

# Conserved plant genes with similarity to mammalian *de novo* DNA methyltransferases

Xiaofeng Cao<sup>†\*</sup>, Nathan M. Springer<sup>†§¶</sup>, Michael G. Muszynski<sup>¶</sup>, Ronald L. Phillips<sup>¶</sup>, Shawn Kaeppler<sup>§</sup>, and Steven E. Jacobsen<sup>†\*\*</sup>

<sup>†</sup>Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA 90095-1606; <sup>§</sup>Department of Agronomy, University of Wisconsin, Madison, WI 53706; <sup>¶</sup>Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108; and <sup>||</sup>Pioneer Hi-Bred International, Inc., 7300 NW 62nd Avenue, Johnston, IA 50131

Contributed by Ronald L. Phillips, February 25, 2000

**DNA methylation plays a critical role in controlling states of gene activity in most eukaryotic organisms, and it is essential for proper growth and development. Patterns of methylation are established by *de novo* methyltransferases and maintained by maintenance methyltransferase activities. The Dnmt3 family of *de novo* DNA methyltransferases has recently been characterized in animals. Here we describe DNA methyltransferase genes from both *Arabidopsis* and maize that show a high level of sequence similarity to *Dnmt3*, suggesting that they encode plant *de novo* methyltransferases. Relative to all known eukaryotic methyltransferases, these plant proteins contain a novel arrangement of the motifs required for DNA methyltransferase catalytic activity. The N termini of these methyltransferases contain a series of ubiquitin-associated (UBA) domains. UBA domains are found in several ubiquitin pathway proteins and in DNA repair enzymes such as Rad23, and they may be involved in ubiquitin binding. The presence of UBA domains provides a possible link between DNA methylation and ubiquitin/proteasome pathways.**

**M**ethylation of the C5 position of cytosine is the most common covalent modification of DNA in higher plants and animals. This methylation is usually associated with transcriptional gene silencing, or so-called epigenetic gene inactivation. There are several examples of epigenetic silencing in the plant kingdom, including paramutation in maize (1), *PAI* (2) and *SUPERMAN* (3) gene silencing in *Arabidopsis*, and transgene silencing in many plants species (4). Genetic experiments in *Arabidopsis* have shown that proper DNA methylation levels are required for normal development (5, 6). In animal systems, DNA methylation plays a prominent role in allele-specific gene expression that occurs during parental genomic imprinting and X chromosome inactivation (7, 8). Methylation is also important in the regulation of genomic parasites such as transposable elements, retrotransposons, and retroviruses (9, 10).

The presence of 5-methylcytosine in genomic DNA is the result of enzymatic activity of the C5 DNA methyltransferases, which catalyze the transfer of a methyl group from *S*-adenosyl-L-methionine (AdoMet). All known C5 DNA methyltransferases are characterized by the presence of several conserved motifs in the region of the protein involved in catalysis (11). This conservation suggests that cytosine methyltransferases share a common evolutionary history.

There are two major types of DNA methyltransferase activities, maintenance and *de novo*. The methylation of hemimethylated symmetrical sequences (CpG and CpXpG) after DNA replication is maintenance methylation. This results in stable patterns of methylation that are maintained throughout development or, in many cases, between generations. Methylation that occurs at previously unmethylated cytosines is known as *de novo* methylation. For symmetric sites, *de novo* methylation need occur only once, after which methylation can be preserved by maintenance activity. However, for maintenance of methylation at asymmetric sites (cytosines in contexts other than CpG and CpXpG) *de novo* methylation must occur continually.

The eukaryotic DNA methyltransferases can be grouped into at least four distinct classes based on sequence homology and function, the Dnmt1/MET1 class, the Dnmt2 class, the CMT class, and Dnmt3 class.

The Dnmt1/MET1 class enzymes act primarily as maintenance methyltransferases. Dnmt1 was originally isolated from mouse (12) and a Dnmt1 homolog (MET1) was subsequently found in *Arabidopsis* (13). Biochemical studies with mammalian Dnmt1 and a pea MET1-like protein have shown a high level of methyltransferase activity on hemimethylated substrates relative to unmethylated substrates (14–16). Genetic evidence indicates that Dnmt1 and MET1 are the predominant maintenance methyltransferases in mouse and *Arabidopsis*, respectively. Mice harboring loss-of-function *Dnmt1* mutations show only one-third the normal level of genomic methylation, and these mice die after 9 days of development (17). However, nullizygous *Dnmt1* cells retain *de novo* methylation activity (18). In *Arabidopsis*, transgenic plants carrying antisense *MET1* RNA constructs have a 90% decrease in overall DNA methylation levels (5, 19). These antisense *MET1* plants are viable but display a number of specific developmental abnormalities. Two of these abnormalities have been studied in detail and are caused by dense ectopic hypermethylation at both symmetric and asymmetric sites within the floral regulatory genes *SUPERMAN* and *AGAMOUS* (3, 20, 21). Therefore, despite an overall reduction in genomic methylation, the antisense-*MET1* plants retain *de novo* methyltransferase activity.

The Dnmt2 class of DNA methyltransferases, found in mammals (Dnmt2), fission yeast (PMT1), and *Drosophila melanogaster* (DmMT2) does not appear to play a significant role in establishing or maintaining DNA methylation patterns. These proteins do not show significant *in vitro* activity, and loss-of-function *Dnmt2* mutations do not show any reduction in the amount of overall DNA methylation, or a reduction in *de novo* methylation (22–26). As neither *Drosophila* nor fission yeast contain detectable amounts of cytosine methylation in their genomes, the function of these Dnmt2 methyltransferase-like proteins is unclear.

The chromomethylases (CMTs) represent a class of DNA methyltransferases that have so far been found only in plants. The distinguishing feature of CMTs is the presence of a chro-

Abbreviations: CMT, chromomethylase; ES cells, embryonic stem cells; UBA, ubiquitin-associated; RACE, rapid amplification of cDNA ends; BAC, bacterial artificial chromosome; EST, expressed sequence tag.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF240695 and AF242320).

