Manuscript Number: PLAN2656R1

Title: Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks

Article Type: Manuscript

Section/Category:

Corresponding Author: Dr Teresa Roldan-Arjona,

Corresponding Author's Institution: Universidad de Cordoba

First Author: Ana-Pilar Ortega-Galisteo

Order of Authors: Ana-Pilar Ortega-Galisteo; Teresa Morales-Ruiz; Rafael R Ariza; Teresa Roldan-Arjona

Abstract: Cytosine DNA methylation is a stable epigenetic mark for maintenance of gene silencing across cellular divisions, but it is a reversible modification. Genetic and biochemical studies have revealed that the Arabidopsis DNA glycosylase domain-containing proteins ROS1 (REPRESSOR OF SILENCING 1) and DME (DEMETER) initiate erasure of 5-methylcytosine through a base excision repair process. The Arabidopsis genome encodes two paralogs of ROS1 and DME, referred to as DEMETER-LIKE proteins DML2 and DML3. We have found that DML2 and DML3 are 5-methylcytosine DNA glycosylases that are expressed ina wi de range of plant organs. We analyzed the distribution of methylation marks at two methylated loci in wild-type and dmlmu tant plants. Mutations in DML2 and/or DML3 lead to hypermethylation of cytosine residues that are unmethylated or weakly methylated in wild-type plants. In contrast, sites that are heavily methylated in wild-type plants are hypomethylated in mutants. These results suggest that DML2 and DML3 are

required not only for removing DNA methylation marks from improperly-methylated cytosines, but also for maintenance of high methylation levels in properly targeted sites.

Suggested Reviewers: Hervé Vaucheret
herve.vaucheret@versailles.inra.fr

Anne B Britt
abbritt@ucdavis.edu

Svend K Petersen-Mahrt
skpm@cancer.org.uk

Hiroshi Sano
sano@gtc.naist.jp

Response to Reviewers:

ANSWERS TO REVIEWER 1:

COMMENT: "since DNA methylation patterns were reported to correlate with environmental stresses, the authors may be able to check whether or not their mutants differently behave upon stresses such as low temperature, salt, drought, pathogen infection etc".
OUR RESPONSE: We have checked whether the dml2 and dml3 mutants behave differently upon salt stress. These data are now included as a new Supplementary figure (Supplementary Figure 1; previous Supplementary Figure 1 is now renamed Supplementary Figure 2) and discussed in pages 12 and 15.

COMMENT: "Also I suggest the authors to show the global methylation status in the mutant lines by biochemical techniques such as HPLC and Southern hybridization".
OUR RESPONSE: Global methylation levels in dml2 or dml3 mutants are not altered, as recently showed by Penterman et al (2007) using a highly-sensitive method, and we did not regard as necessary to replicate this work with our mutant plants using less sensitive techniques. The absence of global methylation changes is now discussed in relation to our previous data with ros1 mutants (Gong et al., 2002) in page 15.

COMMENT: "Abstract: "removing wrongly-deposited DNA methylation". How do authors judge "wrong" patterns? Data simply show different patterns between mutants and wild type plants".
OUR RESPONSE: We judge that the methylation patterns observed in WT plants as compared to mutants are "normal" or "regular" patterns. We have changed "wrongly" to "improperly" (according to its meaning of "not regularly or normally formed"). This change has been made in the Abstract, and also in pages 13 and 16.

COMMENT: "Introduction: "(where N is any nucleotide)", N should be A, T or C".
OUR RESPONSE: The expressions CpNpG and CpNpN have been changed to CpHpG and CpHpH, respectively (where H stands for either A or T or C) throughout text and figures.

COMMENT: "Results: p9; "PCR reactions" should read "PCR", since R is abbreviation of reaction. p9, 3 lines from the bottom, data should be shown. This is critical control experiments"
Our reponse: In Results, p. 10: "PCR reactions" has been changed to "PCR". Also, the lack of activity of mutant DML2 and DML3 proteins is now showed in Figure 2.

COMMENT: "P11: figure 5C needs explanation as to the size and numbers below the photograph".
OUR RESPONSE: The figure 5C is now clarified in the corresponding figure legend (p. 21). Sizes and numbers below the photographs are now fully explained.

COMMENT: "P12: careful explanation on experiments is necessary. I was confused with "at 63 sites of the AtGP1 locus". Did authors randomly selected Cs in the sequenced region?
OUR RESPONSE: We examined all cytosine residues present in both sequenced regions. This is now made clear in page 12 and 13, including a detailed explanation of the bisulfite analysis, and also in the legend to figure 6 (page 21).

COMMENT: "Figure 6: describe clearly the unit or meaning of both vertical and horizontal axes. Also clear description on the method used is preferable".
OUR RESPONSE: The meaning of both vertical and horizontal axes in Figure 6 and the method used is now clearly described in the corresponding figure legend (page 21).

COMMENT: "There is a published paper describing isolation and characterization of tobacco ROS-related genes, which could be helpful for this article (Plant Biotechnology 24: 339-344, 2007)".

OUR RESPONSE: The article describing the isolation and characterization of tobacco ROS-related genes is now cited in the Introduction (page 4).

ANSWERS TO REVIEWER 2:

COMMENT: It is a common error in nomenclature, not limited to the authors of the submitted manuscript, that CpNpG and CpNpN do not define what is meant: correct is CpHpG and CpHpH, since H stands for either A or T or C according to convention. N stands for any of the four bases and therefore includes G, which does not distinguish CpNpG from CpG(pG). Please replace the nomenclature accordingly throughout text and figures".
OUR RESPONSE: "The expressions CpNpG and CpNpN have been changed to CpHpG and CpHpH, respectively (where H stands for either A or T or C), throughout text and figures.

COMMENT: The elegant control that proteins with a mutation in the conserved amino acid loose their in vitro activity is strong evidence for the enzymatic activity of the wild type version. I regret that the authors state this as "data not shown" (p.9) and recommend including the data in the figure 2".
OUR RESPONSE: The lack of activity of mutant DML2 and DML3 proteins is now showed in Figure 2.

COMMENT: "The methylation analysis at FWA and GP1 are the basis for a major statement of the paper, namely the occurrence of hyper- AND hypo-methylation in the mutants. Any claim made on the basis of bisulfite sequencing analysis of DNA methylation requires a tight control about the quality of the chemical conversion. In the current work, I am missing data or a description how this control was done. This is required to strengthen the conclusion".
OUR RESPONSE: As a control for the efficiency of bisulfite conversion we used plasmid DNA containing each target region and propagated in a dcm mutant E. coli strain. The sequence analysis of both unmethylated plasmids confirmed a complete bisulfite conversion of all cytosines in the target sequences under the conditions used. This control is now fully described in the Materials and methods section (pages 8 and 9).

COMMENT: "Further question: there is a discrepancy between Figure 4A (high expression of DML2 and DML3 in flowers in the RT-PCR versus very low expression of the GUS fusion constructs in

young and mature flowers. This should be discussed. Also, I recommend to include the expression data for ROS and DME (stated on p.14 that they exist) in Figure 4A"

OUR RESPONSE: The discrepancy between high expression of DML2 and DML3 in flowers as detected by RT-PCR versus very low expression of the GUS fusion constructs in young and mature flowers (Figure 4A and 4B, respectively) is now discussed in page 14. We have also included the expression data for ROS1 and DME in Figure 4A.

COMMENT: "I am missing references to a few publications from others: there is substantial work on biochemical evidence for active demethylation in animal systems by the group of Jean-Pierre Jost (e.g. Jost et al. Nucleic Acid Res. 2001), and this should be cited in the Introduction. A second publication by Penterman et al. in Plant Physiol. 2007 already makes it clear that the interplay of methylation and demethylation activity is not as simple as thought before. The connection with the smallRNA/target match should be discussed and the paper cited".

OUR RESPONSE: The biochemical evidence for 5-meC excision in animals, including the work by the group of Jean-Pierre Jost, is now cited in the Introduction (page 4). The work by Penterman et al (2007, Plant Physiol, 145, 1549) is now cited, and the connection of methylation/demethylation processes with RNAi pathways discussed, in page 15.

COMMENT: "I recommend a thorough editing of spelling and grammar by a native speaker; this would make the MS easier to read".

OUR RESPONSE: A thorough editing of spelling and grammar has been performed, and we hope that the new version of the manuscript is easier to read.

Title:

Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks

Authors:

Ana Pilar Ortega-Galisteo, Teresa Morales-Ruiz, Rafael R. Ariza, Teresa Roldán-Arjona.

Authors' affiliation:

Departamento de Genética, Universidad de Córdoba, 14071 Córdoba, Spain

Number of words: 7,172
Number of Figures: 6
Supplementary material: 2 Tables and 2 Figures

Abstract

Cytosine DNA methylation is a stable epigenetic mark for maintenance of gene silencing across cellular divisions, but it is a reversible modification. Genetic and biochemical studies have revealed that the Arabidopsis DNA glycosylase domain-containing proteins ROS1 (*REPRESSOR OF SILENCING 1*) and DME (*DEMETER*) initiate erasure of 5-methylcytosine through a base excision repair process. The Arabidopsis genome encodes two paralogs of ROS1 and DME, referred to as DEMETER-LIKE proteins DML2 and DML3. We have found that DML2 and DML3 are 5-methylcytosine DNA glycosylases that are expressed in a wide range of plant organs. We analyzed the distribution of methylation marks at two methylated loci in wild-type and *dml* mutant plants. Mutations in *DML2* and/or *DML3* lead to hypermethylation of cytosine residues that are unmethylated or weakly methylated in wild-type plants. In contrast, sites that are heavily methylated in wild-type plants are hypomethylated in mutants. These results suggest that DML2 and DML3 are required not only for removing DNA methylation marks from improperly-methylated cytosines, but also for maintenance of high methylation levels in properly targeted sites.

Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks

Ana Pilar Ortega-Galisteo, Teresa Morales-Ruiz, Rafael R. Ariza, Teresa Roldán-Arjona.

Departamento de Genética, Universidad de Córdoba, 14071 Córdoba, Spain

Corresponding Author:
Teresa Roldán-Arjona
Departamento de Genética
Edificio Gregor Mendel
Campus de Rabanales s/n
Universidad de Córdoba
14071-Córdoba
SPAIN
Tel: +34 957 218 979
Fax: +34 957 212 072
e-mail: ge2roarm@uco.es

1

Abstract

Cytosine DNA methylation is a stable epigenetic mark for maintenance of gene silencing across cellular divisions, but it is a reversible modification. Genetic and biochemical studies have revealed that the Arabidopsis DNA glycosylase domain-containing proteins ROS1 (*REPRESSOR OF SILENCING 1*) and DME (*DEMETER*) initiate erasure of 5-methylcytosine through a base excision repair process. The Arabidopsis genome encodes two paralogs of ROS1 and DME, referred to as DEMETER-LIKE proteins DML2 and DML3. We have found that DML2 and DML3 are 5-methylcytosine DNA glycosylases that are expressed in a wide range of plant organs. We analyzed the distribution of methylation marks at two methylated loci in wild-type and *dml* mutant plants. Mutations in *DML2* and/or *DML3* lead to hypermethylation of cytosine residues that are unmethylated or weakly methylated in wild-type plants. In contrast, sites that are heavily methylated in wild-type plants are hypomethylated in mutants. These results suggest that DML2 and DML3 are required not only for removing DNA methylation marks from improperly-methylated cytosines, but also for maintenance of high methylation levels in properly targeted sites.

Introduction

DNA methylation is found in the genomes of diverse organisms including both prokaryotes and eukaryotes. In prokaryotes, DNA methylation occurs on both cytosine and adenine bases and encompasses part of the host restriction system (Wilson and Murray, 1991), causing transcriptional activation or repression (Casadesus and Low, 2006). In multicellular eukaryotes, however, methylation seems to be confined to cytosine bases and is associated with an inhibition of gene expression (Bender, 2004; Bird, 2002). Eukaryotic DNA methylation is detected in protists, fungi, plants and animals (Colot and Rossignol, 1999), and plays important roles in the establishment of developmental programs (Holliday and Pugh, 1975; Riggs, 1975) and in genome defense against parasitic mobile elements (Yoder et al., 1997). Hypermethylation of tumour suppressor genes is an important mechanism in the development of many common forms of cancer (Esteller, 2005). DNA methylation is performed by DNA-methyltransferases that catalyze transfer of a methyl group from S-adenosyl-L-methionine to carbon 5 of cytosine bases in DNA, giving rise to 5-methylcytosine (5-meC) (Goll and Bestor, 2005). Whereas mammalian DNA methylation is restricted to symmetrical CpG sequences, in plants cytosine methylation occurs in any sequence context: CpG, CpHpG and CpHpH (where H stands for either A or T or C) (Bird, 2002; Finnegan et al., 1998).

Although DNA methylation is a stable epigenetic mark for maintenance of gene silencing across cellular divisions, it is a reversible modification. Demethylation may take place as a passive process due to lack of maintenance methylation during several cycles of DNA replication, or as an active mechanism in the absence of replication (Kress et al., 2001). Extensive demethylation of the mammalian genome takes place in preimplantation embryos, first in the male pronucleus through an active mechanism independent of DNA replication and subsequently in both paternal and maternal chromosomes through a passive process (Li, 2002). Localized demethylation at specific genes occurs later throughout development and tissue differentiation (Frank et al., 1991).

Despite many attempts to identify the mechanism responsible for active DNA demethylation in animal cells, its enzymatic basis remains controversial (Wolffe et al., 1999; Kress et al., 2001). A proposed mechanism is disruption of the labile N-

glycosidic bond between the 5-meC base and the deoxyribose moiety in DNA, followed by replacement with an unmodified cytosine. A 5-meC-DNA glycosylase activity was first identified in chicken embryos (Jost et al., 1995) and found to copurify with a protein homologous to human thymine DNA glycosylase (TDG) (Zhu et al., 2000b; Neddermann et al., 1996). It was latter reported that methyl CpG binding protein 4 (MBD4), another human DNA glycosylase with no sequence similarity to TDG, also has 5-methylcytosine-DNA glycosylase activity (Zhu et al., 2000a). However, both TDG and MBD4 have a very weak activity on 5meC·G pairs compared to their activities towards U·G and T·G mismatches (Hardeland et al., 2003; Hendrich et al., 1999; Zhu et al., 2000a), and hence their roles in DNA demethylation remain unclear.

In plants, genetic and biochemical studies have revealed that the Arabidopsis DNA glycosylase domain-containing proteins ROS1 (*REPRESSOR OF SILENCING 1*) and DME (*DEMETER*) function as DNA demethylases (Choi et al., 2002; Gong et al., 2002; Morales-Ruiz et al., 2006; Gehring et al., 2006; Agius et al., 2006). ROS1 is required for release of transcriptional silencing of a hypermethylated transgene (Gong et al., 2002) and DME activates the maternal expression of two imprinted genes silenced by methylation (Choi et al., 2002). Both DME and ROS1 catalyze the release of 5-meC from DNA by a glycosylase/lyase mechanism (Morales-Ruiz et al., 2006). DME and ROS1 cleave the phosphodiester backbone at the 5-meC removal site by successive β,δ-elimination, leaving a gap that has to be further processed to generate a 3′-OH terminus suitable for polymerization and ligation (Morales-Ruiz et al., 2006; Gehring et al., 2006; Agius et al., 2006). Interestingly, DME and ROS1 erase 5-meC not only at CpG but also non-CpG sequences, which matches the pattern of DNA methylation in plants (Morales-Ruiz et al., 2006; Gehring et al., 2006). There is also biochemical evidence of 5-meC removal catalyzed by two ROS1-like proteins from tobacco plants (Choi and Sano, 2007).

In addition to ROS1 and DME, the genome of Arabidopsis encodes two additional paralogs, referred to as DEMETER-LIKE proteins DML2 and DML3 (Choi et al., 2002). We report here the isolation, functional characterization and expression analysis of DML2 and DML3. We have found that DML2 and DML3 are 5-meC DNA glycosylases that are expressed in a wide range of plant organs and are required for maintenance of proper DNA methylation patterns.

Materials and methods

Plant material and growth conditions

Arabidopsis thaliana plants (ecotype Columbia) were grown in pots in growth rooms under fluorescent lights (16 hr of light and 8 hr of dark) at 23ºC, or on Murashige and Skoog (MS) nutrient medium with 1 % sucrose and 0.8 % agar under constant white fluorescent light at 23ºC.

Cloning of DML2 and DML3 full-length cDNAs

Truncated clones lacking the 5' portions of *DML2* and *DML3* cDNAs were isolated from an Arabidopsis cDNA library (Kieber et al., 1993) (Arabidopsis Biological Resource Center, Ohio State University). The 5´ends of both cDNAs were obtained by RT-PCR on total RNA isolated from Arabidopsis plants (ecotype Columbia) using primers designed according to their genome sequence information (see Supplementary Table 2). After sequencing, a PCR error at position 976 of *DML3* cDNA was corrected by site-directed mutagenesis (Stratagene QuickChange™ Site-Directed Mutagenesis Kit). Verified fragments were connected by digestion and ligation, and a full-length cDNA was assembled for each gene.

Protein expression and purification

Full-length *DML2 and DML3* cDNAs were inserted into the pMAL-c2X expression vector (New England Biolabs) to obtain a *malE-* in-frame fusion. Expression in *E. coli Rosetta 2* (DE3) (Novagene) was induced by the addition of isopropyl-1-thio-β-D-galactopyranoside and the MBP-DML2 and MBP-DML3 fusion proteins were purified by amylose affinity chromatography using standard protocols.

Site-directed mutagenesis

Site-directed mutagenesis was performed using the QuickChange XL site-directed mutagenesis kit (Stratagene). The D903A mutation was introduced into pMAL-DML2 using the oligonucleotides EN3D903A-F and EN3D903A-R (see Supplementary Table 2). The D672A mutation was introduced into pMAL-DML3 using the primers EN4D672A-F and EN4D672A-R. The mutant sequences were confirmed by DNA sequencing and the constructs were used to transform *E. coli* strain *Rosetta 2* (DE3) (Novagene). Mutant proteins were overexpressed and purified as described above.

DNA substrates

Oligonucleotides used as DNA substrates (see Supplementary Table 1) were synthesized by Operon and were purified by PAGE before use. Double-stranded DNA substrates were prepared by mixing a 5-µM solution of a 5'-fluorescein-labelled oligonucleotide (upper-strand) with a 10-µM solution of an unlabelled oligomer (lower-strand), heating to 95ºC for 5 min and slowly cooling to room temperature.

