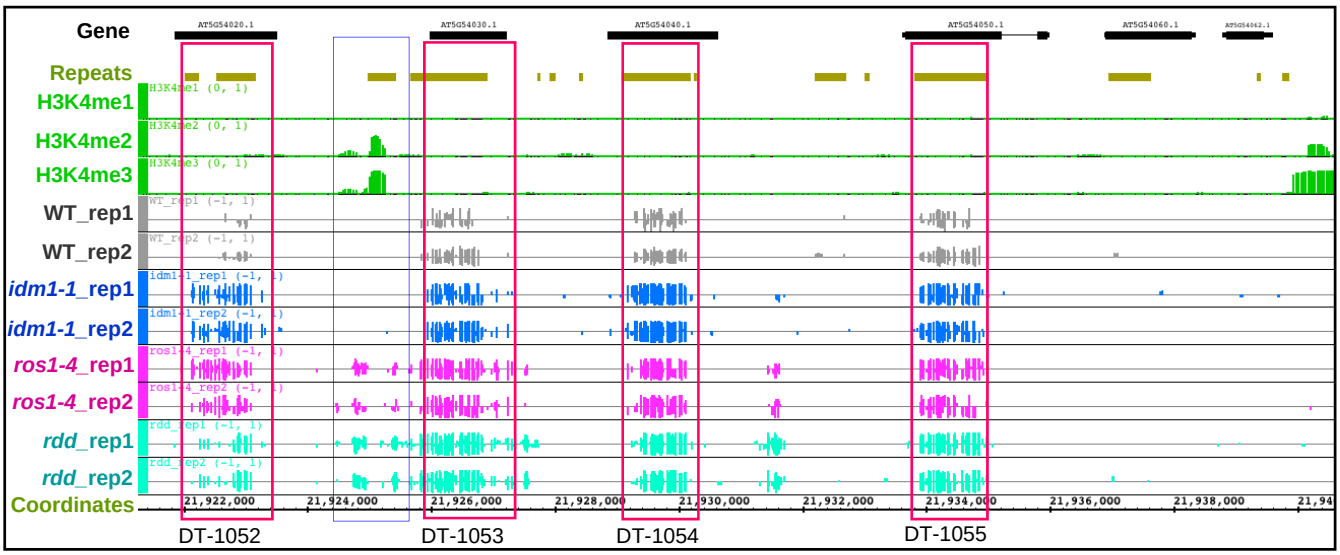
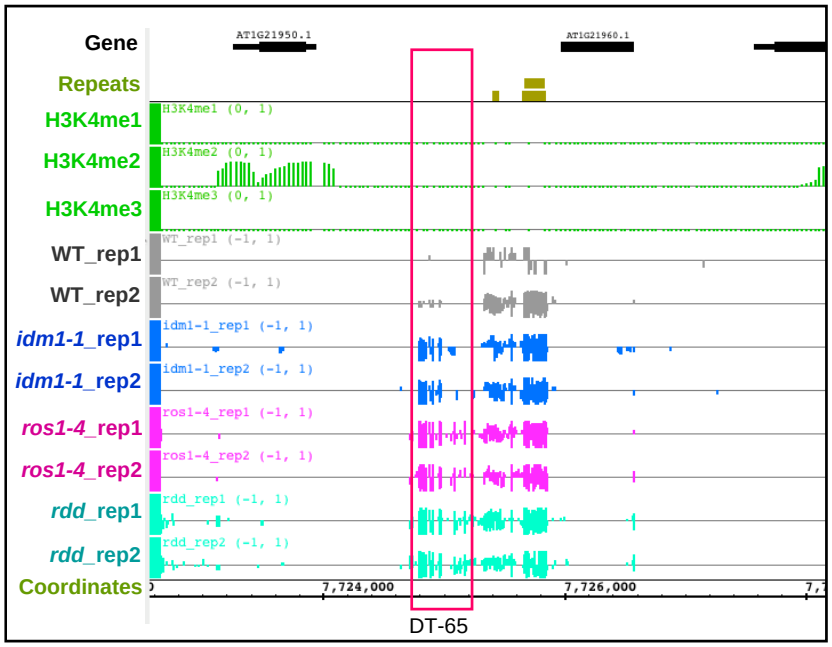


fig. S8. Examples of whole genome bisulfite sequencing data showing DNA hypermethylation and other features in *idm1-1* and other mutant plants and the effects of the mutations on expression of the hypermethylated genes or nearby genes. (**A**) Upper panel: Five members of an FAD-binding berberine gene family are hypermethylated in *idm1-1*, *ros1-4* and *rdd* mutant plants. Hypermethylated regions are highlighted with red boxes. The marker locus, *At1g26400*, was also identified as one of the hypermethylated loci in the mutants by the whole genome bisulfite sequencing (*DT-77*). Repeats were annotated according to RepeatMasker. H3K4 methylation data was downloaded from Zhang et al (*10*). Lower panel: Confirmation of the whole genome bisulfite sequencing results by CHOP PCR. (**B**) Four members of the cysteine/histidine-rich C1 domain gene family are hypermethylated in *idm1-1*, *ros1-4* and *rdd* mutant plants. Hypermethylated regions are highlighted with red boxes, where all *idm1*-dependent hypermethylation sites show a lack of H3K4 methylation; in contrast, a *ros1*-dependent but *idm1*-independent hypermethylated region (between *DT-1052* and *DT-1053*) does have high levels of H3K4me2 and H3K4me3, which is highlighted with a blue box. (**C-H**) Examples of DT loci and effects on the expression of the hypermethylated genes or nearby genes in 2-week-old *idm1-1* and *ros1-4* seedlings.

B



C



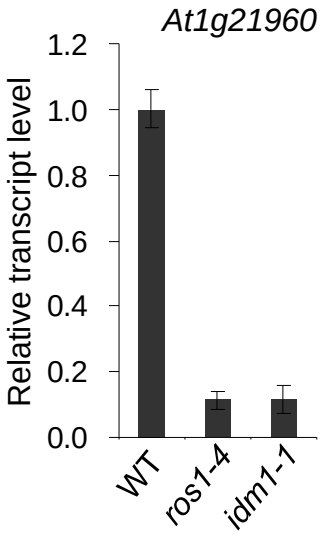
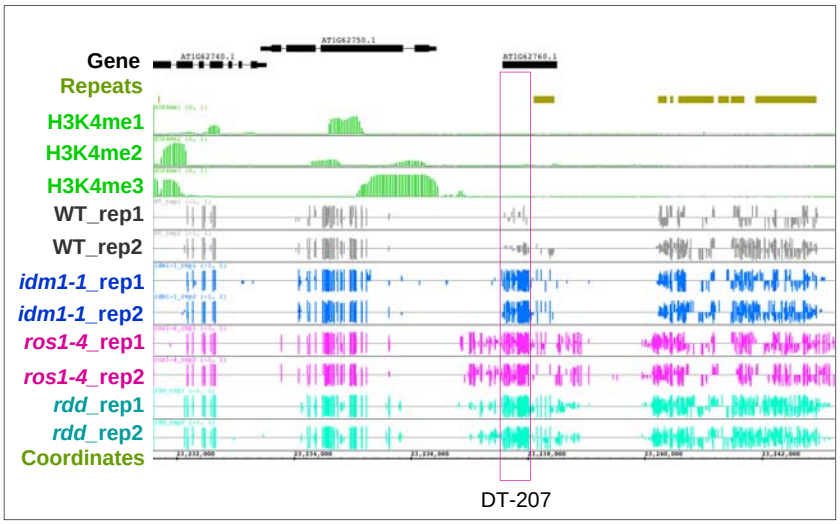
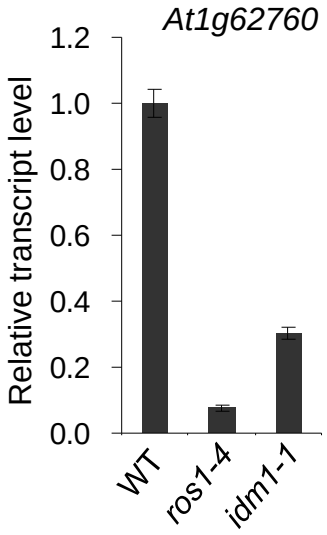