\*X.C. and N.M.S. contributed equally to this work.

\*\*To whom reprint requests should be addressed: E-mail: jacobsen@ucla.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

modomain embedded between catalytic motifs I and IV (27, 28). A loss-of-function mutation in a maize *CMT*-like gene, *ZMET2*, specifically reduces CpXpG methylation (Charles Papa, N.S., and S.K., unpublished results). This observation provides an explanation for the fact that plants contain a high level of CpXpG methylation in their genomes relative to animals (29), and it suggests that the CMTs may act as a specialized type of plant specific maintenance methyltransferase.

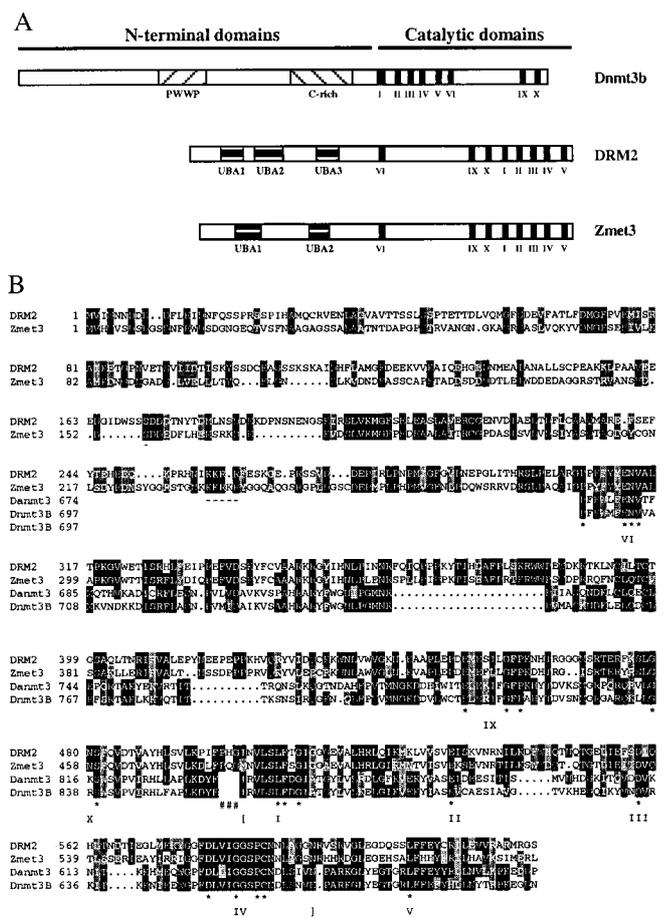
Recently the Dnmt3 class of DNA methyltransferases was identified in mouse, human, and zebra fish (16, 30). Several lines of evidence suggest that the Dnmt3 proteins act as *de novo* methyltransferases. Recombinant Dnmt3a and Dnmt3b enzymes from mouse displayed *de novo* activity when tested on unmethylated DNA templates *in vitro* (16). Furthermore, expression of Dnmt3a in *Drosophila melanogaster* caused *de novo* methylation of its normally unmethylated genome (31). When both *Dnmt3a* and *Dnmt3b* were inactivated by gene targeting, the resulting embryonic stem (ES) cells and early embryos lacked *de novo* methylation activity (32). Mutations in the human *Dnmt3b* gene were found to be the cause of ICF syndrome (immunodeficiency, centromeric instability, and facial anomalies syndrome) (32–34). ICF patients show decreased CpG methylation within satellite repeats of the centromeres of chromosomes 1, 9, and 16, coupled with centromeric instability. This suggests that Dnmt3b may be critical for methylation of some types of repetitive DNA.

In this report, we describe genes in maize and *Arabidopsis* that display a high degree of sequence similarity to *Dnmt3*. They display a distinct arrangement of the DNA methyltransferase catalytic motifs, and they contain a series of ubiquitin-associated (UBA) domains in their N termini.

## Materials and Methods

**cDNA Cloning and Rapid Amplification of cDNA Ends (RACE) Analysis.** *Arabidopsis* genomic DNA was PCR amplified to generate a hybridization probe corresponding to the TAMU bacterial artificial chromosome (BAC) survey sequence B62154, which was used to screen an *Arabidopsis* young seedling cDNA library (kindly provided by the Arabidopsis Biological Resource Center, Columbus, OH). Hybridization procedures were as previously described (35) but the washes were done under moderately stringent conditions;  $0.1 \times$  SSPE/0.5% SDS for 15 min at 55°C ( $1 \times$  SSPE = 0.18 M NaCl/10 mM phosphate, pH 7.4/1 mM EDTA). After the initial identification of the maize *Dnmt3*-like expressed sequence tag (EST) sequence, RACE PCR was performed on Marathon cDNA (CLONTECH), using Advantage2 DNA polymerase (CLONTECH) to clone the full-length *ZMET3* cDNA sequence. One-week-old Mo17 seedling RNA was used to construct the cDNA. The primers used for RACE were Dmt3F1 (5'-ATCCGTATGCCAAGC-CTGTGGAGAGC-3'), Dmt3F2 (GATGGACTTGACG-GCGTGTAAAGATCC-3'), Zmet3RACE1 (5'-GGAG-GAAGTGGCAGAGGAGGAGG-3'), and Zmet3RACE2 (5'-GGAGGCACTGGACGGCGTGG-3').