Oligonucleotide incision assay

Double-stranded oligodeoxynucleotides (2 pmol) were incubated at 30°C for 3 hr in a reaction mixture containing 50 mM Tris-HCl at pH 8.0, 1 mM EDTA, 1 mM DTT, 0.1 mg/mL BSA, and 5 pmol of protein in a total volume of 100 µL. Reactions were stopped by adding EDTA to 20 mM, sodium dodecyl sulfate to 0.6 %, and proteinase K to 200 µg/mL, and the mixtures were incubated for 30 min at 37ºC. DNA was extracted with phenol:chloroform:isoamyl alcohol (25:24:1) and ethanol precipitated at -20ºC in the presence of 0.3 mM NaCl and 16 µg/mL glycogen. Samples were resuspended in 10 µL of formamide dye mix (80% formamide, 1 mg/mL bromophenol blue, 10 mM EDTA), and heated at 95ºC for 5 min. Reaction products were separated in a 12% denaturing polyacrylamide gel containing 7 M urea. Fluorescein-labelled DNA was visualized using the blue fluorescence mode of the FLA-5100 imager and analysed using Multigauge software (Fujifilm).

Methylation-sensitive PCR assay

Arabidopsis genomic DNA (1 µg) was isolated from 4-weeks plants and incubated at 30 °C for 3 hr in a reaction mixture containing 50 mM Tris-HCl at pH 8.0, 1 mM EDTA, 1 mM DTT, 0.1 mg/mL BSA, and 0.5-2.0 pmol of DML2 protein in a total volume of 20 µL. The treated DNA was cleaned up and 2 µl were used for PCR to amplify *AtGP1*, *CACTA* and *ACTIN1* sequences. Primers used are listed in Supplementary Table 2.

Reverse Transcriptase–Mediated PCR

Total RNA from different plant tissues of WT plants (ecotype Columbia) was isolated with the RNeasy plant kit (Qiagen) and treated with RNAse-free DNAse (Amersham Pharmacia Biotech). Reverse transcription was performed using Moloney Murine Leukemia virus reverse transcriptase (GIBCO, BRL) according to the

manufacturer instructions. The complementary strand (1 µl) was used for PCR amplification, using specific primers (Supplementary Table 2). The cDNAs were amplified by denaturation at 94ºC for 30s, annealing at 62ºC for 30 s, and extension at 72ºC for 1 min, with an initial denaturation step of 5 min at 94ºC and a final elongation step at 72ºC for 10 min. Aliquots of the PCR products were analyzed by electrophoresis in 2 % agarose gels. Equal loading of each amplified gene sequence was confirmed by equal intensities of the *ACTIN1* control PCR product.

Real-time amplifications of *MET1* and *ACTIN1* were performed in a MyiQ Single Color Real Time PCR Detection System (Bio-Rad) with iQ SYBR Green Supermix using specific primers (Supplementary Table 2) and 1 µg of cDNA in a total volume of 25 µl. The cDNAs were amplified as described above. Each experiment, from RNA extraction to quantification, was performed twice independently, using triplicate samples. Data were analyzed using the $\Delta\Delta$Ct method, as described by (Livak and Schmittgen, 2001). The fold-change in *MET1* expression was calculated as $2^{-\Delta\Delta Ct_{WT-mutant}}$, where Ct is the cycle threshold value for each gene and $\Delta\Delta Ct_{WT-mutant} = (Ct_{MET1}-Ct_{ACTIN1})_{WT} - (Ct_{MET1}-Ct_{ACTIN1})_{mutant}$.

GUS staining

The upstream regions of *DML2* and *DML3* genes (+15 to –2269 and +13 to -1300 from first ATG, respectively) were PCR-amplified and inserted into the pCAMBIA1381 Xa binary vector in order to place the ß-glucuronidase (GUS) gene under the control of the *DML2 or DML3* promoters. Constructs were introduced into *Agrobacterium* strain LB4404 and transformed into WT Arabidopsis. Five independent hygromycin resistant transgenic plants were obtained and analyzed for each gene. Staining for detection of GUS activity employed Staining Solution [5-bromo-4-chloro-3-indolyl-ß-D-glucuronide (X-gluc) dissolved at 1 mg/ml in 100 mM sodium phosphate buffer (pH 7.0) containing 0.1% Triton X-100 and 0.05% NaN$_3$]. Specimens were placed in 1.5 mL Eppendorf tubes containing Staining Solution, twice vacuum-infiltrated for 10 min, then incubated for 12 hr at 37ºC. Stained samples were then washed twice with 70% ethanol and several times with 95% ethanol. The GUS staining was visualized using a Leica MZ FLIII stereomicroscope.

Molecular characterization of DML2 and DML3 insertional mutants

The following primer combinations (see Supplementary Table 2) were used to amplify DNA flanking the T-DNA. For the *dml2* mutation, the left border–flanking DNA was amplified with LBa1 and EN3248-0R, and the right border-flanking DNA with RBa1 and EN3248-0L. For the *dml3* mutation, the left border–flanking DNA was amplified with LBa1 and EN4440R. The DNA flanking the right border in the *dml3* insertion mutant could not be amplified with primers located on the T-DNA. PCR products were gel purified using a QIAquick Gel Extraction kit (Qiagen), cloned using the bacterial vector pGEM-T (Promega) and sequenced.

RT-PCR was performed on whole-plant total RNA of WT, *dml2*[-/-] and *dml3*[-/-] plants as described above. The following primer-pairs were used: DML2p3 and DML2p4 to amplify a 490-bp fragment of the 3´coding region of the *DML2* cDNA; Lba1 and DML2p1 to amplify a 280-bp bp junction fragment of the T-DNA and the 5´coding region of the *DML2* cDNA; Rba1 and DML2p2 to amplify a 1104-bp junction fragment of the T-DNA and the 3´coding region of the *DML2* cDNA; DML2p1 and DML2p2 to amplify the 5´and 3´coding regions of the *DML2* cDNA that flank the T-DNA insertion. DML3p3 and DML3p4 to amplify a 375-bp fragment of the 3´coding region of the *DML3* cDNA; Lba1 and DML3p2 to amplify a 492-bp junction fragment of the T-DNA and the 3´coding region of the *DML3* cDNA; DML3p1 and DML3p2 to amplify the 5´and 3´coding regions of the *DML3* cDNA that flank the T-DNA insertion. PCR conditions were as follows: 5 min at 94°C, 35 cycles of 30 s at 94°C, 30 s at 60°C and 45 s at 72°C, followed by 10 min at 72°C.

Bisulfite sequencing

Genomic DNA, isolated from plants of each genotype, was cleaved with HindIII and purified. Bisulfite sequencing was performed using the EZ DNA Methylation Kit (Zymo Research). DNA was treated with sodium bisulfite and cleaned up following the manufacturer's instructions. The target region was amplified using primers described previously (Zhu et al., 2007) (see Supplementary Table 2). PCR products were cloned in pGEM-T Easy (Promega) or pMBL-T (Dominion MBL) and sequenced. For each sample, at least 15 clones were sequenced. Sequences were aligned using the ContigExpress and AlignX modules of Vector NTI Advance 10 Software Package (Invitrogen) and compared to the DNA sequence from untreated samples. Two control

reactions for the efficiency of bisulfite conversion were set up using plasmid DNA containing each sequenced region and propagated in a *dcm* mutant *E. coli* strain. The sequence analysis of both unmethylated plasmids confirmed a complete bisulfite conversion of all cytosines in the target sequences under the conditions used.

Results

DNA glycosylase activity of DML2 and DML3

DML2 and *DML3* full-length cDNAs were isolated from purified Arabidopsis leaf mRNA by RT–PCR with primers designed according to partial cDNA clones and predicted gene annotations (Locus Tag At3g10010 and At4g34060, respectively). They encode proteins of 1332 and 1105 amino acids, respectively. Sequence analysis showed that both proteins are closely related to DME and ROS1, with high sequence similarity at the DNA glycosylase domain and the C-terminal region (Figure 1).

The DNA glycosylase domain contains a helix-hairpin-helix Gly/Pro-rich (HhH-G/PD) motif with essential amino acids for substrate recognition and catalysis. Both DML2 and DML3 possess in this region an aspartic acid residue that is invariant among all HhH-G/PD proteins, and a lysine conserved in the sub-class of bifunctional DNA glycosylase/lyases. Also like DME and ROS1, both DML2 and DML3 present downstream the HhH-G/PD motif a characteristic pattern of four cysteines (Cys-X_6-Cys-X_2-Cys-X_5-Cys), which in *E. coli* endonuclease III ligate an iron-sulfur [4Fe-4S] cluster (Thayer et al., 1995).

We fused the *DML2* and *DML3* cDNAs to the maltose-binding protein (MBP) gene and expressed the fusion proteins in *E. coli*. The purified proteins were used in oligonucleotide incision assays on DNA substrates containing 5-meC in diverse sequence contexts (Figure 2A). We found that DML3 incised an oligonucleotide duplex containing a single 5-meC at CpG, CpHpG or CpHpH sequences. In contrast, no incision was detected with DML2 at any of these contexts. The incision activity of DML3 was similar to that of DME and ROS1, and generated the same two types of fragments, which represent β- and β,δ-elimination products (Morales-Ruiz et al., 2006).