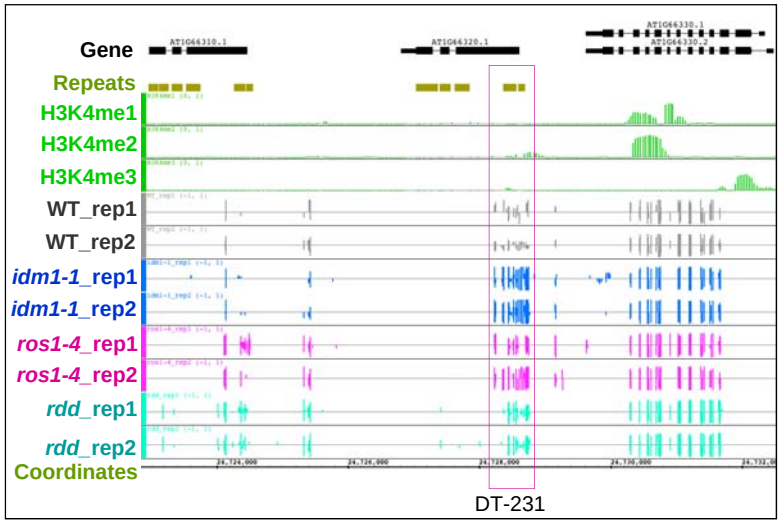
fig.S8 continued

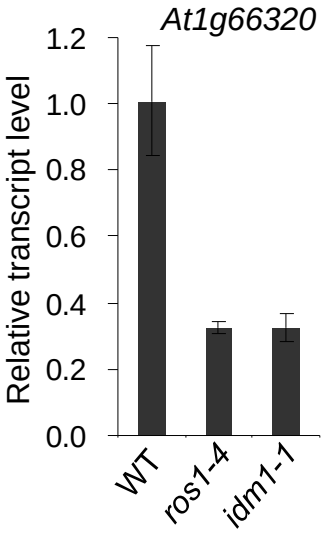
D



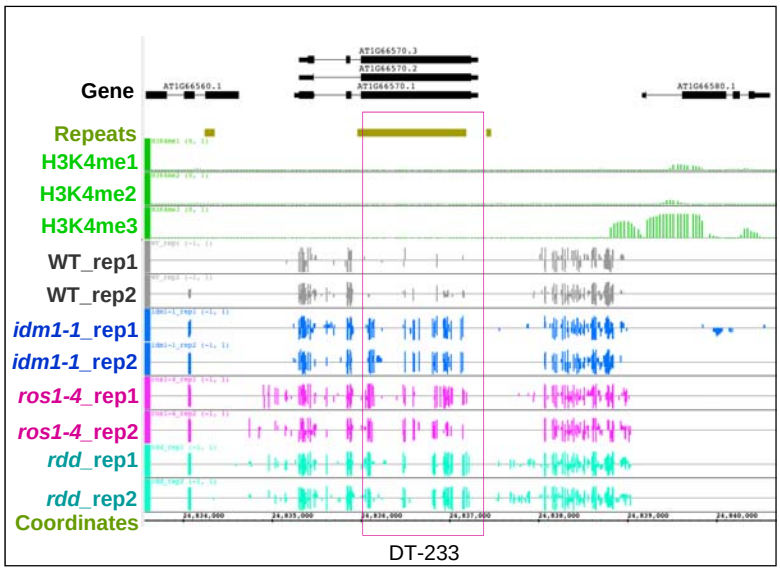


E





F



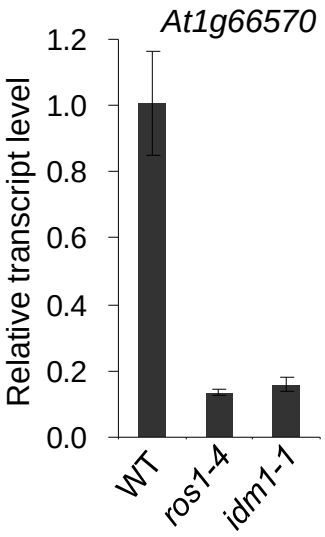
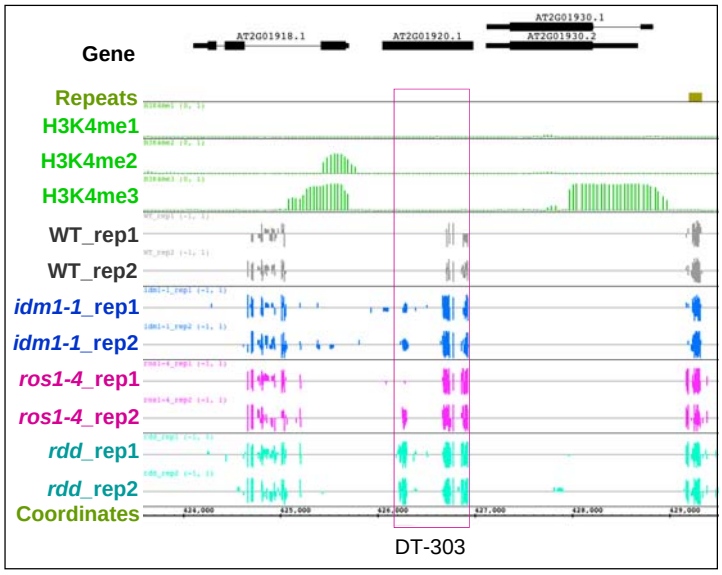
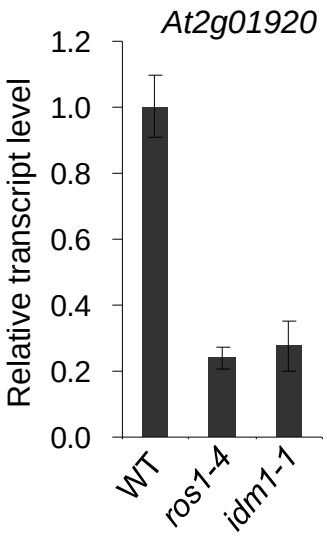