**Phylogenetic Analysis.** Alignments were performed with a region of each methyltransferase starting at the conserved catalytic motif I and ending at motif IV (brackets in Fig. 1B) by using CLUSTALX 1.8 and default parameters. Since the chromodomain present in the CMT proteins represents a large insertion relative to all of the other methyltransferases, we introduced a deletion in this region so that all of the sequences aligned unambiguously (total length of alignment, 81 amino acids). Fig. 3 shows a bootstrap tree based on amino acids (unordered character states). Bootstrapping was done by using heuristic searches in PAUP\* (4.0b2 written by David L. Swofford, Smithsonian Institution, Washington, DC) with 1000 replicates and 10 random additions per replicate. Similar trees were also obtained with



**Fig. 1.** (A) Schematic diagram of the domain structures of Dnmt3b, DRM2, and Zmet3. Figure is drawn to scale. Shaded boxes show the different motifs present in these proteins, including the PWWP and cysteine rich (C-rich) motifs present in Dnmt3b and the UBA domains present in DRM2 and Zmet3. Roman numerals denote the motifs of the methyltransferase catalytic domains. (B) Alignment of DRM2 of *Arabidopsis thaliana* with Zmet3 of *Zea mays* and the methyltransferase catalytic domains of mouse Dnmt3b (GenBank accession no. AF068628) and *Danio rerio* Zmt3 (Danmt3; accession no. AF135438). Alignments were done in CLUSTALX 1.8 using default parameters and shaded in MACBOXSHADE 1.0.8 (Michael D. Baron, Institute for Animal Health, Surrey, U.K.). Identical residues are shown with a black background and similar residues with a gray background. Dashes show putative nuclear localization sequences that are conserved in the plant proteins. Pound symbols (#) show the point of rearrangement of the plant proteins relative to the animal proteins. Here the numbering of the animal methyltransferases begins at amino acid 581 for Dnmt3b and 558 for Danmt3. Conserved catalytic motifs I–VI and IX–X are marked. Asterisks denote conserved amino acids present in each motif (11). Brackets show the exact region used in the alignments that were used to produce the tree shown in Fig. 3.

alignments in which the chromodomain sequences were not deleted (S.E.J., unpublished results).

**Mapping.** To map *DRM2*, the 5' 970 base pairs of the *DRM2* cDNA was PCR amplified, labeled with  $^{32}$ P, and used to probe an ordered Institut für Genbiologische Forschung Berlin BAC filter (kindly provided by the Arabidopsis Biological Resource Center). *DRM1* was mapped in a similar fashion by using a probe corresponding to the genomic sequence contained in GenBank accession no. B62154.

**RNA Blot Analysis.** Total RNA from *Arabidopsis* leaves or inflorescences from 4-week-old plants or whole roots grown in culture

was isolated with Tri Reagent (Molecular Research Center, Cincinnati). Thirteen micrograms of total RNA was subjected to electrophoresis on a formaldehyde-containing 0.8% agarose gel, blotted onto a Hybond-N<sup>+</sup> (Amersham) membrane, and probed with a full-length *DRM2* cDNA. Hybridization procedures were as previously described (35) and washes were at high stringency: 0.1× SSPE/0.5% SDS for 15 min at 65°C.

## Results

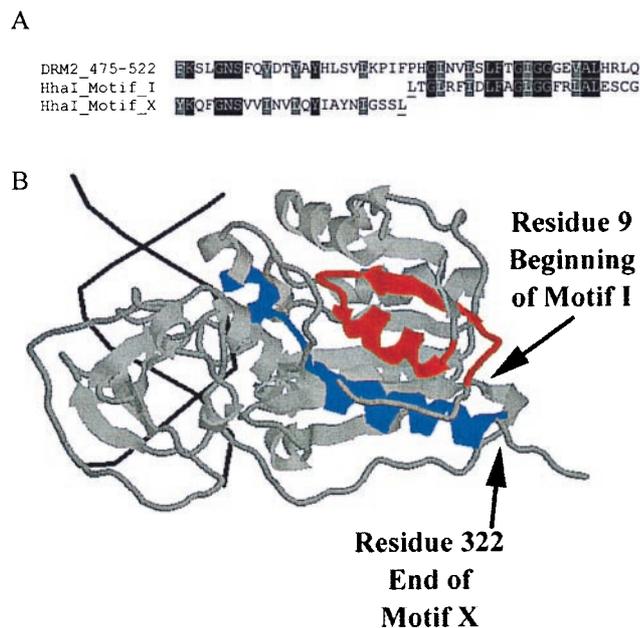
We searched *Arabidopsis* and maize databases for genes similar to *Dnmt3*. In *Arabidopsis*, a BAC end sequence (GenBank accession no. B62154) showing similarity to the catalytic domain was found. A probe derived from this BAC end sequence was used to screen a cDNA library at moderate stringency. No cDNA clones corresponding to this sequence were found. However, several clones corresponding to a closely related sequence were isolated. Three of the cDNA clones appeared to be full length, as they contained an in-frame stop codon in the 5' untranslated region (GenBank accession no. AF240695). In maize, mouse *Dnmt3* was used as a query against an EST database provided by Pioneer Hi-Bred International. An EST contig containing the 3' 612 base pairs of a maize cDNA was identified. RACE-PCR was used to clone and identify the 5' sequence of the mRNA (GenBank accession no. AF242320).

The predicted proteins from the *Arabidopsis* and maize cDNA sequences are similar to each other along their entire length (Fig. 1). They exhibit 28% amino acid identity in the N-terminal domains and 66% identity in the C-terminal catalytic domains. The catalytic domains of these proteins are most similar (an average of 28% amino acid identity) to the catalytic domains of the animal *Dnmt3* proteins (Fig. 1B). TBLASTN searches of the *Arabidopsis* and maize sequences against the National Center for Biotechnology Information EST database showed that the most closely related mammalian EST sequences are those corresponding to the *Dnmt3* genes of mouse and human. We did not detect significant similarity between the N-terminal domains of the plant and animal proteins by using BLASTP or CLUSTALX, and both *DRM2* and *Zmet3* lack the PWWP and cysteine-rich motifs present in the *Dnmt3* methyltransferases (30, 33).

