The catalytic core of DEMETER guides active DNA demethylation in Arabidopsis

Changqing Zhang\textsuperscript{a,b,1}, Yu-Hung Hung\textsuperscript{a,b,1}, Hyun Jung Rim\textsuperscript{c,d,e,1}, Dapeng Zhang\textsuperscript{i}, Jennifer M. Frost\textsuperscript{b,2}, Hosub Shin\textsuperscript{c,d,e}, Hosung Jang\textsuperscript{f,d,e}, Fang Liu\textsuperscript{a,b,h}, Wenyuan Xiao\textsuperscript{i}, Lakshminarayan M. Iyer\textsuperscript{i}, L. Aravind\textsuperscript{i}, Xiang-Qian Zhang\textsuperscript{a,b,3}, Robert L. Fischer\textsuperscript{b,3}, Jin Hoe Huh\textsuperscript{c,d,e,3}, and Tzung-Fu Hsieh\textsuperscript{a,b,3}

\textsuperscript{a}Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695; \textsuperscript{b}Plants for Human Health Institute, North Carolina State University, North Carolina Research Campus, Kannapolis, NC 28081; \textsuperscript{c}Department of Plant Science, Seoul National University, 08826 Seoul, Republic of Korea; \textsuperscript{d}Research Institute for Agriculture and Life Sciences, Seoul National University, 08826 Seoul, Republic of Korea; \textsuperscript{e}Plant Genomics and Breeding Institute, Seoul National University, 08826 Seoul, Republic of Korea; \textsuperscript{f}Department of Biology, St. Louis University, St. Louis, MO 63103; \textsuperscript{g}Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720; \textsuperscript{h}State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioreources, College of Agriculture, Guangxi University, 530004 Nanning, China; \textsuperscript{i}National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894; and \textsuperscript{j}College of Forestry and Landscape Architecture, South China Agricultural University, 510642 Guangzhou, China

Contributed by Robert L. Fischer, July 12, 2019 (sent for review April 29, 2019; reviewed by James J. Giovannoni and Julie A. Law)

The Arabidopsis DEMETER (DME) DNA glycosylase demethylates the maternal genome in the central cell prior to fertilization and is essential for seed viability. DME preferentially targets small transposons that flank coding genes, influencing their expression and initiating plant gene imprinting. DME also targets intergenic and heterochromatic regions, but how it is recruited to these differing chromatin landscapes is unknown. The C-terminal half of DME consists of 3 conserved regions required for catalysis in vitro. We show that this catalytic core guides active demethylation at endogenous targets, rescuing dme developmental and genomic hypermethylation phenotypes. However, without the N terminus, heterochromatin demethylation is significantly impeded, and abundant CG-methylated genic sequences are ectopically demethylated. Comparative analysis revealed that the conserved DME N-terminal domains are present only in flowering plants, whereas the domain architecture of DME-like proteins in nonvascular plants mainly resembles the catalytic core, suggesting that it might represent the ancestral form of the 5mC DNA glycosylase found in plant lineages. We propose a bipartite model for DME protein action and suggest that the DME N terminus was acquired late during plant evolution to improve specificity and facilitate demethylation at heterochromatin targets.

DNA methylation is a covalent modification that influences the transcription of nearby genes and regulates important processes in eukaryotic genomes, including cell differentiation, transposable element (TE) silencing, and genomic imprinting (1, 2). Plant DNA methylation occurs in CG, CHG, and CHH sequence contexts (H = A, C, or T) and is targeted primarily to TEs. Flowering plants and mammals can also exhibit gene body methylation (gbM) in the CG context, generally in constitutively expressed genes, but the function of gbM is not fully understood (3, 4).

DNA methylation homeostasis is essential for genome stability, notably in maintaining TE silencing, and for the stable inheritance of epigenetic information (5, 6). In plants, this is achieved by maintenance and de novo DNA methylation, as well as by active DNA demethylation (1, 7, 8). Active DNA demethylation is catalyzed by a family of DNA glycosylases, including REPRESSOR OF SILENCING 1 (ROSI), DEMETER (DME), and DEMETER-LIKE 2 (DML2) and DML3, through a base excision repair pathway (9–11). However, epigenetic profiles are dynamic in response to biotic and abiotic stress and during reproduction and development. Similar to mammals, flowering plants require epigenetic reprogramming during gamete formation, which is characterized by extensive DNA demethylation in Arabidopsis by DME (12–14).