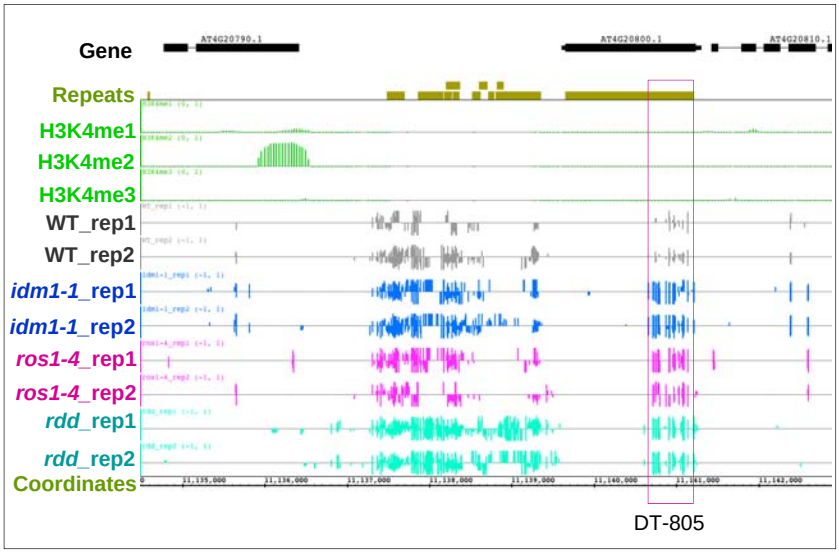
fig.S8 continued

G





H



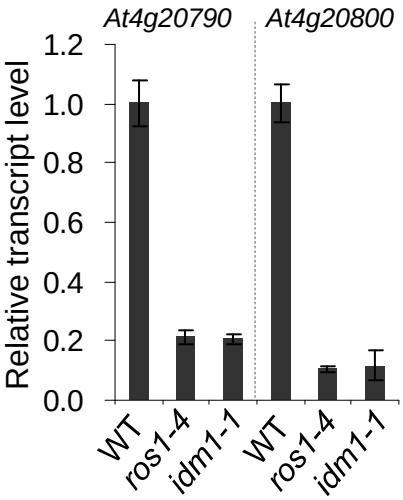


fig.S8 continued

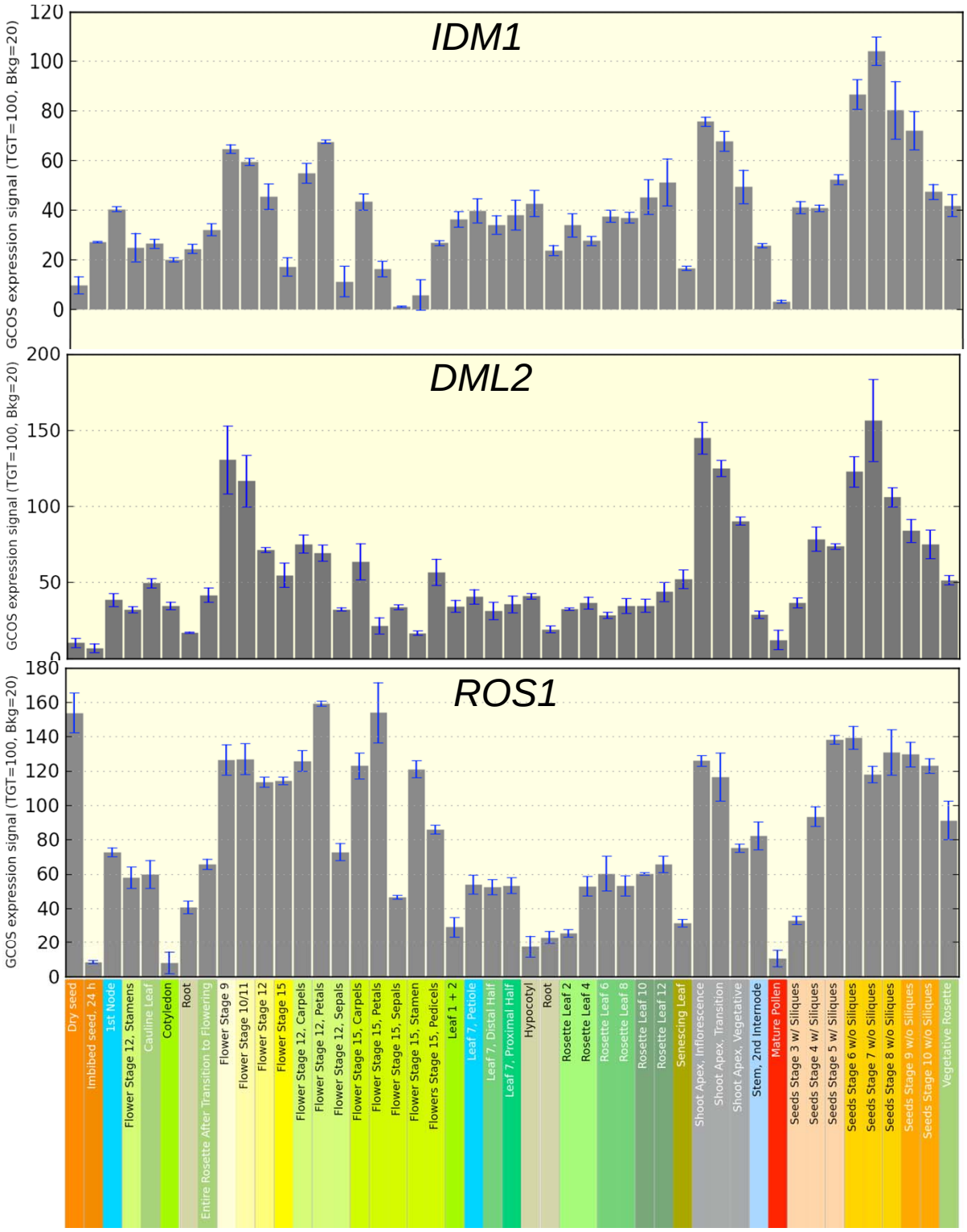
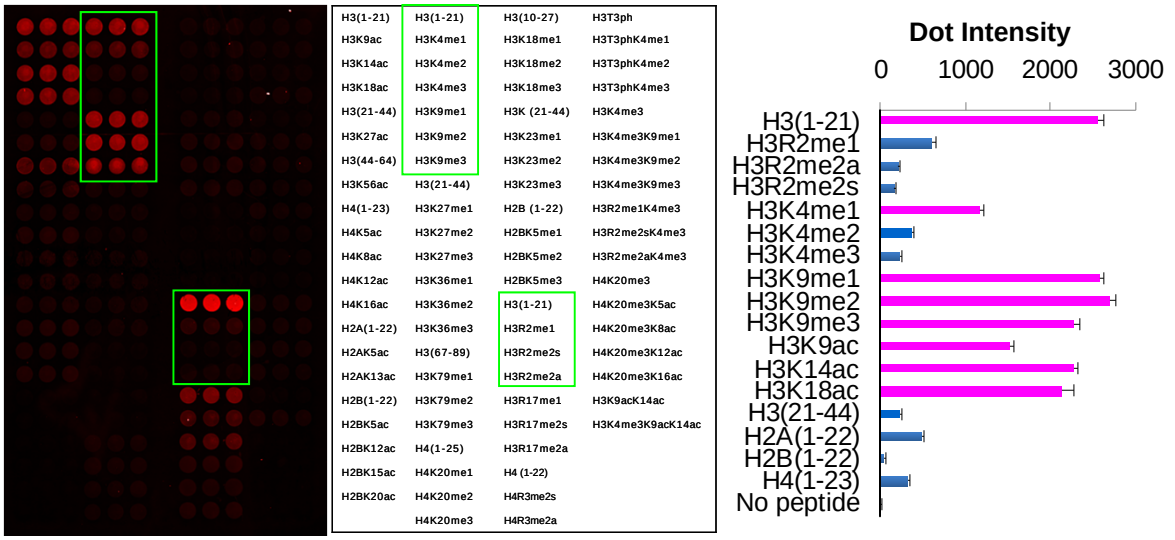


fig. S9. Expression patterns for *IDM1*, *ROS1* and *DML2*. Gene expression data in different tissues or at different developmental stages were downloaded from the TAIR website (http://www.arabidopsis.org/servlets/Search?action=new_search&type=expression).

A



B

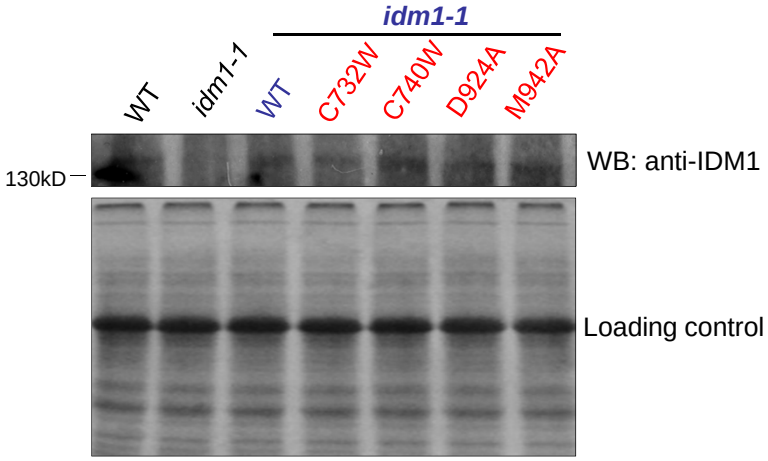
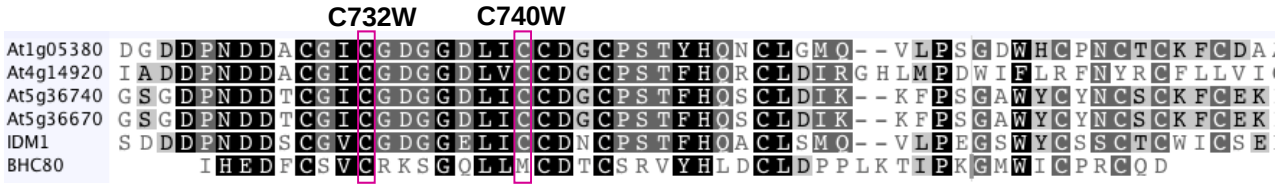
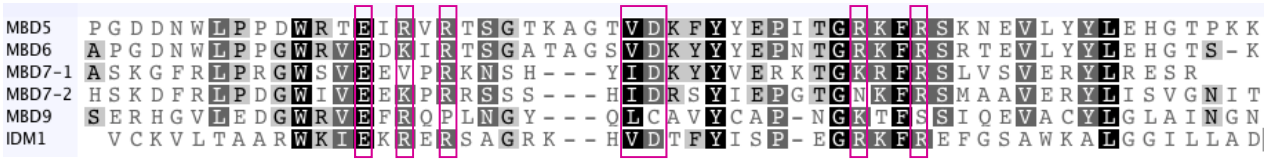