Relative to known eukaryotic methyltransferases, the plant proteins show a novel arrangement of the conserved catalytic motifs. Most methyltransferases, including *Dnmt3*, contain motifs I, II, III, IV, V, VI, IX, X, and Y from the N terminus to the C terminus of the protein (motifs VII and VIII are not highly conserved and are difficult to distinguish in many methyltransferases). However, both the *Arabidopsis* and maize sequences display an altered arrangement of these motifs, VI, IX, X, I, II, III, IV, V (Fig. 1A). The location of the rearrangement can be pinpointed to a region of several amino acids between motifs X and I (Fig. 1B). Because of this rearrangement, the *Arabidopsis* proteins have been named the domains rearranged methyltransferases (DRMs). The first observed homolog from BAC sequence B62154 is named *DRM1* and the sequence reported in Fig. 1 is named *DRM2*. The maize sequence has been named *ZMET3*, as this represents the third class of methyltransferase to have been isolated from maize.

A search of other plant ESTs by using *DRM* and *ZMET3* as queries revealed the presence of a soybean 3' cDNA sequence (accession no. A1736568) which displays a high level of identity to both the *Arabidopsis* and maize sequences. This partial sequence predicts a polypeptide that encodes the methyltransferase catalytic motifs IX, X, I, II, III, IV, and V, which is the same order seen in both *Arabidopsis* and maize.

We sought to determine the possible effect of the motif rearrangement seen in *DRM2/Zmet3* on the protein structure, relative to known structures for DNA methyltransferases. Specifically, we looked at the possible implications of the juxtaposition of motifs X and I in the primary sequence. The solved



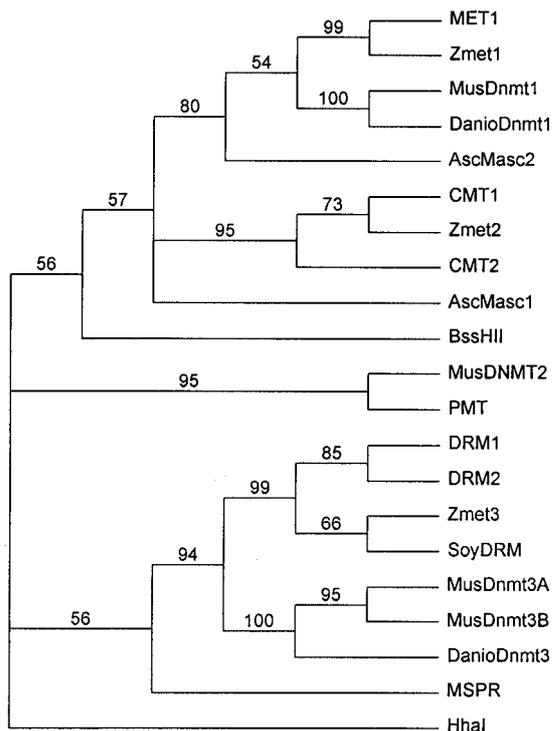
**Fig. 2.** (A) An alignment of motifs X and I of *HhaI* with the sequence of *DRM2* spanning the point of motif rearrangement. *HhaI* residues underlined are the last residue of motif X and the first residue of motif I and correspond the positions in *B* marked by arrows. Alignments were done in CLUSTALX and shaded in BOXSHADE. (B) RASMOL-generated image (Roger Sayle, Glaxo Research and Development, Greenford, Middlesex, U.K.) of Protein Database ID 5MHT (ternary structure of *HhaI* methyltransferase with hemimethylated DNA and S-adenosylhomocysteine) (36, 37). DNA helix is shown in black, the majority of *HhaI* is shaded gray, motif I is red, and motif X is blue. The colored regions correspond to those shown in the alignment in A, which are residues 9–32 for motif I and residues 299–322 for motif X. Arrows show the last residue of motif X and first residue of motif I.

structure of a prokaryotic *HhaI* methyltransferase (36, 37) is shown in Fig. 2 with motifs I and X highlighted. These motifs lie parallel to one another in the tertiary structure and are physically associated. Furthermore, the C terminus of *HhaI* motif X (residue 322, marked with arrow in Fig. 2B) and the N terminus of *HhaI* motif I (residue 9, marked with arrow in Fig. 2B) are very close together in the three-dimensional structure. Because these amino acids are directly adjacent to one another in the primary sequence of *DRM2* and *Zmet3*, it is conceivable that, despite the motif rearrangement, the overall fold of the plant proteins is similar to that of *HhaI*.

To examine the relationships between the plant methyltransferases described here and other known methyltransferases, we performed alignments with the conserved catalytic motifs I–IV (Fig. 1B), which were then used to generate a phylogenetic tree (Fig. 3). Representatives of four classes of animal and plant DNA methyltransferases were used in the alignments, including enzymes of the *Dnmt1/MET1* maintenance methyltransferase class, as well as the *Dnmt2*, *CMT*, and *Dnmt3* classes. *DRM1*, *DRM2*, *Zmet3*, and the related soybean EST sequence group with a 94% bootstrap value to the clade containing the *de novo* methyltransferase proteins *Dnmt3a* and *Dnmt3b* from mammals and zebrafish (*Danio rerio*).

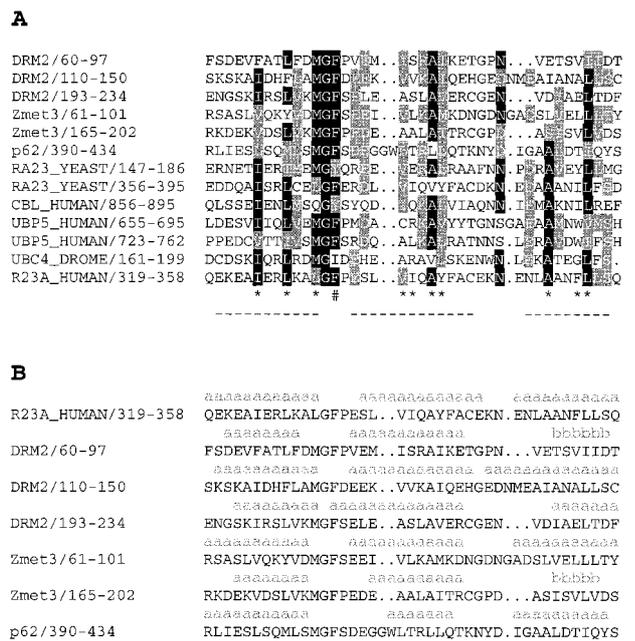
Consistent with their putative function as DNA methyltransferases, both the *DRM2* and *Zmet3* proteins are predicted by PSORT (38) to reside in the nucleus and contain conserved nuclear targeting sequences of the simian virus 40 large T antigen type. These lie in the N terminus of the protein (underlined in Fig. 1B).