DME encodes a bifunctional 5mC DNA glycosylase/lyase that is essential for reproduction (9, 15). Paralogs ROS1, DML2, and DML3 function primarily in the sporophyte to counteract the spread of DNA methylation mediated by RNA-dependent DNA methylation (16, 17). The A, glycosylase, and B regions of the C-terminal half of DME are conserved among the DME/ROS1 DNA glycosylase clade and are absolutely required for DME 5mC excision in vitro, composing the catalytic core for its enzymatic activity (9, 15). DME acts primarily in the central cell and the vegetative nucleus (15, 18, 19). The vegetative nucleus contributes to germination and growth of the pollen tube, which

Significance

Flowering plants reproduce via a unique double-fertilization event, producing the zygote and the nutritive endosperm. The genome of the central cell, the precursor of the endosperm, undergoes extensive demethylation prior to fertilization. This epigenetic reconfiguration, directed by the DEMETER (DME) glycosylase at thousands of loci in Arabidopsis, differentiates the epigenetic landscapes of parental genomes and establishes parent-of-origin expression of many imprinted genes in endosperm essential for seed development. However, how DME is targeted to various locations remains unknown. Here we show that the multidomain DME is organized into 2 functional regions: the C-terminal region, which guides localization and catalysis, and the N-terminal region, which likely recruits chromatin remodelers to facilitate demethylation within heterochromatin.


The authors declare no conflict of interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo/ (accession no. GSE130559).

C.Z., Y.-H.H., and H.J.R. contributed equally to this work.

Present address: Blizzard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, E1 2AT London, United Kingdom.

To whom correspondence may be addressed. Email: aacrav@163.com, rfischer@berkeley.edu, huhjh@snu.ac.kr, or thsieh3@ncsu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1907290116/-/DCSupplemental.

Published online August 13, 2019.
delivers the sperm cells to the female gametophyte. Following double fertilization, the egg and central cell develop into the embryo and the nutritive endosperm, respectively, the latter of which accumulates starch, lipids, and storage proteins to nourish the developing embryo. The endosperm is the site of plant genomic imprinting, resulting from allelic inheritance of differential epigenetic states (20–22). DNMT1 homolog MET1-mediated DNA methylation and DNA-mediated demethylation are important regulators of plant gene imprinting. For example, DME demethylation is required to activate MEA, FIS2, and FWA expression in the central cell, which persists in the endosperm, while MET1 maintains the silencing of FIS2 and FWA paternal alleles (20–22). Imprinting is essential for reproduction in Arabidopsis, and seeds that inherit a maternal DME allele abort due to failure to activate MEA and FIS2, essential components of the Polycomb Repressive Complex 2 (PRC2) required for seed viability (20–22).

Although DME preferentially targets small AT-rich and nucleosome-poor eukaryotic transposons, it also demethylates intergenic and heterochromatin targets (13). How DME is recruited to target sites with various chromatin structures is unknown, although the Facilitates Chromatin Translocations (FACT) histone chaperone is required at heterochromatic targets and some imprinted loci (23, 24). Other than the glycosylase domain, the catalytic core region of DME contains multiple conserved globular domains of unknown function.

Here we show that expressing a nuclear-localized DME catalytic region controlled by a native DME promoter complements dme seed abortion and pollen germination defects and partially rescues the DNA hypermethylation phenotype in endosperm. Our DNA methylation analysis revealed that the majority of canonical DME target sites are demethylated by the catalytic core, indicating that this region is sufficient to direct DME localization. However, without the N-terminal region, the degree of demethylation is reduced, and demethylation of heterochromatin targets is particularly impeded. In addition, we observed prevalent ectopic demethylation, particularly at genomic sites, by the catalytic core. Thus, the N-terminal region of DME is likely needed for the full breadth and depth of demethylation and to prevent gene body demethylation. We show that the N-terminal conserved domains are specific to the angiosperm lineage, acquired late during land plant evolution, potentially to ensure robust demethylation in nucleosome-rich heterochromatin targets.

Results

Nuclear-Localized Catalytic Core of DME Rescues dme Developmental Defects. DME produces at least 2 alternatively spliced variants, encoding 2 hypothetical polypeptides of 1,729 (DME.1) and 1,987 (DME.2) amino acids in length (11, 15). The amino acid positions denoted in this study correspond to DME.2, the predominant isoform expressed in floral tissues (18). First, we detailed the domain structure and predicted characteristics of the catalytic and noncatalytic regions of the DME protein. The N-terminal half of DME consists of a large portion of unstructured, low-complexity sequences (amino acids 364–947; Fig. 1A), a stretch of basic amino acid-rich repeats (amino acids 291–363; basic stretch) known to direct nuclear localization (15), and a 120-aa N-terminal domain (amino acids 1–120, DmeN) of unknown function present only among the angiosperm DME/ROSI-like proteins. Within the DmeN domain is a sWIPxTPxKs motif that is highly conserved (SI Appendix, Fig. S1) but is absent in the shorter DME.1 isoform. This motif has a hydrophobic core and may mediate protein–protein interactions and/or be subjected to posttranslational modification (i.e., phosphorylation or methylation at the lysine residue). The basic stretch region is highly conserved among angiosperm DME-like proteins and is reminiscent of the AT-hook motifs that can bind DNA in a nonsequence-specific manner (25), suggesting that the basic stretch might also bind DNA along with directing DME to the nucleus.