fig.S10. Functional characterization of the PHD finger domain of IDM1 and detection of IDM1 protein from transgenic plants. (**A**) Test of histone binding specificity of the PHD finger of IDM1 using a histone peptide array. Samples of interests are highlighted with green box and the array data was quantified according to the dot intensity. Error bars represent standard error (n=3). Note that the binding to H3 peptide amino acids 44–64 is nonspecific, as was previously shown (*16*). Red spots indicate positive binding. H2A, histone H2A; H2B, histone H2B; H3, histone H3; H4, histone H4. me, methylation; ac, acetylation; ph, phosphorylation. (**B**) Detection of wild type IDM1 protein and site-directed mutant IDM1 proteins in transgenic Arabidopsis plants. IDM1 proteins were detected by Western blotting of samples immunoprecipitated by anti-IDM1 antibody from extracts from the transgenic plants. Coomassie-stained SDS-PAGE gel of 25 microgram aliquots of the plant extracts used for immunoprecipitation is shown as loading control.

A



B



C

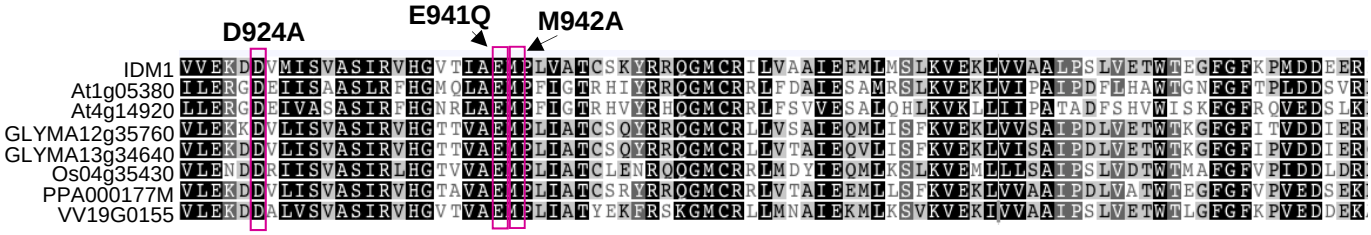
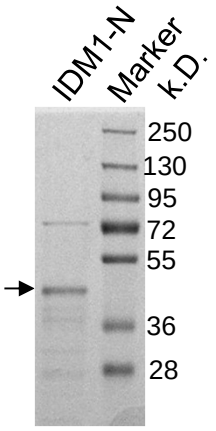


Fig. S11. Amino acid sequence alignments for the PHD, MBD, and HAT domains of IDM1. (**A**) Multiple sequence alignment for PHD domains from Arabidopsis and human proteins. The PHD finger of BHC80 binds unmethylated H3K4, and M502 (corresponding to C740 of IDM1) is important for the binding. C732 is conserved in all of the PHD domains. Red boxes indicate the amino acid residues mutated in Fig. 2A and 2B. Conserved residues are highlighted in dark or gray. Amino acid sequences shown for At1g05380, At4g14920, At5g36740, At5g36670, IDM1, and BHC80 are 619–674, 646–703, 644–699, 644–699, 720–776, and 486–534, respectively. (**B**) Multiple sequence alignment for MBD domains of Arabidopsis proteins. Conserved amino acid residues for DNA binding are marked by red boxes. Amino acid sequences shown for MBD5, MBD6, MBD7-1, MBD7-2, MBD9, and IDM1 are 29–85, 75–130, 110–160, 176–229, 262–314, and 247–299, respectively. (**C**) Multiple sequence alignment for HAT domains of plant proteins. D924, E941 and M942 (Fig. 3C and 3D), which are marked by red boxes, are conserved in all listed HAT domains. Sequences for IDM1, At1g05380, and At4g14920 are from Arabidopsis; sequences for GLYMA12g35760 and GLYMA13g34640 are from soybean; Os04g35430 is from rice; PPA000177M is from peach; and VV19G0155 is from grape. Amino acid sequences shown for IDM1, At1g05380, At4g14920, GLYMA12g35760, GLYMA13g34640, Os04g35430, PPA000177M, and VV19G0155 are 919–1002, 821–904, 851–934, 941–1024, 835– 918, 1239–1322, 1095–1178, and 906–989, respectively.

A



B

fig. S12. Purification of the N-terminal MBD domain-containing fragment of IDM1 (1–400 aa) and methylated oligonucleotides used for EMSA. (**A**) Purification of the N-terminal fragment of IDM1 (1–400 aa), which includes the putative MBD domain. Arrow indicates the IDM1-N band. (**B**) Oligonucleotide probes used in the methylcytosine-binding assays shown in Fig. 2C and 2D. Asterisks indicate the positions of 5mC residues.