To determine whether *DRM2* or *Zmet3* contains any recognizable domains in their N termini, we tested the protein



**Fig. 3.** Phylogenetic relationships of DNA methyltransferases. Tree shows nodes with >50% bootstrap support. Abbreviations and accession numbers, in parentheses, are as follows: MET1, *Arabidopsis thaliana* maintenance methyltransferase MET1 (P34881); Zmet1, *Zea mays* MET1-like protein (AF0063403); MusDnmt1, mouse maintenance methyltransferase Dnmt1 (P13864); DanioDnmt1, zebrafish Dnmt-like protein (AF097875); AscMasc2, *Ascobolus immersus* Masc2 (AF030976); CMT1, *Arabidopsis thaliana* chromomethylase 1 (AF039372); Zmet2, *Zea mays* CMT-like protein (AF243043); CMT2, *Arabidopsis thaliana* chromomethylase 2 (AL021711); AscMasc1, *Ascobolus immersus* Masc1 (AF025475); BssHII, *Bacillus stearothermophilus* BssHII methylase (AF020002); MusDNMT2, mouse DNMT2 (AF045889); PMT, *Schizosaccharomyces pombe* PMTp1 (P40999); DRM1, *Arabidopsis thaliana* DRM1 (B62154); DRM2, *Arabidopsis thaliana* DRM2; Zmet3, *Zea mays* Zmet3; SoyDRM, soybean EST clone with similarity to DRM and Zmet3 (A1736568); MusDnmt3A, mouse *de novo* methyltransferase Dnmt3a (AF068625); MusDnmt3B, mouse *de novo* methyltransferase Dnmt3b (AF068628); DanioDnmt3, zebra fish Dnmt3-like protein (AF135438); MSPR, *Bacillus subtilis* bacteriophage SPR methyltransferase (P00476); and HhaI, bacterial (*Haemophilus haemolyticus*) HhaI methyltransferase used as an outgroup (P05102).

sequences on both the PFAM and SMART (39) protein prediction web servers. Both programs predicted a series of UBA domains in DRM2 (three separate domains) and Zmet3 (two domains) (Fig. 1A). UBA domains are found in several ubiquitination pathway enzymes, in proteins involved in nucleotide excision repair (such as Rad23), and in some protein kinases (40). The NMR structure of a UBA domain from the human homolog of Rad23 (HHR23A) shows that it folds into a compact three-helix bundle (41). Fig. 4A shows an alignment of the DRM2 and Zmet3 UBA domains with those of several other proteins. They contain several conserved residues that are thought to participate in the formation of the hydrophobic core as well as the sharp turn in the loop between the first and second alpha helices. To determine whether the DRM2 and Zmet3 UBA domains are likely to have a structure similar to HHR23A, we tested the sequences in a nearest-neighbor secondary structure prediction program, NNSSP (Fig. 4B). This algorithm predicted a secondary structure for HHR23A that is largely similar to the known NMR structure, and it predicted that two

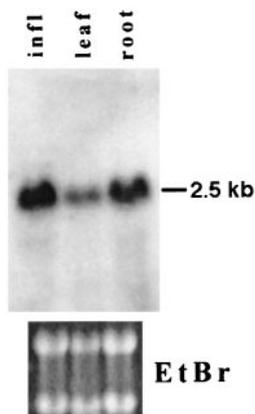


**Fig. 4.** (A) CLUSTALX alignment of the UBA domains predicted by PFAM (<http://pfam.wustl.edu/>) of DRM2 and Zmet3 with those of human p62 (accession no. U46751), yeast Rad23 (accession no. S66117), human Cbl (accession no. A43817), human ubiquitin C-terminal hydrolase 5 (accession no. P45974), *Drosophila melanogaster* ubiquitin-conjugating enzyme E2 (accession no. P52486), and human HHR23A (accession no. S44443). Below the alignment is the known structure of the second UBA domain of HHR23A, showing the three  $\alpha$ -helices (---) (41) (Thomas Mueller and Juli Feigon, personal communication). \*, Residues making important contributions to the hydrophobic core. #, Phe residue connecting the loop between helices  $\alpha$ 1 and  $\alpha$ 2 to the C terminus of helix  $\alpha$ 3. (B) Results of nearest-neighbor secondary structure prediction (NNSSP) on the Baylor College of Medicine Protein Web server (<http://dot.imgen.bcm.tmc.edu:9331/pssp/prediction/pssp.html>) (54), which predicts the presence of  $\alpha$ -helices (a) or  $\beta$ -sheets (b). Gaps were introduced in the sequences as in the alignments in A.

of the DRM2 UBA domains and one of the Zmet3 UBA domains are likely to have a structure similar to HHR23A (Fig. 4B). The remaining two UBA domains, DRM2 amino acids 60–97 and Zmet3 amino acids 165–202, were predicted to contain a  $\beta$ -sheet in the place of the third  $\alpha$ -helix predicted in HHR23A.