It was determined that the C-terminal half of DME (amino acids 936–987; hereinafter the DME CTD) constitutes the catalytic core for 5mC excision in vitro (9, 26). This core contains the glycosylase domain of the HhH (helix-hairpin-helix) motif, followed by the [4Fe-4S] cluster loop (FCL) motif (27). The conserved B region contains an RNA recognition motif fold (RRMF) and a divergent, circularly permuted version of a methylated CpG-discriminating CXXC domain (27). The presence of the permuted CXXC and RRMF raises the possibility that the enzymatic core might contain intrinsic targeting information (28, 29). To test this possibility, we investigated whether DME CTD can rescue the dme seed abortion phenotype when expressed by an endogenous DME promoter. A classical SV40 nuclear localization signal (NLS) (PKKKPKV) was included to ensure robust nuclear localization (Fig. 1A).

Fig. 1. Complementation results. (A) Domain architecture and the positions of conserved domains along DME protein. nDME CTD is the construct used for complementation and methylome analyses. (B) In dme-2/dme-2 siliques, >99% of seeds are aborted. A single copy of the nDME CTD transgene reduces seed abortion to 50%, and in the dme-2/dme-2; nDME CTD/nDME CTD siliques, all seeds develop normally. (Scale bar: 0.5 mm.) (C) Percentages of viable seeds in DME/ dme-2 or in dme-2/dme-2 plants that were complemented by nDME CTD transgene. Error bars represent SD. (D) The nDME CTD transgene restores DME target gene FWA and FIS2 expression. WT: Col-0; nDME CTD: dme-2/dme-2; nDME CTD/nDME CTD, dme-2: dme-2 homozygotes. Total RNA was isolated from stage F1 to F12 floral buds.
resulting transgene, ndME<sup>CTD</sup>, was transformed into DME/dme-2 heterozygous Col-gl plants. Self-pollinated DME/dme-2 plants produce 50% viable (inherited DME maternal allele) and 50% aborted (inherited dme-2 maternal allele) F1 seeds. In contrast, transgenic lines that are DME/dme-2 and hemizygous for a single-locus ndME<sup>CTD</sup> transgene produced a 3:1 ratio of viable to aborted F1 seeds when self-pollinated (1:2:9:2:6; 3:1; \( \chi^2 = 0.12, P > 0.77, \) SI Appendix, Table S1), indicating that ndME<sup>CTD</sup> fully complemented the dme seed abortion phenotype. When ndME<sup>CTD</sup> was transformed into dme-2/dme-2 homozygous plants (SI Appendix, Materials and Methods), the T1 transgenic lines (dme-2/dme-2/ndME<sup>CTD</sup>−/) displayed a 50% to 75% rate of viable seeds when self-pollinated (SI Appendix, Table S1), compared with selfed nontransformed dme-2/dme-2 plants, which bear <0.05% viable seeds, again indicating full seed abortion complementation by ndME<sup>CTD</sup> (Fig. 1 B and C and SI Appendix, Table S1). These results show that ndME<sup>CTD</sup> rescues the dme-2 seed abortion phenotype. Since seed abortion is partially due to defects in activating imprinted PRC2 genes (22), we used qRT-PCR to measure the expression of PRC2 subunit FIS2 and of FWA, whose maternal expression is enabled by DME-mediated DNA demethylation, and found that ndME<sup>CTD</sup> restored the expression of these genes (Fig. 1D).