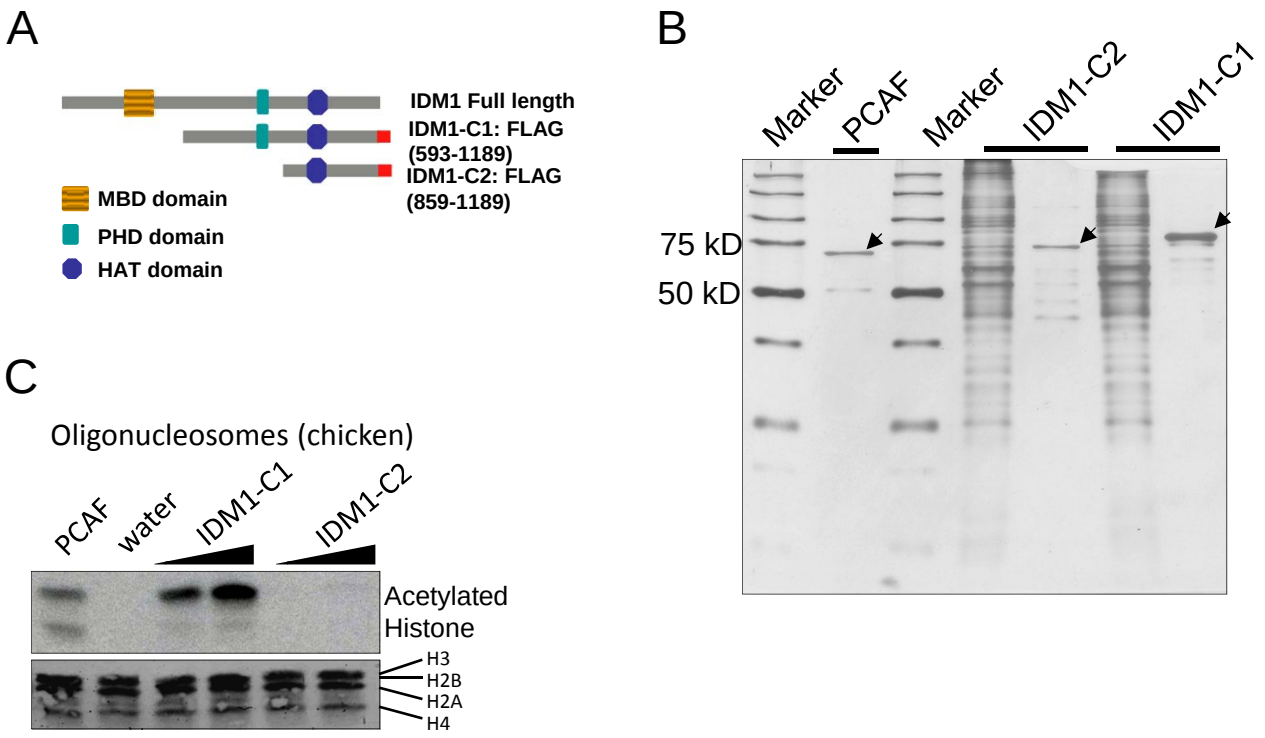
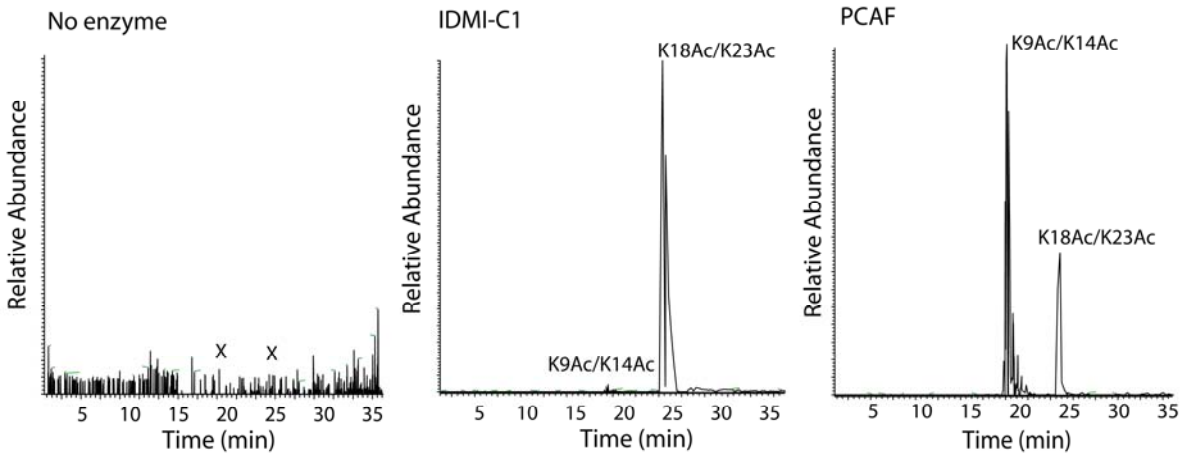
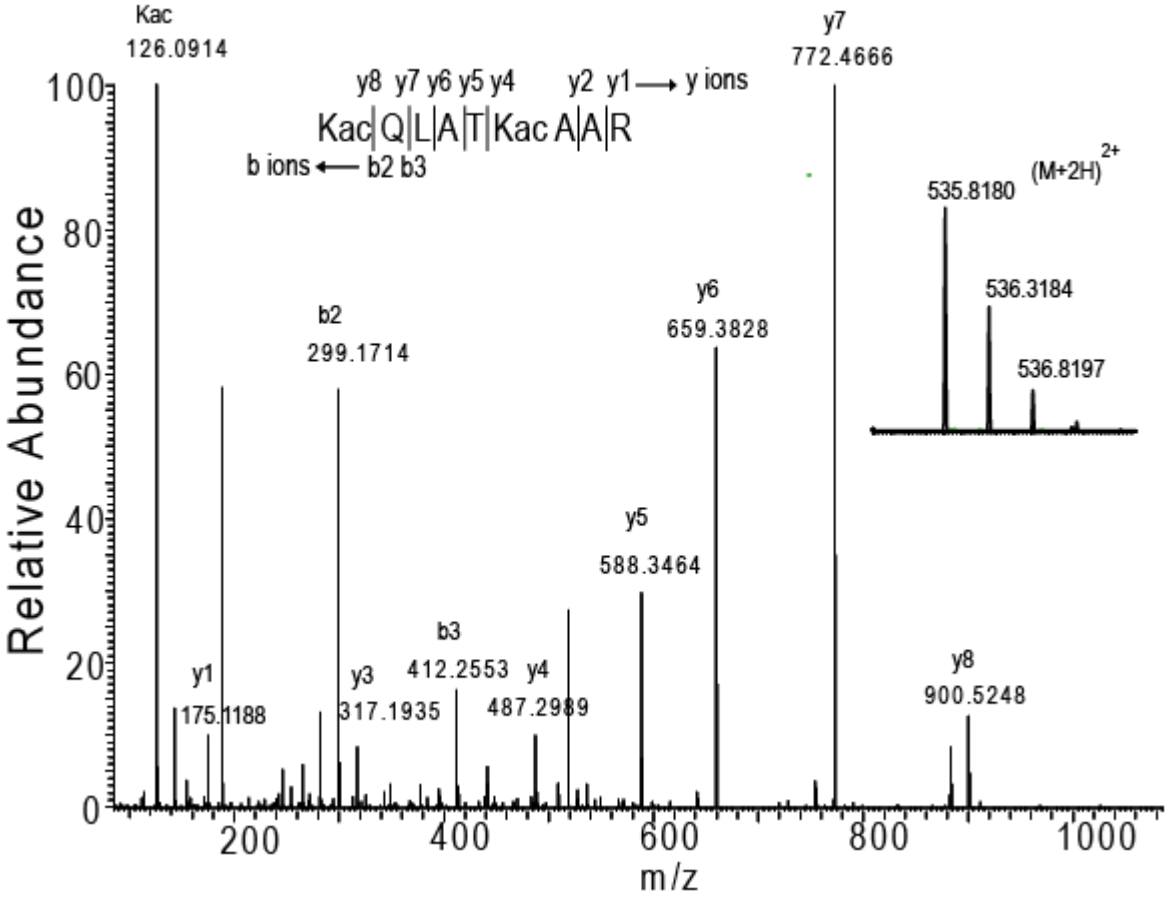


fig. S13. Purification and characterization of the HAT domain fragment of IDM1. (**A**) Schematic of the IDM1 protein and derivatives for expression and purification from insect cells and for histone acetyltransferase activity assays. (**B**) Purification of IDM1 C-terminal fragments from insect cells. Arrows point to the PCAF control protein or IDM1 C-terminal fragments. (**C**) HAT activity assay using chicken oligonucleosomes as substrate. Upper: autoradiograph for acetylated histones; Lower: Coomassie blue-stained histone proteins in the substrate. (**D**) Selective display (relative ion abundance versus retention time of peptides on a HPLC column) of ions at *m/z* 493.2749 ± 0.1000 corresponding to H3 K9/K14 diacetylated peptide KAcSTGGKAcAPR and at *m/z* 535.8195 ± 0.1000 corresponding to H3 K18/K23 diacetylated peptide KAcQLATKAcVAR. PCAF was used as a positive control enzyme. (**E**) MS/MS spectrum of a precursor ion at *m/z* 535.8180 ($[M+2H]^{2+}$) from (**D**). This peptide was determined to span residues 18-26 and was found to contain two acetylation (ac) modifications on K18 and K23. b and y type ions are labeled correspondingly to the sequence shown on the top of the figure. Inset: mass spectrum of the precursor ion with the experimental mass at *m/z* 535.8180 that is 2.8 ppm deviated from the calculated mass (*m/z* 535.8195) of the above-defined acetylated peptide.