We assayed the complexity of the gene families encoding DRM and Zmet3 type proteins by using Southern blot analysis and BLAST searching. In both *Arabidopsis* and maize, Southern blot analysis of genomic DNA detected several hybridizing bands suggesting the presence of small gene families (data not shown). TBLASTN searches with the full-length DRM2 protein identified three additional sequences with significant similarity within the available *Arabidopsis* genome sequences (85% of the genome was sequenced at the time of this writing). These are accession nos. AB022216, AC012375, and T22J18. All three of these sequences reside in sequenced regions of the genome and appear to encode DRM pseudogenes (S.E.J., unpublished observation). A TBLASTN search of GenBank using the full-length Zmet3 sequence detected a maize EST sequence encoding a related protein (accession no. A1947339). This sequence lacks the highly conserved PC site in motif IV of the catalytic domain, suggesting that it may be a pseudogene.

The DRM genes were mapped by hybridizing them to a filter containing ordered *Arabidopsis* BAC clones (42). A DRM2 probe hybridized strongly to six clones (F6K20, F17H18, F17N1, F19B7, F27E1, and F19H5), all of which map to an overlapping



**Fig. 5.** RNA blot analysis showing the size and abundance of *DRM2* in different tissues. RNA from roots, leaves, or inflorescences was blotted and hybridized to a full-length *DRM2* clone. Ethidium bromide staining of the ribosomal RNA bands is shown below as a loading control.

position on the top of chromosome V between markers mi174 and mi322. *DRM1* hybridized strongly to a different set of overlapping clones (F17C20, F25G21, F7P21, F1K12, F12A2, F27F12, F14F13, F5A22, F19D5, F9D20, F4K15, F3H2, F111, and F8M21) which map to a nearby region of chromosome V. From the physical map the estimated distance between *DRM1* and *DRM2* is about 230 kb, suggesting a map distance of approximately 1 centimorgan (details can be found at <http://genome-www3.stanford.edu/cgi-bin/AtDB/Pmap?contig=F24C6-T7-F24L21-Sp6&clone=F28E9>). These map positions do not correspond to known methylation mutants in *Arabidopsis*. The *DRM* probes also detected several weakly hybridizing clones, all of which mapped to two different regions on chromosome I. These groups consist of F6F2, F7M24, F24I22, F24L16, F12N13, F28N2, and F4P4, which map to an overlapping region near the *PAI3* marker, and F27D22, F11L21, and F11P6, which map to an overlapping region near the *PAI1* marker. These two regions correspond exactly to the pseudogenes mentioned above, AC012375 and T22J18, respectively. Thus, all hybridizing clones can be accounted for by DNA sequences present in the database.

RNA blot analysis was used to detect expression of *DRM2* in different tissues. A 2.5-kb transcript was easily detected on blots of total RNA from roots, leaves, or inflorescences (Fig. 5). This finding suggests that *DRM2* is expressed in most tissues. Probing of the same blot with *DRM1* did not detect any message, suggesting that *DRM1* is expressed at a lower level than *DRM2*.

## Discussion

**Identification of a Distinct Family of DNA Methyltransferases.** We describe a distinct type of putative C5 DNA methyltransferase conserved in *Arabidopsis* and maize. *DRM2* and *Zmet3* display the highest level of similarity to each other in their C-terminal regions, which contain the DNA methyltransferase catalytic domains. The amino acid identity in these regions is 66%, whereas it is only 28% in the N-terminal regions. This observation suggests selection for the conservation of methyltransferase function, supporting the hypothesis that *Zmet3* and *DRM2* are functional DNA methyltransferases. This hypothesis is further supported by the observation that both *Zmet3* and *DRM2* contain several amino acids thought to be critical for the function of cytosine methyltransferases, including the Phe-Xaa-Gly-Xaa-Gly residues present in motif I involved in S-adenosylmethionine binding, the invariant Pro-Cys dipeptide of the catalytic site at motif IV, and the Glu-Asn-Val residues at motif VI, which interact with the target cytosine (11, 36, 37, 43, 44).

*DRM2* and *Zmet3* are characterized by a rearrangement of

the catalytic methyltransferase motifs. The presence of the same arrangement in *Arabidopsis*, maize, and soybean indicates that the permutation occurred before the divergence of monocots and dicots. There are at least two processes that could have given rise to structure seen in these genes. The first is a transposition event resulting in a swap between motifs I–V and motifs VI–X. A second possibility is gene duplication followed by deletions to remove motifs I–V of the first gene, the intervening sequence between the two genes, and motifs VI–X of the second gene.

We know of no other examples of eukaryotic genes displaying rearranged DNA methyltransferase motifs. However, there are examples of domain permutations within the bacterial C5 DNA methyltransferases (45). One example is *BssHIII* in which motifs IX and X precede motif I–VIII (46). This example is similar to *DRM2* and *Zmet3* in the sense that motifs X and I are juxtaposed. In a second case, *AquI*, motifs IX and X are located in a separately encoded subunit (47). The fact that these prokaryotic enzymes maintain DNA methyltransferase activity shows that the usual arrangement of DNA methyltransferase motifs is not necessary for function. Furthermore, the specific permutation present in *DRM2* and *Zmet3*, juxtaposition of motifs I and X in the primary sequence, may have little overall effect on the folding or function of the methyltransferase catalytic domains, because these motifs are found adjacent to each other in the tertiary structures of *HhaI* (36, 37) (Fig. 2) and *HaeIII* (44).

*DRM2* and *Zmet3* contain a series of UBA domains in their N termini. The function of the UBA domain is presently unclear. UBA domain-containing proteins show a variety of associations with the ubiquitin pathway. Some are components of the ubiquitination machinery such as ubiquitin C-terminal hydrolases, ubiquitin-conjugating enzymes (E2), and ubiquitin protein ligases (E3) (40). Others are involved in DNA repair, such as Rad23, which is itself ubiquitinated and degraded during the transition between the G<sub>1</sub> and S phases of the cell cycle (48). The UBA-containing protooncprotein Cbl shows reversible ubiquitination that is associated with its cycling between the plasma membrane and cytosolic fractions of the cell (49, 50). Cbl also acts as ubiquitin ligase for receptor protein-tyrosine kinases (50, 51).