Although we did not detect any activity of DML2 in the oligonucleotide incision assay, this could be due to the limitations of this method, such as the use of a short target DNA molecule and/or specific context sequences. To determine if DML2 was

9

indeed able to incise methylated DNA we incubated the enzyme with Arabidopsis genomic DNA and performed PCR to amplify either unmethylated or methylated sequences (Figure 2C). In this assay, disappearance or reduced levels of the PCR product after incubation with the enzyme indicates accumulation of DNA incisions at the studied region. We found that incubation with increasing amounts of DML2 did not affect the amplification of an unmethylated region, such as the *ACTIN1* gene, but strongly reduced the amplification of highly methylated regions such as the *AtGP1* and *CACTA* transposons (Figure 2C). We therefore conclude that both DML2 and DML3 possess DNA glycosylase/lyase activity on methylated DNA.

To verify that the DNA incision activity is intrinsic to DML2 and DML3 we generated mutant proteins in which a conserved aspartic acid residue in the glycosylase domain (Figure 1) was changed to alanine. Preparations of both mutant proteins, purified by the same procedure as the WT proteins described above, lacked any detectable incision activity (Figure 2B and 2D).

We next analyzed the DNA glycosylase activity of DML3 on DNA substrates containing a 5-meC•G pair or a T•G mispair positioned in different sequence contexts. DML3 efficiently processed 5-meC residues and T•G mismatches when positioned in a CpG context but was less active when they were displaced one nucleotide away to a CpC context (Figure 3, upper panel). DML3 also incised 5-meC and thymine in a symmetric CpApG context, but showed a reduced activity on both types of residues when they were located in an asymmetric context (Figure 3, lower panel). Overall, the incision activity on the upper strand for all symmetric sequences was not significantly affected by the presence of a methylated cytosine in the lower strand (Figure 3 and data not shown). We conclude that DML3 has 5-meC and thymine-DNA glycosylase activity with a preference for CpG and CpHpG sequences.

Analysis of DML2 and DML3 expression

We performed RT-PCR analysis of *DML2* and *DML3* expression in different plant organs, and found that both genes are expressed in cauline leaves, flowers, stems, siliques and roots (Figure 4A). Neither *DML2* nor *DML3* transcripts were detected in mature seeds. We also analyzed *ROS1* and *DME* expression by RT-PCR and found that their transcripts were present in all organs tested. This is in agreement with our earlier analysis of ROS1 expression in *ROS1:GUS* transgenic plants (Gong et al., 2002), but

challenges a previous suggestion that *DME* is expressed primarily in the central cell of the female gametophyte (Choi et al., 2002).

In order to determine the tissue and developmental expression pattern of *DML2* and *DML3* genes, their promoters were fused with the β-glucuronidase reporter gene (GUS), and the resulting constructs were introduced into WT Arabidopsis plants (Figure 4B). Histochemical localization of GUS activity revealed that *DML2* and *DML3* promoters were very active in a variety of tissues. Young seedlings recently emerged from their seed coats (2 d after imbibition), showed strong GUS staining in the cotyledons and in the hypocotyl adjacent region. As the seedlings developed, staining was observed in leaf primordia and growing leaves. No staining was observed in flower buds, but as the flower developed GUS activity appeared in sepals and stamen filaments. Staining was also evident in mature siliques.

Isolation of Arabidopsis dml2 and dml3 mutants

To analyze the function of *DML2* and *DML3 in vivo* we searched for mutants in public T-DNA insertion line collections. Two Arabidopsis mutant lines with insertions in the *DML2* (SALK_N641248) and *DML3* (SALK_N556440) coding sequences were found in the SIGnAL T-DNA insertion Arabidopsis mutant database (Alonso et al., 2003).

Plants homozygous for *dml2* and *dml3* T-DNA insertions were identified by PCR. In order to characterize the insertions molecularly, the DNA flanking the T-DNA was amplified and the PCR products sequenced (Figure 5). The *dml2* allele T-DNA is inserted in intron 10, and is combined with a 113 bp deletion that includes the complete exon 11 and the first 2 bp of exon 12, as well as the addition of filler DNA on both sides of the T-DNA (Figure 5B). The *dml3* allele is more complex: the T-DNA is inserted in exon 6, with an intact left-border and a truncated right-border, also accompanied by the insertion of filler DNA from other regions of the genome (Figure 5B).

Plants homozygous for the T-DNA insertions were selected by PCR, and RT-PCR analysis performed to assess the presence of *DML2* and *DML3* transcripts in total RNA isolated from WT and mutant plants (Figure 5C). RT-PCR of *DML2* and *DML3* coding regions with T-DNA- and *DML2*- or *DML3*-specific primers revealed the presence of a transcript containing the T-DNA in both *dml2*[-/-] and *dml3*[-/-] plants. Sequence analysis showed that the mutant *DML2* transcript would encode a truncated protein of 960 amino

acids. This putative mutant protein would lack both the DNA glycosylase domain and the C-terminal region that is highly conserved in DML2, DML3, DME and ROS1 wild-type proteins. The T-DNA insertion in exon 6 of DML3 also prevents the synthesis of a full-length polypeptide. If these mutant proteins are synthesized at all, they are most probably non-functional. In addition, RT-PCR analysis with DML2- or DML3-specific primers corresponding to the 5´end region of both genes detected a product in WT plants but not in *dml2*[-/-] or *dml3*[-/-] mutant plants. This suggests that the mutant transcripts are unstable and/or the T-DNA insertion prevents in both cases the synthesis of full-length mRNA. We conclude that the T-DNA insertions in *DML2* and *DML3* inactivate the wild-type function of these genes, leading to a null phenotype.

We also constructed a double homozygous *dml2*[-/-] *dml3*[-/-] mutant line. Neither the single nor the double mutant plants showed any obvious phenotypic alterations under the grown conditions used in this study. Since DNA methylation changes have been reported during plant responses to stress (Boyko and Kovalchuk, 2008), we tested whether the *dml2*[-/-] or *dml3*[-/-] mutants showed altered sensitivity to salt stress (Supplementary Figure 1). No significant difference in the sensitivity of mutant compared to WT plants was found in a Student's t-test.

DML2 and DML3 are required for appropriate distribution of 5-meC in methylated sequences

To determine the function of DML2 and DML3 in demethylation, we carried out bisulfite sequencing analysis of the gypsy-class LTR (long-terminal repeat) retroelement *AtGP1* (Lippman et al., 2003) and a region of approximately 1.4 Kb just upstream of the translation start site of the *FWA* gene (Soppe et al., 2000) in WT and mutant plants (Figure 6).

We analyzed methylation at all 63 cytosine residues present in the sequenced region of the *AtGP1* locus, including 15 CpG, 29 CpHpG and 19 CpHpH sites. We found that in WT plants the CpG sites are heavily methylated (average 81 % methylation), whereas CpHpG sites show lower methylation levels (29 %) and CpHpH are mostly not methylated (3 %) (Figure 6A). The upstream region of *FWA* shows 99 cytosine residues, distributed in 20 CpG, 13 GpHpG and 66 CpHpH sites. WT plants displayed strong methylation of the CpG sites (average 85 % methylation), and almost no methylation at CpHpG and CpHpH sites (7% and 4%, respectively) (Figure 6A).

We next examined whether the deficiency in DML2 and/or DML3 had any effect in the methylation profiles of *AtGP1* and *FWA*. For every cytosine we calculated the increase or decrease in methylation observed in each mutant genotype (mutant- minus WT-methylation) (Figure 6B). In all three mutant genotypes we found that sites showing low or null methylation in the WT plants were mostly hypermethylated in mutants, whereas sites with strong methylation levels in WT plants showed a lower methylation level in DML2- and/or DML3-defficient plants. Hypomethylation in mutant plants affected mainly to heavily methylated CpG sites, but it was also observed for some strongly methylated CpHpG sites at the *AtGP1* locus in *dml3* and *dml2 dml3* mutants (Figure 6A). By other hand, CpG sites in *FWA* with intermediate to low methylation levels in WT were mostly hypermethylated in mutants. We conclude that deficiency of DML2 and/or DML3 has opposing effects on sites with divergent methylation levels, irrespective of the context.

Since DML3 not only removes 5-meC but also its deamination product thymine, we asked if the hypomethylation of heavily methylated sites observed in mutant plants was due to an increase in C $\rightarrow$ T mutations generated by spontaneous deamination of 5-meC to T. We amplified the *AtGP1* and *FWA* loci from DNA not treated with bisulfite and analyzed 15-20 clones for each locus and genotype. We did not detect any mutation in the analyzed region (data not shown). Thus, the hypomethylation observed for heavily methylated sites in mutants is not due to a C $\rightarrow$ T decay caused by a deficiency in thymine-DNA glycosylase activity.

We also reasoned that this hypomethylation could be due to a decrease in the expression of METHYLTRANSFERASE 1 (MET1), which maintains CpG methylation during replication (Chan et al., 2005 1558). We therefore quantified the expression of *MET1* by realt-time RT-PCR in WT and mutant plants (Supplementary Figure 2). No significant difference was found between different genotypes. Thus, hypomethylation of highly methylated sites is not due to down-regulation of *MET1* in the mutant backgrounds.

Altogether, these results suggest that both DML2 and DML3 are required not only for removing methylation marks from improperly-methylated cytosines, but also for maintaining high methylation levels in properly targeted sites.

Discussion

In this work we describe the expression analysis and functional characterization of DML2 and DML3, two Arabidopsis paralogs of DME and ROS1, and also examine the effect that a deficiency in DML2 and/or DML3 has on the methylation profile of two methylated loci in Arabidopsis.