D



E



figS13 continued

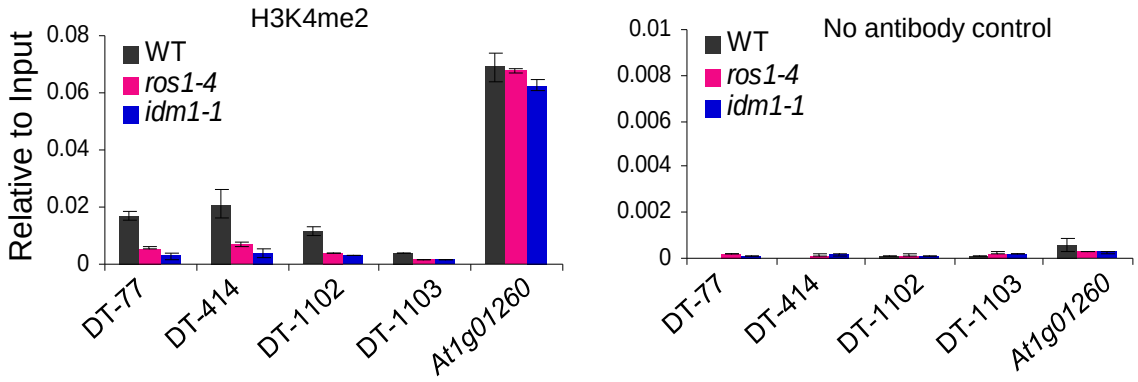


fig. S14. H3K4me2 levels at some IDM1 target regions and control region. ChIP was performed in wild-type, *ros1-4*, and *idm1-1* plants with anti-H3K4me2 antibody. The ChIP signal was quantified as relative to input DNA. The no-antibody precipitates served as negative controls. Two biological replicates were performed, and very similar results were obtained. Standard errors were calculated from three technical repeats.

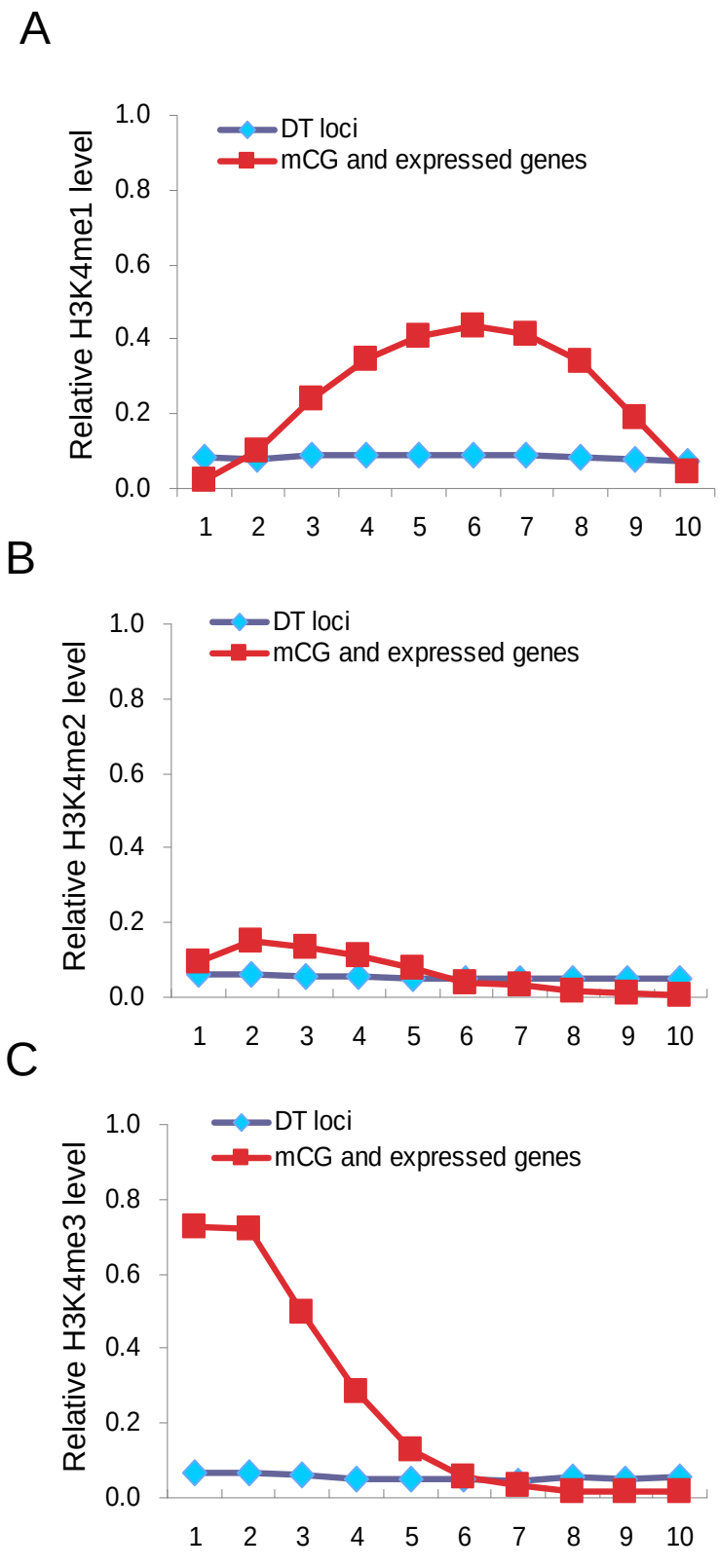


fig. S15. Analysis of H3K4 methylation levels at DT loci and expressed genes with high CG methylation. See Materials and Methods for details.

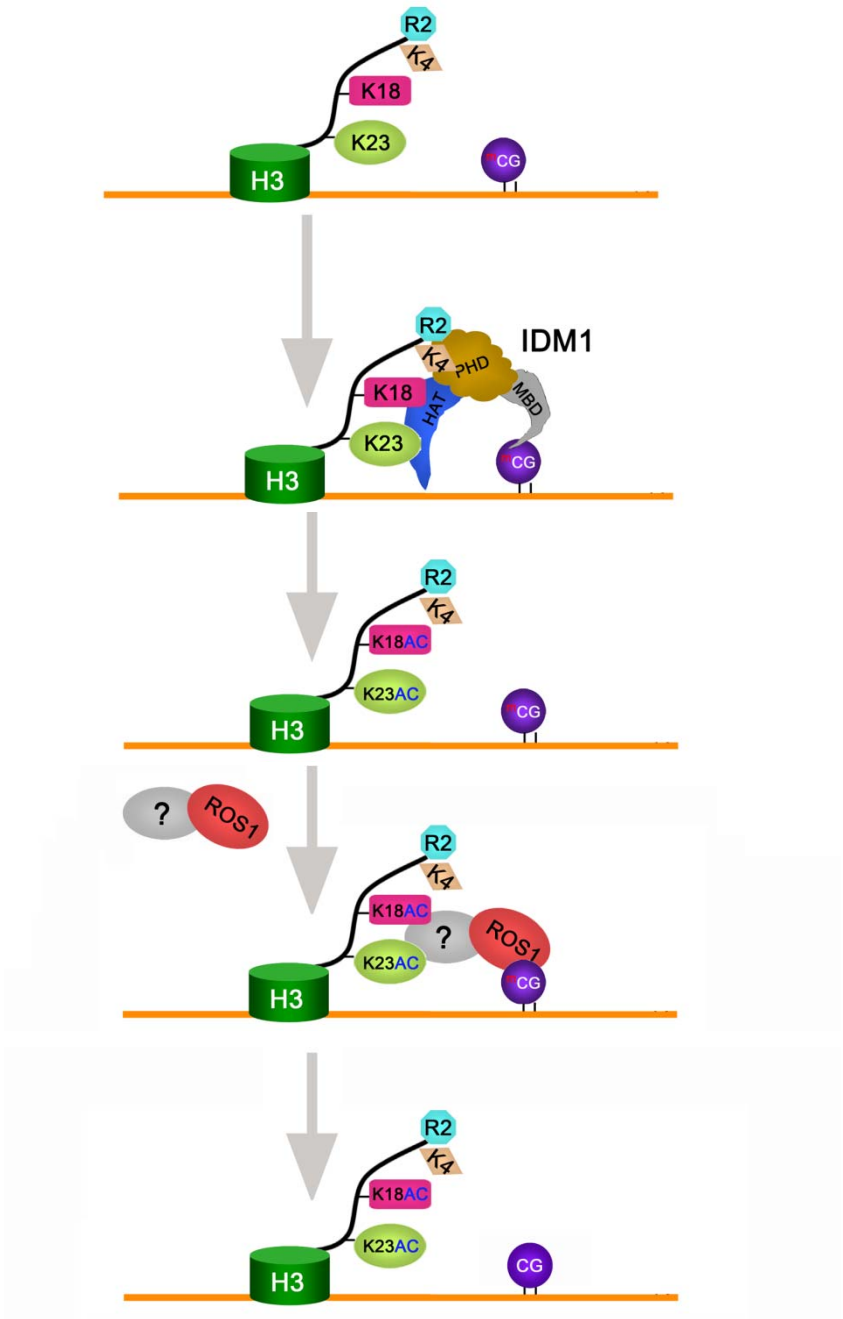


fig. S16. Working model for IDM1 function in DNA demethylation. A subset of active DNA demethylation target loci are characterized by CG methylation as well as a lack of histone H3K4 or H3R2 methylation. IDM1 recognizes these epigenetic features through its MBD (methyl-DNA binding domain) and PHD (plant homeodomain) finger domains. IDM1 then acetylates histone H3 at K18 and K23 through its HAT (histone acetyltransferase) domain. An as yet unidentified ROS1-interacting protein (indicated by question mark) may recognize the acetylated H3K18 and H3K23 marks and help direct ROS1 (other members of the ROS1 subfamily of 5-methycytosine DNA glycosylases) to the loci for active DNA demethylation.