The C-terminal UBA domain present in the human Rad23 homolog HHR23A was shown to bind *in vitro* to the HIV-1 protein Vpr (52). Furthermore, an 80-amino acid region of the p62 protein, which consists largely of a UBA domain (39), was shown to display noncovalent binding to ubiquitin (53). These results suggest that UBA domains may serve as protein interaction interfaces, and that one specific function of the UBA domain is to bind ubiquitin.

The presence of UBA domains in *DRM2* and *Zmet3* suggests a link between DNA methylation and ubiquitin/proteasome pathways. One possibility is that, like Rad23, *DRM2/Zmet3* proteins vary throughout the cell cycle through ubiquitin-mediated protein degradation. A second possibility is that ubiquitination alters the cellular localization of the *DRM2/Zmet3* proteins in response to external signals, the cell cycle, or transposon or retroviral activity. As UBA domains are not found in other classes of methyltransferases or in the mammalian or fish Dnmt3 proteins (S.E.J., unpublished observation), this possible association with the ubiquitin pathway may be restricted to the Dnmt3-like methyltransferases of plants.

**Origins of Eukaryotic DNA Methyltransferases.** Our phylogenetic analysis suggests that the Dnmt3/*DRM2/Zmet3* enzymes form a distinct class of proteins that are closer to each other than they are to other types of methyltransferases, including the Dnmt1/MET1 class, the CMT class, and the Dnmt2 class. This observation suggests that these different types of methyltransferases formed early in eukaryotic evolution, before the divergence of plants and animals, and may therefore share common functions.

One possible scenario is that the different classes of methyltransferases could have originated from separate prokaryotic lineages. In particular, Xie *et al.* (30) have suggested that the *Dnmt3* class of genes may have evolved from an ancestor related to the *Bacillus subtilis* bacteriophage SPR methyltransferase, whereas the Dnmt1 class of DNA methyltransferases may have evolved from a separate prokaryotic ancestor. This conclusion is supported by the observation that SPR methyltransferase groups more closely with the Dnmt3/DRM2/Zmet3 sequences than with other eukaryotic methyltransferases (Fig. 3).

**De Novo Methylation in Plants.** Given the relationship of the plant genes to *Dnmt3*, we propose that DRM2 and Zmet3 act as plant *de novo* methyltransferases. Several well-characterized examples of *de novo* methylation occur in plants. One case is the extensive methylation at the *SUPERMAN* locus in the *Arabidopsis clark kent* mutants and in plants containing antisense-*MET1* constructs (3). *SUPERMAN* remains methylated and silenced in most tissues of the plant, indicating a requirement for *de novo* methyltransferase activity in somatic cells of the shoot. The expression profile of *DRM2* fits this requirement, since *DRM2* RNA is found in all of the major tissues of the plant. The methylation at *SUPERMAN* is very dense (over 50% of the cytosines are methylated), and at mostly asymmetric sites. How-

ever, different hypermethylated *superman* alleles show largely similar patterns of methylation (21). It is difficult to envision a mechanism by which this strange pattern of methylation is inherited. One hypothesis is that *SUPERMAN* is continually methylated by a *de novo* methyltransferase system in response to a "seed" of preexisting methylation. In this way, the *DRM* genes might serve as maintenance methyltransferases for genes containing asymmetric methylation. Other examples of *de novo* methylation in plants include paramutation at *R*, transposable element inactivation, and transgene silencing. It seems possible that Zmet3/DRM2-type enzymes could play a major role in these phenomena as well. Biochemical and genetic studies should allow determination of the function of these plant DNA methyltransferases.

We thank the Arabidopsis Biological Resource Center for materials, and Mike Frohlich, Thomas Mueller, and Juli Feigon for helpful discussions. This work was supported by National Institutes of Health Grant GM60398 and Jonsson Cancer Center Foundation/UCLA Seed Grant ACS #IRG 78-001-21 to S.E.J. Maize research was supported by Pioneer Hi-Bred International, Inc., and the University of Wisconsin Graduate School, and this was carried out jointly in the laboratories of S.K. and R.L.P. Questions and material requests for maize work should be directed to S.K. N.S. was supported by a U.S. Department of Agriculture National Needs Fellowship (98-38420-5832).