We found that DML3 catalyzes *in vitro* the release of 5-meC from DNA by a glycosylase/lyase mechanism, and also removes thymine mismatched to guanine. The DNA glycosylase activity displayed by DML3 is quite similar to those of DME and ROS1 (Morales-Ruiz et al., 2006): it removes with similar efficiency both 5-meC and its deamination product, preferentially from symmetrical sequence contexts (CpG and CpHpG), and generates a mixture of β- and β,δ-elimination products upon base excision. We could not detect 5-meC DNA glycosylase/lyase activity in DML2 using the same oligonucleotide incision assay. This result agrees with data reported in a recent analysis of DML2 activity *in vitro* (Penterman et al., 2007b), which failed to detect significant incision activity of DML2 in the same test. However, we found that DML2 was able to specifically incise methylated but not unmethylated genomic regions in Arabidopsis DNA, as detected through a PCR-mediated assay. The causes of the inability to detect DML2 incision activity on methylated DNA oligonucleotides *in vitro* are currently unknown. Since both the oligonucleotide and the PCR-mediated incision assays have similar detection limits (our unpublished data), we speculate that perhaps DML2 activity is better detected in longer DNA substrates containing multiple methylation sites.

Both DML2 and DML3 are expressed in a variety of plant organs and tissues throughout the plant life cycle, as detected through fusion of their promoters to a GUS reporter gene. Although *DML2* and *DML3* promoters displayed a very similar activity pattern, RT-PCR analysis showed a wider organ distribution of *DML3* trascripts compared to those of *DML2*. There is also a disparity between the weak expression observed for *DML2:GUS* and *DML3:GUS* fusions in young and mature flowers and the high expression level detected using RT-PCR. These discrepancies may be due to differences in mRNA stability for both genes in some tissues. Similarly to *DML2* and *DML3*, *ROS1* and *DME* transcripts are also present in a wide range of tissues. Altogether, our results suggest that all four DNA glycosylases of the DME family have

a broad range of expression in different plant organs and tissues. There are some, but still limited, data about the expression of other DNA glycosylases in Arabidopsis plants. The *AtOGG1* and *AtNTH1* genes encode small DNA glycosylases/lyases that remove oxidized purines and pyrimidines, respectively, and they show a broad pattern of expression in different plant organs (Garcia-Ortiz et al., 2001; Roldan-Arjona et al., 2000). The 3-methyladenine glycosylase (*αMAG*) gene of Arabidopsis is also expressed in many tissues, including those with low rates of cell division, such as leaves (Shi et al., 1997). Such a broad expression pattern has been related to a genome maintenance function required in both replicating and non-replicating cells (Garcia-Ortiz et al., 2001; Roldan-Arjona et al., 2000; Shi et al., 1997).

We have obtained mutant plants homozygous for a T-DNA insertion in *DML2* and/or *DML3*. We were not able to detect any significant differences in phenotype between the single or double mutants and the WT plants. All mutants germinated, grew and developed normally, and were fertile. Furthermore, dml2 and dml3 mutants do not show any altered sensitivity to salt stress. However, we have found that DML-deficient plants display alterations in the distribution of methylation marks at methylated loci, compared to the normal pattern observed in WT plants.

A recent report based on 5-meC immunocapturing followed by genome-tiling microarrays analysis (Penterman et al., 2007b) has suggested that an important function of DML glycosylases is to protect the genome from excess methylation, such as previously proposed for ROS1 (Zhu et al., 2007). The demethylation activity of DML proteins and ROS1 does not target all loci, since global methylation levels are not altered in *dml* or *ros1* mutants compared to WT plants (Penterman et al., 2007b; Gong et al., 2002). There is genetic evidence that DML enzymes remove methylation directed by the PolIV/RDR2/DCL3/AGO4 RNAi pathway against transposable elements and repetitive sequences, and that ROS1 is also able to remove DNA methylation directed by an RDR2-independent RNAi pathway (Penterman et al., 2007a). To gain insight in the function of DML2 and DML3 we have performed a detailed examination of the DNA methylation status at two loci: the retrotransposon *AtGP1* and a region of two repetitive sequences upstream the *FWA* gene. Rather than averaging for different sequence contexts, we have compared the methylation level in mutants to the methylation level in WT plants for each cytosine residue. We found that DML-deficient plants display hypermethylation across different cytosine residues, irrespective of their

sequence context, and this is in agreement with our data showing that DML3 5-meC glycosylase targets both CpG and non-CpG sequences. Our results also agree with the hypermethylation found for the same loci in *ros1* mutants (Zhu et al., 2007). However, and unlike *ros1* mutants, *dml2* and/or *dml3* mutants display hypomethylation in cytosines that are heavily methylated in WT plants. We have ruled out the possibility that this hypomethylation is due to a steady accumulation of C $\rightarrow$ T mutations in strongly methylated sequences and/or to the downregulation of the *MET1* gene in mutant plants. Thus, our results suggest that DML2 and DML3 are required not only for removing methylation marks from improperly-methylated sites, but also for maintaining high methylation levels in properly targeted sites.

The requirement of DML2 and DML3 for maintenance of heavily methylated sites is intriguing. One possibility is that DML proteins play a direct role during the methylation process. In mammals, for example, the T:G mismatch DNA glycosylase (TDG), a multifunctional protein that is involved in both DNA repair and transcriptional regulation, interacts with DNA methyltransferases Dnmt3a and 3b on untranscribed chromatin (Gallais et al., 2007; Li et al., 2007). This interaction might allow fast remethylation of sites affected by methylcytosine deamination (Li et al., 2007). Another possibility is that the hypermethylation of unproperly targeted sites in *dml* mutants leads to changes in chromatin modification that in turn affect targeting of CpG and CpHpG methylation. In this scenario, the accumulation of unproperly methylated sites would indirectly lead to a reduced methylation of properly methylated sites. In any case, the opposing effects of a DML2 or DML3 deficiency on sites with divergent methylation levels indicate that the roles of both enzymes in DNA methylation/demethylation dynamics may be more complex than previously suspected.

Acknowledgements

References

Agius F, Kapoor A, Zhu JK (2006). Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. Proc Natl Acad Sci USA 103: 11796-11801.

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, *et al.* (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301: 653-657.

Bender J (2004). DNA methylation and epigenetics. Annu Rev Plant Physiol Plant Mol Biol 55: 41-68.

Bird A (2002). DNA methylation patterns and epigenetic memory. Genes Dev 16: 6-21.

Boyko A, Kovalchuk I (2008). Epigenetic control of plant stress response. Environ Mol Mutagen 49: 61-72.

Casadesus J, Low D (2006). Epigenetic gene regulation in the bacterial world. Microbiol Mol Biol Rev 70: 830-856.

Chan SW, Henderson IR, Jacobsen SE (2005). Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat Rev Genet 6: 351-360.

Choi C, Sano H (2007). Identification of tobacco genes encoding proteins possessing removal activity of 5-methylcytosines. Plant Biotechnology 24: 339–344.

Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer RL (2002). DEMETER, a DNA glycosylase domain protein, Is required for endosperm gene imprinting and seed viability in Arabidopsis. Cell 110: 33-42.

Colot V, Rossignol JL (1999). Eukaryotic DNA methylation as an evolutionary device. Bioessays 21: 402-411.

Esteller M (2005). Aberrant DNA methylation as a cancer-inducing mechanism. Annu Rev Pharmacol Toxicol 45: 629-656.

Finnegan EJ, Genger RK, Peacock WJ, Dennis ES (1998). DNA methylation in plants. Annu Rev Plant Physiol Plant Mol Biol 49: 223-247.

Frank D, Keshet I, Shani M, Levine A, Razin A, Cedar H (1991). Demethylation of CpG islands in embryonic cells. Nature 351: 239-241.

Gallais R, Demay F, Barath P, Finot L, Jurkowska R, Le Guevel R, Gay F, Jeltsch A, Metivier R, Salbert G (2007). Dnmt 3a and 3b associate with the nuclear orphan receptor COUP-TFI during gene activation. Mol Endocrinol 21: 2085-2098.

Garcia-Ortiz MV, Ariza RR, Roldan-Arjona T (2001). An OGG1 orthologue encoding a functional 8-oxoguanine DNA glycosylase/lyase in *Arabidopsis thaliana*. Plant Mol Biol 47: 795-804.

Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL (2006). DEMETER DNA glycosylase establishes *MEDEA* polycomb gene self-imprinting by allele-specific demethylation. Cell 124: 495-506.

Goll MG, Bestor TH (2005). Eukaryotic cytosine methyltransferases. Annu Rev Biochem 74: 481-514.

Gong Z, Morales-Ruiz T, Ariza RR, Roldan-Arjona T, David L, Zhu JK (2002). ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. Cell 111: 803-814.

Hardeland U, Bentele M, Jiricny J, Schar P (2003). The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs. Nucleic Acids Res 31: 2261-2271.

Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature 401: 301-304.

Holliday R, Pugh JE (1975). DNA modification mechanisms and gene activity during development. Science 187: 226-232.

Jost JP, Siegmann M, Sun L, Leung R (1995). Mechanisms of DNA demethylation in chicken embryos. Purification and properties of a 5-methylcytosine-DNA glycosylase. J Biol Chem 270: 9734-9739.

Kieber JJ, Rothenberg M, Roman G, Feldmann KA, Ecker JR (1993). CTR1, a negative regulator of the ethylene response pathway in Arabidopsis, encodes a member of the raf family of protein kinases. Cell 72: 427-441.

Kress C, Thomassin H, Grange T (2001). Local DNA demethylation in vertebrates: how could it be performed and targeted? FEBS Lett 494: 135-140.

Li E (2002). Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet 3: 662-673.