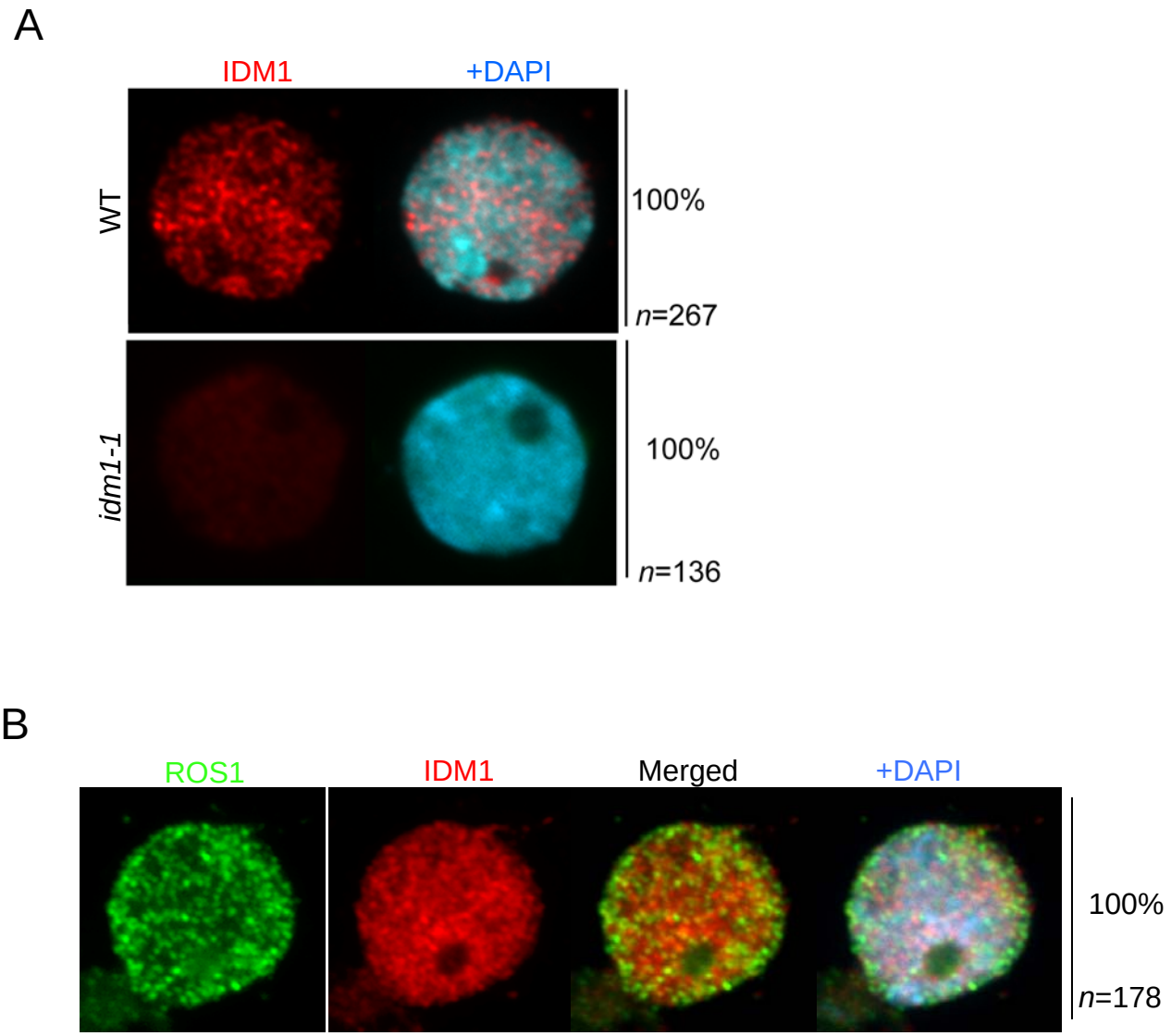


fig. S17. Sub-nuclear localization of IDM1 and its co-localization with ROS1. (**A**) Detection of IDM1 (red) in WT and *idm1-1* mutant nuclei by immunostaining using anti-IDM1. (**B**) Dual immunolocalization of ROS1 (green) and IDM1 (red). DNA was stained with DAPI (blue). The frequency of nuclei displaying each interphase pattern is shown on the right.

Table S7. List of primers and probes.

Primer Name	Sequence	Purpose
At1g26400/ DT-77-F1 *At1g26400/ DT-77*-R1	TGACCTGCATAGGCTATAACACA ATTGGAATCAATCCGAGTGG	CHOP PCR
At4g18650-F1 *At4g18650*-R1	TCCACATGAGCCATCAACTC AAGGGTACCAGTTTGGGAAAA	CHOP PCR
LUC-F *LUC*-R	AAGCGAAGGTTGTGGATCTGGA TCAATCAAGGCGTTGGTCGCTT	Real-time PCR
NPTII-F *NPTII*-R	ACCTTGCTCCTGCCGAGAAAGTAT ATGCGATCTTTCGCTTGGTGGT	Real-time PCR
ROS1-F *ROS1*-R	AAGGACCAACTTGTTGCGAC AGGACTCTATTAGCACTGAGC	Real-time PCR
ROS3-F *ROS3*-R	CCTTCACGACTAAGCTTTCC TCTATGTGCGATGACACAGC	Real-time PCR
TUB8-F *TUB8*-R	ATAACCGTTTCAAATTCTCTCTCTC TGCAAATCGTTCTCTCCTTG	Rea-ltime PCR
At1g21960-F *At1g21960*-R	ACCTCGGATTGACTCGTTTG TCGCGTCAGAGTTGGATTTC	Real-time PCR
At1g62760-F *At1g62760*-R	CTCAATCTCACCCTCTCATCC AGTCTCCAATCTCTTCAACGC	Rea-ltime PCR
At1g66320-F *At1g66320*-R	GCCCGAATGTGAAATCTCTTG TCATCTCCAACTTGTCCCATG	Real-time PCR
At1g66570-F *At1g66570*-R	GATCGCATGGTTCCCTTTTC CATCAATCCCAATGCACCAAC	Real-time PCR
At2g01920-F *At2g01920*-R	CATACAGAGCAACCCTTCCAG TCAAACATTTCAAAGCCACCG	Real-time PCR
At4g20790-F *At4g20790*-R	CTCTCAGTATTCTCCGCAACC AGCATTGGATCACGAAGACAG	Real-time PCR
At4g20800-F *At4g20800*-R	GAAAGTACGGAATCTCGGTCG ACGAGAGGATAACGCCAAAG	Real-time PCR
IDM1-F1 *IDM1*-R1	TTTACATATCGCCAGAGGGACG GGAAGCTGGTTTGATATCGGAT	Genotyping
IDM1-F2 *IDM1*-R2	TTGGGTCAGCTTGGAAAGCATTGG AAGGCTCTAAGGCACTCCACCAAA	Real-time PCR
IDM1-F3 IDM1-R3	CTGGATCCGATGATGATCCAAATGATG CTGTCGACTTACAACTCGCTGCATATCCA	expression PHD domain with GST tag
IDM1-C732W-F IDM1-C732W-R	GATGATTCTTGTGGAGTTTGGGGTGATGGGGGTGAGCTGATCTGC GCAGATCAGCTCACCCCCATCACCCCAAACTCCACAAGAATCATC	site direct mutagenesis
IDM1-C740W-F IDM1-C740W-R	GTGATGGGGGTGAGCTGATCTGGTGTGATAATTGTCCATCTACC GGTAGATGGACAATTATCACACCAGATCAGCTCACCCCCATCAC	site direct mutagenesis
IDM1-D924A-F IDM1-D924A-R	CAGTGGTTGTGGAAAAGGATGCTGTGATGATCTCAGTAGCATCTATC GATAGATGCTACTGAGATCATCACAGCATCCTTTTCCACAACCACTG	site direct mutagenesis
IDM1-E941Q-F IDM1-E941Q-R	TGCACGGAGTAACTATTGCACAGATGCCTCTTGT ACAAGAGGCATCTGTGCAATAGTTACTCCGTGCA	site direct mutagenesis
IDM1-M942A-F IDM1-M942A-R	GGAGTAACTATTGCAGAGGCGCCTCTTGTAGCTACGTGCAGCA TGCTGCACGTAGCTACAAGAGGCGCCTCTGCAATAGTTACTCC	site direct mutagenesis
H3-K14R-F H3-K14R-R	CTCGCAAGTCCACCGGCGGCCGGGCCCCGCGCAAGCAGCTGG CCAGCTGCTTGCGCGGGGCCCGGCCGCCGGTGGACTTGCGAG	site direct mutagenesis
H3-K18-F H3-K18-R	GGCGGCAAGGCCCCGCGCCGGCAGCTGGCCACCAAGGCT AGCCTTGGTGGCCAGCTGCCGGCGCGGGGCCTTGCCGCC	site direct mutagenesis
H3-K23-F H3-K23-R	CAAGCAGCTGGCCACCCGGGCTGCCCGCAAGAGCG CGCTCTTGCGGGCAGCCCGGGTGGCCAGCTGCTTG	site direct mutagenesis
5S rDNA-F *5S rDNA*-R,	GGATGCGATCATACCAG GAGGGATGCAMCACSAG	Southern probe amplification