- Martienssen, R. A. & Richards, E. J. (1995) *Curr. Opin. Genet. Dev.* **5**, 234–242.
- Bender, J. & Fink, G. R. (1995) *Cell* **83**, 725–734.
- Jacobsen, S. E. & Meyerowitz, E. M. (1997) *Science* **277**, 1100–1103.
- Matzke, M. A. & Matzke, A. J. (1998) *Cell. Mol. Life Sci.* **54**, 94–103.
- Finnegan, E. J., Peacock, W. J. & Dennis, E. S. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 8449–8454.
- Richards, E. J. (1997) *Trends Genet.* **13**, 319–323.
- Li, E., Beard, C. & Jaenisch, R. (1993) *Nature (London)* **366**, 362–365.
- Beard, C., Li, E. & Jaenisch, R. (1995) *Genes Dev.* **9**, 2325–2334.
- Walsh, C. P. & Bestor, T. H. (1999) *Genes Dev.* **13**, 26–34.
- Yoder, J. A., Walsh, C. P. & Bestor, T. H. (1997) *Trends Genet.* **13**, 335–340.
- Cheng, X. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24**, 293–318.
- Bestor, T., Laudano, A., Mattaliano, R. & Ingram, V. (1988) *J. Mol. Biol.* **203**, 971–983.
- Finnegan, E. J. & Dennis, E. S. (1993) *Nucleic Acids Res.* **21**, 2383–2388.
- Pradhan, S., Cummings, M., Roberts, R. J. & Adams, R. L. (1998) *Nucleic Acids Res.* **26**, 1214–1222.
- Bestor, T. H. (1992) *EMBO J.* **11**, 2611–2617.
- Okano, M., Xie, S. & Li, E. (1998) *Nat. Genet.* **19**, 219–220.
- Li, E., Bestor, T. H. & Jaenisch, R. (1992) *Cell* **69**, 915–926.
- Lei, H., Oh, S. P., Okano, M., Juttermann, R., Goss, K. A., Jaenisch, R. & Li, E. (1996) *Development (Cambridge, U.K.)* **122**, 3195–3205.
- Ronemus, M. J., Galbiati, M., Ticknor, C., Chen, J. & Dellaporta, S. L. (1996) *Science* **273**, 654–657.
- Jacobsen, S. E. (1999) *Curr. Biol.* **9**, R617–R619.
- Jacobsen, S. E., Sakai, H., Finnegan, E. J., Cao, X. & Meyerowitz, E. M. (2000) *Curr. Biol.* **10**, 179–186.
- Pinarbasi, E., Elliott, J. & Hornby, D. P. (1996) *J. Mol. Biol.* **257**, 804–813.
- Wilkinson, C. R., Bartlett, R., Nurse, P. & Bird, A. P. (1995) *Nucleic Acids Res.* **23**, 203–210.
- Okano, M., Xie, S. & Li, E. (1998) *Nucleic Acids Res.* **26**, 2536–2540.
- Yoder, J. A. & Bestor, T. H. (1998) *Hum. Mol. Genet.* **7**, 279–284.
- Hung, M. S., Karthikeyan, N., Huang, B., Koo, H. C., Kiger, J. & Shen, C. J. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 11940–11945.
- Rose, T. M., Schultz, E. R., Henikoff, J. G., Pietrokovski, S., McCallum, C. M. & Henikoff, S. (1998) *Nucleic Acids Res.* **26**, 1628–1635.
- Henikoff, S. & Comai, L. (1998) *Genetics* **149**, 307–318.
- Gruenbaum, Y., Naveh-Manly, T., Cedar, H. & Razin, A. (1981) *Nature (London)* **292**, 860–862.
- Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W. W., Okumura, K. & Li, E. (1999) *Gene* **236**, 87–95.
- Lyko, F., Ramsahoye, B. H., Kashevsky, H., Tudor, M., Mastrangelo, M. A., Orr-Weaver, T. L. & Jaenisch, R. (1999) *Nat. Genet.* **23**, 363–366.
- Okano, M., Bell, D. W., Haber, D. A. & Li, E. (1999) *Cell* **99**, 247–257.
- Xu, G. L., Bestor, T. H., Bourc'his, D., Hsieh, C. L., Tommerup, N., Bugge, M., Hulten, M., Qu, X. Y., Russo, J. J. & Viegas-Pequignot, E. (1999) *Nature (London)* **402**, 187–191.
- Hansen, R. S., Wijmenga, C., Luo, P., Stanek, A. M., Canfield, T. K., Weemaes, C. M. & Gartler, S. M. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 14412–14417.
- Chang, C., Bowman, J. L., DeJohn, A. W., Lander, E. S. & Meyerowitz, E. M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6856–6860.
- Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W. & Roberts, R. J. (1993) *Cell* **74**, 299–307.
- Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. (1994) *Cell* **76**, 357–369.
- Nakai, K. & Kanehisa, M. (1992) *Genomics* **14**, 897–911.
- Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 5857–5864.
- Hofmann, K. & Bucher, P. (1996) *Trends Biochem. Sci.* **21**, 172–173.
- Dieckmann, T., Withers-Ward, E. S., Jarosinski, M. A., Liu, C. F., Chen, I. S. & Feigon, J. (1998) *Nat. Struct. Biol.* **5**, 1042–1047.
- Mozo, T., Dewar, K., Dunn, P., Ecker, J. R., Fischer, S., Kloska, S., Lehrach, H., Marra, M., Martienssen, R., Meier-Ewert, S. & Altmann, T. (1999) *Nat. Genet.* **22**, 271–275.
- Posfai, J., Bhagwat, A. S., Posfai, G. & Roberts, R. J. (1989) *Nucleic Acids Res.* **17**, 2421–2435.
- Reinisch, K. M., Chen, L., Verdine, G. L. & Lipscomb, W. N. (1995) *Cell* **82**, 143–153.
- Jeltsch, A. (1999) *J. Mol. Evol.* **49**, 161–164.
- Xu, S., Xiao, J., Posfai, J., Maunus, R. & Benner, J., 2nd (1997) *Nucleic Acids Res.* **25**, 3991–3994.
- Karremann, C. & de Waard, A. (1990) *J. Bacteriol.* **172**, 266–272.
- Schauber, C., Chen, L., Tongaonkar, P., Vega, I., Lambertson, D., Potts, W. & Madura, K. (1998) *Nature (London)* **391**, 715–718.
- Wang, Y., Yeung, Y. G., Langdon, W. Y. & Stanley, E. R. (1996) *J. Biol. Chem.* **271**, 17–20.
- Lee, P. S., Wang, Y., Dominguez, M. G., Yeung, Y. G., Murphy, M. A., Bowtell, D. D. & Stanley, E. R. (1999) *EMBO J.* **18**, 3616–3628.
- Joazeiro, C. A., Wing, S. S., Huang, H., Levenson, J. D., Hunter, T. & Liu, Y. C. (1999) *Science* **286**, 309–312.
- Withers-Ward, E. S., Jowett, J. B., Stewart, S. A., Xie, Y. M., Garfinkel, A., Shibagaki, Y., Chow, S. A., Shah, N., Hanaoka, F., Sawitz, D. G., *et al.* (1997) *J. Virol.* **71**, 9732–9742.
- Vadlamudi, R. K., Joung, I., Strominger, J. L. & Shin, J. (1996) *J. Biol. Chem.* **271**, 20235–20237.
- Salamov, A. A. & Solovyev, V. V. (1995) *J. Mol. Biol.* **247**, 11–15.