Li YQ, Zhou PZ, Zheng XD, Walsh CP, Xu GL (2007). Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair. Nucleic Acids Res 35: 390-400.

Lippman Z, May B, Yordan C, Singer T, Martienssen R (2003). Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS Biol 1: E67.

Livak KJ, Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ Method. Methods 25: 402-408.

Morales-Ruiz T, Ortega-Galisteo AP, Ponferrada-Marin MI, Martinez-Macias MI, Ariza RR, Roldan-Arjona T (2006). *DEMETER* and *REPRESSOR OF SILENCING 1* encode 5-methylcytosine DNA glycosylases. Proc Natl Acad Sci USA 103: 6853-6858.

Neddermann P, Gallinari P, Lettieri T, Schmid D, Truong O, Hsuan JJ, Wiebauer K, Jiricny J (1996). Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. J Biol Chem 271: 12767-12774.

Penterman J, Uzawa R, Fischer RL (2007a). Genetic interactions between DNA demethylation and methylation in Arabidopsis. Plant Physiol 145: 1549-1557.

Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, Fischer RL (2007b). DNA demethylation in the Arabidopsis genome. Proc Natl Acad Sci USA 104: 6752-6757.

Riggs AD (1975). X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet 14: 9-25.

Roldan-Arjona T, Garcia-Ortiz MV, Ruiz-Rubio M, Ariza RR (2000). cDNA cloning, expression and functional characterization of an Arabidopsis thaliana homologue of the Escherichia coli DNA repair enzyme endonuclease III. Plant Mol Biol 44: 43-52.

Shi L, Kent R, Bence N, Britt AB (1997). Developmental expression of a DNA repair gene in Arabidopsis. Mutat Res 384: 145-156.

Soppe WJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, Peeters AJ (2000). The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Mol Cell 6: 791-802.

Thayer MM, Ahern H, Xing D, Cunningham RP, Tainer JA (1995). Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. EMBO J 14: 4108-4120.

Wilson GG, Murray NE (1991). Restriction and modification systems. Annu Rev Genet 25: 585-627.

Wolffe AP, Jones PL, Wade PA (1999). DNA demethylation. Proc Natl Acad Sci USA 96: 5894-5896.

Yoder JA, Walsh CP, Bestor TH (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13: 335-340.

Zhu B, Zheng Y, Angliker H, Schwarz S, Thiry S, Siegmann M, Jost JP (2000a). 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. Nucleic Acids Res 28: 4157-4165.

Zhu B, Zheng Y, Hess D, Angliker H, Schwarz S, Siegmann M, Thiry S, Jost JP (2000b). 5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex. Proc Natl Acad Sci USA 97: 5135-5139.

Zhu J, Kapoor A, Sridhar VV, Agius F, Zhu JK (2007). The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in Arabidopsis. Curr Biol 17: 54-59.

Figure legends

Figure 1. DML2 and DML3 have a DNA glycosylase domain and are closely related to DME and ROS1. (Upper) Diagram of the four proteins showing the conserved regions as colored sections. (Lower) Amino-acid sequence alignment of *A. thaliana* DML2, DML3, DME, ROS1 and Nth1, *E. coli* Nth, and *H. sapiens* MutY and Ogg1. Asterisk marks the lysine residue that is diagnostic of a glycosylase/lyase activity; triangle indicates the conserved aspartic acid residue in the active site; diamonds label the cysteine residues that in *E. coli* Nth ligate a [4Fe-4S] cluster.

Figure 2. Enzymatic activity of DML2 and DML3 on methylated DNA. (A) Double-stranded oligonucleotide substrates containing 5-meC in CpG, CpHpG or CpHpH contexts were incubated with purified DME, ROS1, DML2 or DML3 as described in Methods. Reaction products were separated in a 12 % denaturing polyacrylamide gels. β and δ indicate β- and β,δ-elimination products, respectively. (B) Double-stranded oligonucleotide substrates containing 5-meC in CpG, CpHpG or CpHpH contexts were incubated with purified WT DML3 or mutant DML3 D672A (mDML3). (C, D) Arabidopsis genomic DNA was incubated with increasing amounts of purified WT DML2 (C) or mutant DML2 D903A (mDML2) (D) and amplified by PCR using specific primers for the indicated sequences.

Figure 3. DML3 excise 5-meC and thymine in different sequence contexts. Double-stranded oligonucleotide substrates containing 5-meC (M), or T•G mismatches in CpG (CXGG) and non-CpG (XCGG, XAG and AXT) sequence contexts were incubated with purified DML3 as described in Materials and Methods. Reaction products were separated in a 12 % denaturing polyacrylamide gel. β and δ indicate β– and β,δ-elimination products, respectively.

Figure 4. Expression analysisis of *DML2* and *DML3* (A) RT-PCR was performed using specific primers for *DML2*, *DML3*, *DME* or *ROS1* on total RNA from multiple plant tissues and aliquots (10 μl) of the products were analysed on 1.5 % agarose gel. *ACTIN1* gene was used as a control for constitutive expression (B) Histochemical localization of GUS activity in *DML2–GUS* (1, 3, 5, 7, 9 and 11) and *DML3-GUS* fusion (2, 4, 6, 8, 10, and 12) transgenic plants. 1 and 2: two-day seedlings; 3 and 4: five-day seedlings; 5 and 6: cauline leaves; 7 and 8: flower buds; 9 and 10: mature flowers; 11 and 12: siliques.

Figure 5. Molecular analysis of T-DNA insertions in the *DML2* and *DML3* loci. (A) Genomic organization of the *DML2* and *DML3* loci. Exons (grey boxes) and the positions of T-DNA insertions are shown. Primers used for genotyping and RT-PCR analysis are indicated as arrowheads. (B) Sequences of the T-DNA/plant genome junctions. Filler DNA is in bold lowercase. In the *dml3* allele, the DNA flanking the right border could not be amplified with primers located on the T-DNA. (C) RT-PCR analysis of the disrupted DML2 and DML3 genes. RT-PCR was carried out with the primers pairs indicated above the panels, using total RNA isolated from WT (lanes 1, 3, 5 and 7, both panels) *dml2$^{-/-}$* (lanes 2, 4, 6 and 8, left panel), and *dml3$^{-/-}$* (lanes 2, 4, 6 and 8, right panel) plants. A control reaction with no DNA is shown in lane 9, both panels. Sizes of fragments from a 100-bp ladder are indicated.

Figure 6. Bisulfite sequencing analisis of DNA methylation at *AtGP1* and *FWA* loci in WT plants, single *dml2$^{-/-}$* or *dml3$^{-/-}$* mutants, and double *dml2$^{-/-}$ dml3$^{-/-}$* mutants. In all cases, 15 to 20 clones were analyzed for each locus and genotype. (A) Percent methylation at all cytosine residues present in the sequenced regions of *AtGP1* and *FWA* loci in WT plants. Vertical axis measures percent methylation. Horizontal axis represents all cytosine residues present in the sequenced regions of *AtGP1* (63 sites, left panel) and *FWA* (99 sites, right panel), ordered from left to right according to their increasing methylation level. Black, red and green bars represent CpG, CpHpG and asymmetrical sites, respectively. (B) Increase or decrease in methylation observed for every site in each mutant genotype. The difference in percent methylation between mutant and wild-type plants was calculated for every cytosine residue in each locus. Vertical axis measures percent methylation (mutant- minus WT-methylation). Horizontal axis shows all cytosine residues present in the sequenced regions of *AtGP1* and *FWA* loci, ordered as in (A).

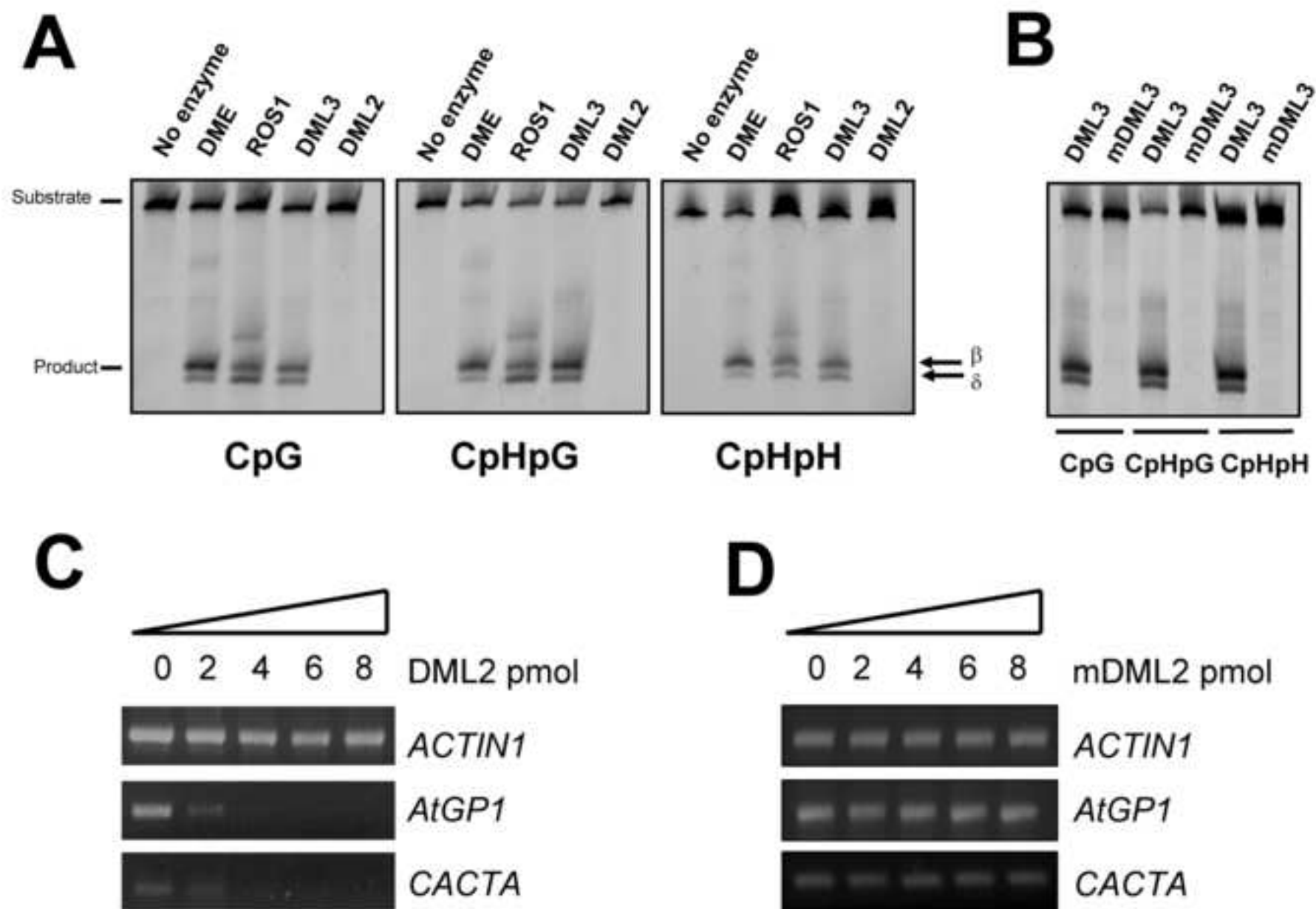**Figure 1**
Click here to download high resolution image