45S rDNA-F *45S rDNA*-R	AATTCCTAGTAAGCGCGAGTC CTGCTTCACAATGATAGGAAG	Southern probe amplification
180 bp-F 180 bp-R	ACCATCAAAGCCTTGAGAAGCA CCGTATGAGTCTTTGTCTTTGTATCTTC	Southern probe amplification
DT-1102-F1 DT-1102-R1	TGGATCTTCCTCCTAAAAGCA CCATCCTCATTGCACAAAAA	ChIP PCR
DT-77-F1 DT-77-R1	CCTTTGCGTCCAACCATATC GATGGATCGGATTCCAGCTA	ChIP PCR
DT-65-F1 DT-65-R1	TGACTAGCTTTATGGACAAAATTGC TCGACTACCCTAAAACTTCAACG	ChIP PCR
DT-414-F1 DT-414-R1	GGCATCGAAAAAGGGAATTT TGAACATTGTCGGAGCAAAA	ChIP PCR
At4g18650-F3 *At4g18650*-R3	AAACGTCACATATTTGCAGTCG GACATGACAGATGGAAGAGAAGG	ChIP PCR
DT-214-F DT-214-R	TATGGGAGGAGGAGGCCTAT TTTCAATGACTTTGCGCTTG	ChIP PCR
DT-221-F DT-221-R	CGGGTTTTTATTTCGGGTTC GGACGTCAATCACACCAACA	ChIP PCR
DT-265-F DT-265-R	TGAATTTTCCCTGGAGTTGC CAAAGCTTCAGCTGGGAAGT	ChIP PCR
DT-271-F DT-271-R	GCTATGGCTGGGACTACGAC tgtacagGGAAGACGTGGTG	ChIP PCR
DT-323-F DT-323-R	TGGGGAAGCAGAATAACACC CCATTCTCCCTGGTTTAGGG	ChIP PCR
DT-678-F DT-678-R	GCGGTGTTCATAGGGTCGTA CTCCGACAACATCTCCGACT	ChIP PCR
DT-713-F DT-713-R	TGGCCATGAGGTGAATAACA GGACTGTCCGTTTTTGTTGG	ChIP PCR
DT-763-F DT-763-R	CGTTTATTGGACCACGTGAA TGTACGATGTCGCATGTTGA	ChIP PCR
DT-912-F	CTCGTCTGTCACCCAAGACA	ChIP PCR
DT-912-R	TGGGAACGTCCTCTCAGTTC	
At1g01260-F *At1g01260*-R	TCCGAATAATGGGAACAGGT GGCCTCGTGTAGTTCTCAGG	ChIP PCR
At1g10950-F *At1g10950*-R	GAGGAGCAGGTCACCCTATG ACCACCCCATTTGTGAACAT	ChIP PCR
At1g30440-F *At1g30440*-R	AGCATCGACAGACCCAAACT AAACACTGCCACCTGGACTC	ChIP PCR
DT-75-F DT-75-R	CATGTTGGAAATTTTGTACCCG GAGTATTATGAAGGAGCTCTACGA	CHOP PCR
DT-76-F DT-76-R	GAACCTGAGTGCTGAAGGTTACTC CAAGATTCAATACTTTACGACATG	CHOP PCR
DT-78-F DT-78-R	GTACGTTTACACGTGTATTTTATT GTAGCGTCTCGTTGCAATTCAAC	CHOP PCR
At1g26400-F2 *At1g26400*-R2	GTAGTTTGAGATGATTAATGATAGAGTT AAACTTATTCAATCTTCAATACTCTAC	Bisulfite sequencing
At4g18650-F4 *At4g18650*-R4	GTTGAATGAATGTTTTTATTTGTTATTTTTGG CTTCCCTTAAAATTTTAATACAAAATCCTATC	Bisulfite sequencing
DT-414-F2 DT-414-R2	GTTTAAGTTTGGGTTTTTGAAtTtATGG TTTAATATTCTCCTTAAATTCCCTATATACAAC	Bisulfite sequencing
DT-65-F2 DT-65-R2	GAGTATGATTGGTTTATTGAAGTTTTAGTT TTCCTTTCCTTTATCTTATTTACCTTTATTTC	Bisulfite sequencing
DT-52-F1 DT-52-R1	GGTAGAGAGTTGGTTTGGTATTG CTAATCCTCATCCTCATCCTCATC	Bisulfite sequencing
DT-1102-F2 DT-1102-R2	TGTGTTGTTGTATTTGAGAG CCATCCTCATTACACAAAAAAATAC	Bisulfite sequencing

DT-1100-F DT-1100-R	TTGTGGAAATtGtTTATGGGTTGAAT ATAAACCAAACTCTCTATACCACTAAACC	Bisulfite sequencing
DT-963-F DT-963-R	TTTGTGTTGTGTGTTTTGTTTGG AACAATATAAACTAACCAACACAAAAAACC	Bisulfite sequencing
DT-1036-F DT-1036-R	AAAAGTGGATAGTTTATGTTGGTGAA TCTCATCCATTATTCCCTTAAACCAC	Bisulfite sequencing
DT-528-F DT-528-R	tGGATTGAATTTAGGATATTGGTATAATTTG cTCTACACAAAAATTATAAATCTCTATCTACTC	Bisulfite sequencing
DT-239-F DT-239-R	CAAAAACAAATAATCCACCCTTACC GATTATTGTAGGTGGTAAAATGGATAG	Bisulfite sequencing
DT-1099-F DT-1099-R	AAGAGTGAAGTGAGAAGAGATATATAG AAACAATATTCTTCACAATTCCCTTC	Bisulfite sequencing
DT-1101-F DT-1101-R	GAAAATGTTGTTTGTGTGGAAG TTCAATCCCATTCCATATCCC	Bisulfite sequencing