Figure 2
Click here to download high resolution image

Figure 2

Figure 3
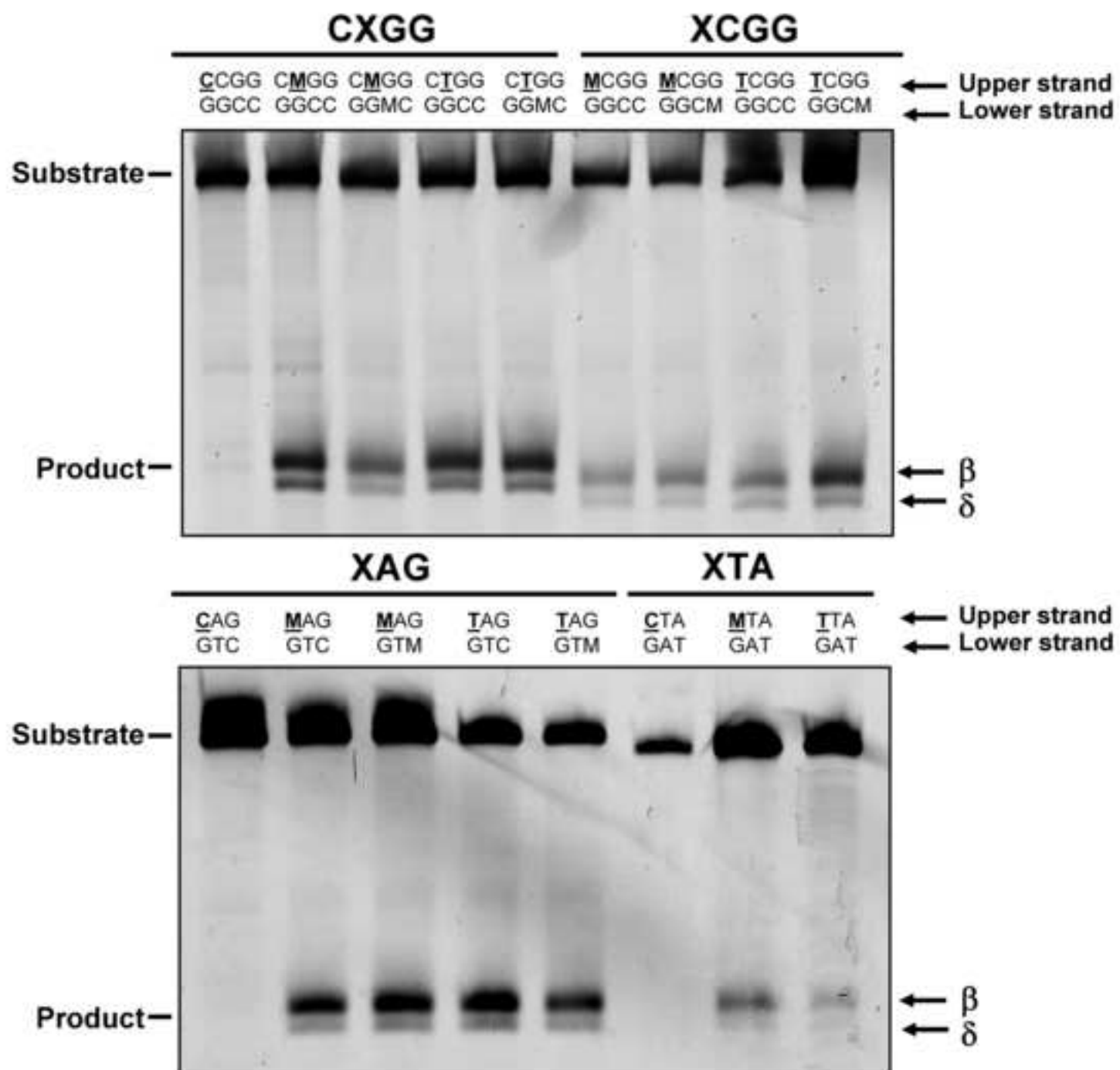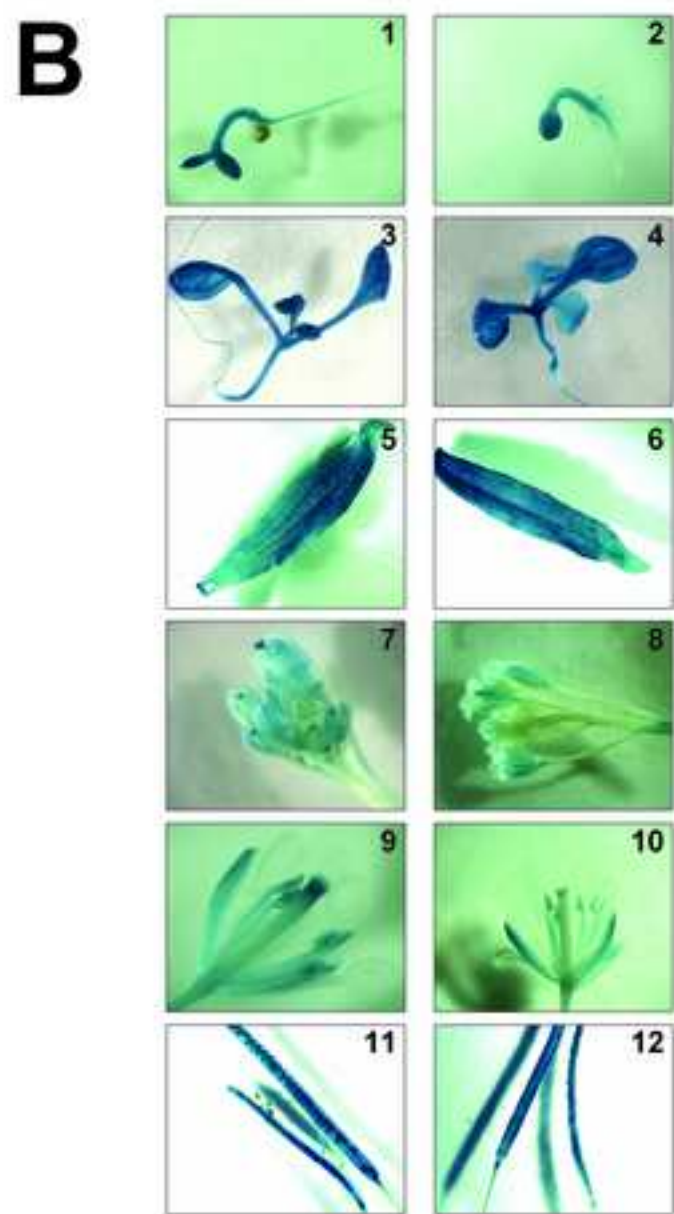**Click here to download high resolution image**

Figure 4
[Click here to download high resolution image](#)
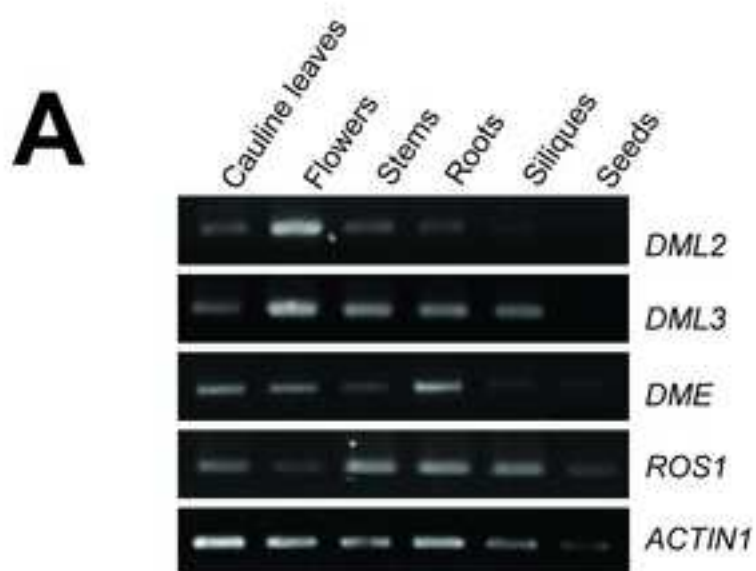
Figure 5
Click here to download high resolution image
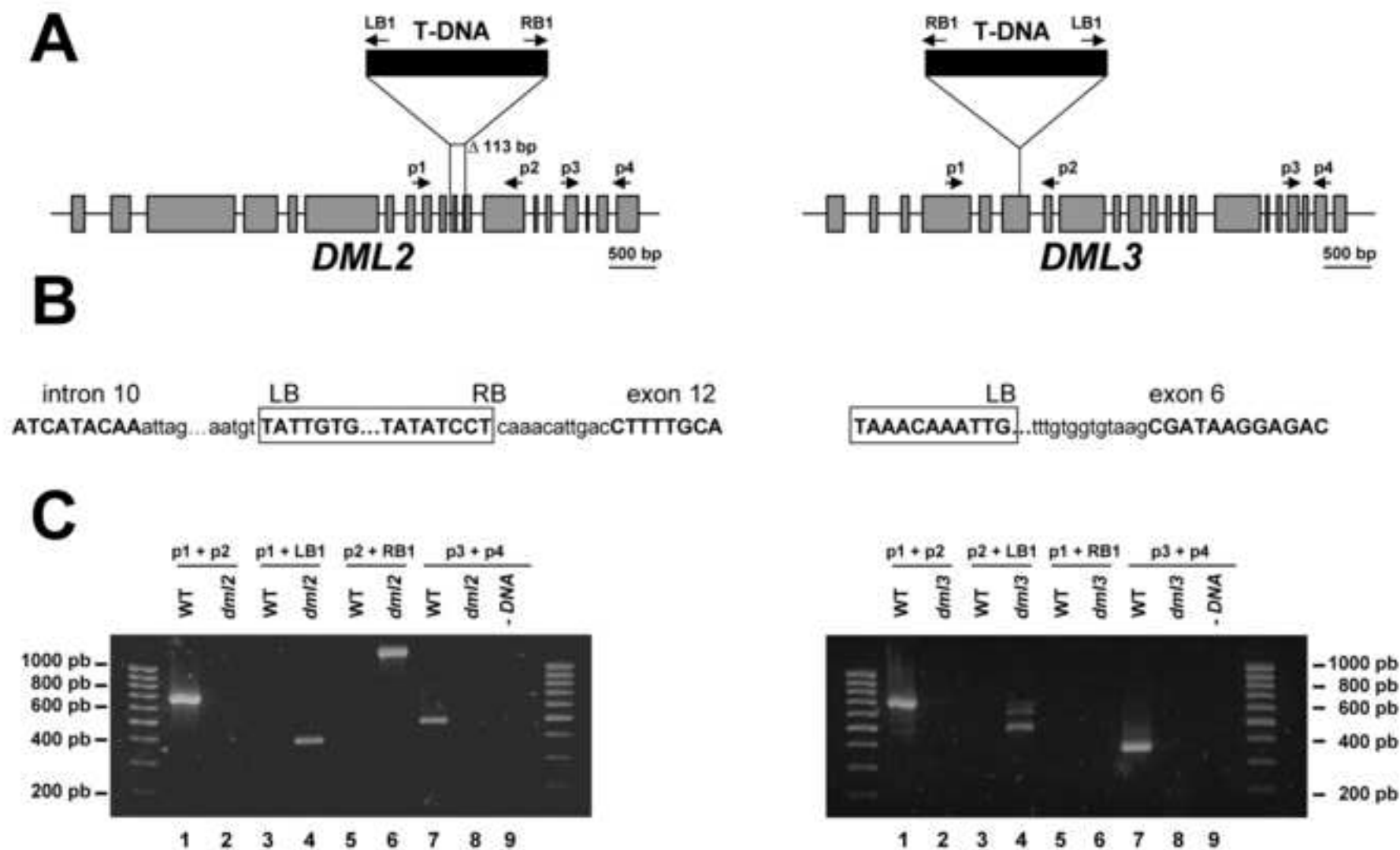
A

LB1 T-DNA RB1

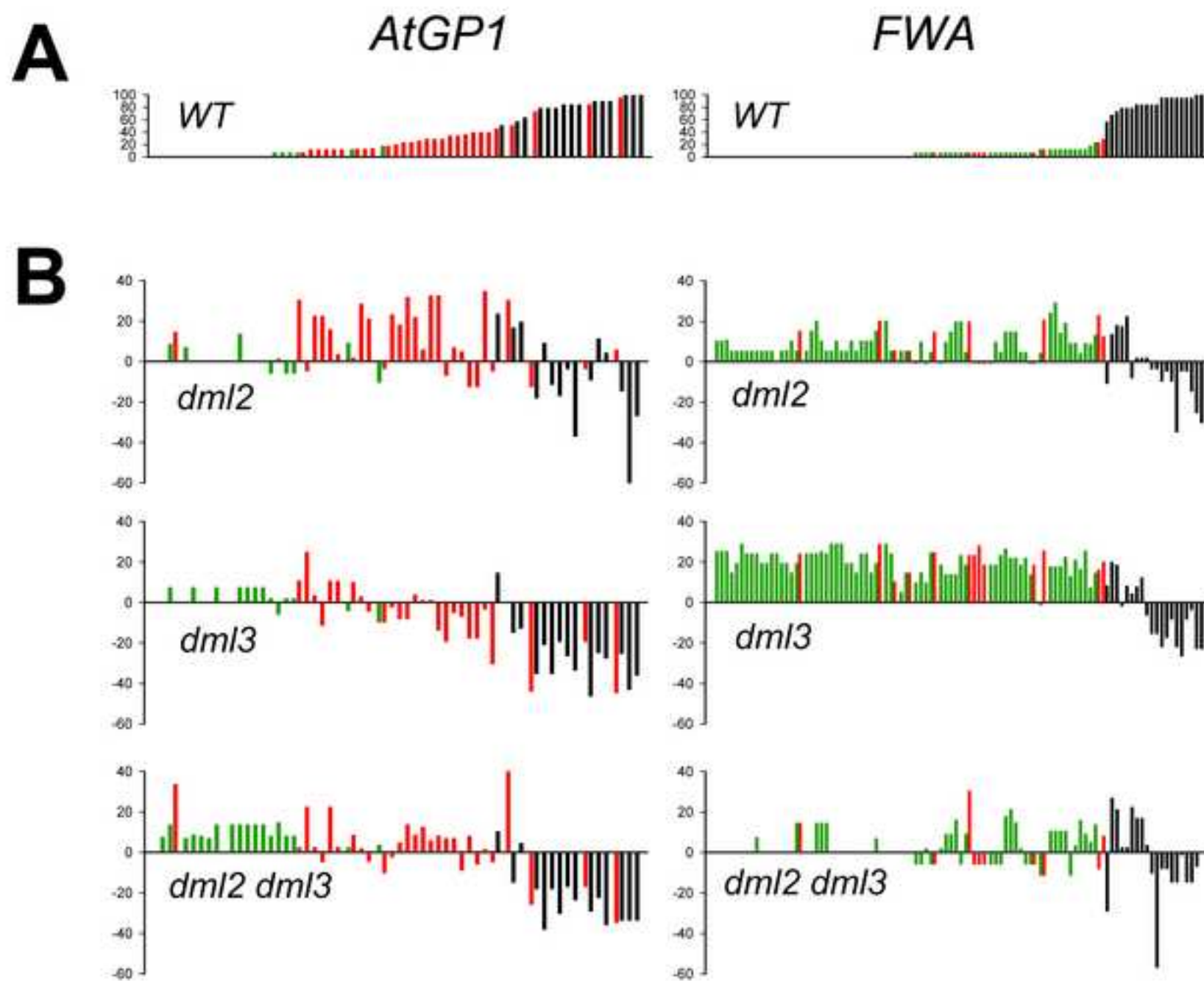Δ 113 bp

p1 p2 p3 p4

DML2

500 bp

RB1 T-DNA LB1

p1 p2 p3 p4

DML3

500 bp

B

intron 10   LB   RB   exon 12

ATCATACAAattag…aatgt TATTGTG…TATATCCT caaacattgacCTTTTGCA

LB   exon 6

TAAACAAATTG…tttgtggtgtaagCGATAAGGAGAC

C

p1 + p2   p1 + LB1   p2 + RB1   p3 + p4

WT dml2 WT dml2 WT dml2 WT dml2 - DNA

1000 pb
800 pb
600 pb

400 pb

200 pb

1 2 3 4 5 6 7 8 9

p1 + p2   p2 + LB1   p1 + RB1   p3 + p4

WT dml3 WT dml3 WT dml3 WT dml3 - DNA

1000 pb
800 pb
600 pb

400 pb

200 pb

1 2 3 4 5 6 7 8 9

Figure 6

Supplementary Table 1

Supplementary Table 2

Click here to download Supplementary material: Suppl Table 2.pdf

Supplementary Figure 1

Supplementary Figure 2