DME is also expressed in the vegetative cells of pollen, and mutations in DME reduce pollen germination, resulting in lower transmission of the paternal dme allele in certain ecotypes (19). Under our growth conditions, wild-type and dme-2 heterozygous Col-gl plants were self-pollinated, approximately 20% to 30% (compared with the expected 50% when the dme-2 allele is normally transmitted) of the F1 progeny were dme-2 heterozygotes (SI Appendix, Table S2). To test whether ndME<sup>CTD</sup> can rescue the dme pollen phenotype, we pollinated wild-type Col-0 with pollens from transgenic lines that are dme-2 homozygous and hemizygous for the ndME<sup>CTD</sup> transgene (dme-2/dme-2; ndME<sup>CTD</sup>−/−) with ~50% seed abortion rates (SI Appendix, Table S1). If ndME<sup>CTD</sup> does not complement dme-2 pollen germination defects, we would expect roughly one-half of the F1 progeny to carry the ndME<sup>CTD</sup> transgene (hygromycin-resistant), because mutant pollen with or without the transgene would germinate with equal frequency. Instead, we observed that 65% to 90% of the F1 progeny are hygromycin-resistant (resistant:sensitive = 190:52, 1:1, \( \chi^2 = 7.3E-19, P = 7.7E-75, \) SI Appendix, Table S5), indicating that ndME<sup>CTD</sup> rescues dme-2 pollen defects. These results show that ndME<sup>CTD</sup> can rescue the dme developmental phenotype, indicating that DME targeting information is contained within the catalytic core.

**Canonical DME Target Loci Are Demethylated by ndME<sup>CTD</sup>**

The molecular cause of DME mutant phenotypes is a loss of DME-mediated DNA demethylation. To test the extent of ndME<sup>CTD</sup> demethylation, we compared the methylation of the ndME<sup>CTD</sup>-complemented endosperm with the methylation of wild-type and dme-2 endosperm (13). ndME<sup>CTD</sup>-complemented endosperm methylation from 3 independent lines (dme-2/dme-2; ndME<sup>CTD</sup>−/) were generated and the reads combined for downstream analyses. Pearson correlation coefficients between independent lines showed that they were highly concordant (SI Appendix, Table S4). We compared the differentially methylated regions (DMRs) between dme-2 and wild-type endosperm (canonical DME targets; \( n = 8.872 \)), and the DMRs between dme-2 and ndME<sup>CTD</sup>-complemented endosperm (ndME<sup>CTD</sup> targets; \( n = 8.939 \)). Looking genome-wide, but excluding all DME and ndME<sup>CTD</sup> targets, the Pearson correlation coefficients between our combined independent lines and the previously published wild-type and dme-2 endosperm datasets ranged between 0.92 and 0.94 (SI Appendix, Table S5), indicating a high level of concordance between the methylation types used in this study.

We found several DME-regulated imprinting control regions of maternally and paternally expressed genes (MEGs and PEGs, respectively) to be hypomethylated in the ndME<sup>CTD</sup>-complemented endosperm compared with dme-2 endosperm (SI Appendix, Fig. S2), suggesting that ndME<sup>CTD</sup> is active at these loci. CG methylation does not return to wild-type levels, however, indicating that the genome is demethylated to a lesser degree by ndME<sup>CTD</sup> than by wild-type DME (Fig. 2 A and SI Appendix, Fig. S3A). The DME and ndME<sup>CTD</sup> DMRs largely overlap (Fig. 2B), and for the DMRs that appear unique to DME, the same regions are also demethylated by ndME<sup>CTD</sup> (Fig. 2C, black solid line trace), but to a reduced degree (the solid black peak is on the left of the dotted peak) that falls below our DMR cutoff (fractional CG methylation difference ≥0.3; \( P < 10^{-10} \), Fisher’s exact test). The shared DMRs are also slightly less demethylated in ndME<sup>CTD</sup>-complemented endosperm compared with wild-type endosperm (Fig. 2D, red trace the left of the black trace). Taken together, these data show that ndME<sup>CTD</sup> rescues the dme hypermethylation phenotype, but only partially. As far as we could ascertain, this was not a result of lower transgene expression, since qRT-PCR analyses of endosperm tissue showed abundant expression of ndME<sup>CTD</sup> (SI Appendix, Fig. S3B).

To investigate whether chromatin features influence ndME<sup>CTD</sup> demethylation, we assessed histone marks and genomic characteristics (30) in target sites that are ndME<sup>CTD</sup>-unique, DME-unique, or shared between the 2 types. Compared with wild-type, DME and ndME<sup>CTD</sup> shared DMRs, DME-unique target sites are highly enriched for heterochromatin states 8 and 9 (\( \chi^2 \) test, \( P = 8.29E^{-7} \), kb as unit of length; Fig. 2E). ndME<sup>CTD</sup> DMRs (unique and shared) are enriched for open chromatin states (\( \chi^2 \) test, \( P = 4.46E^{-117} \)), but show significant reductions in heterochromatic chromatin states 8 and 9 compared with DME DMRs (\( \chi^2 \) test, \( P = 8.345E^{-153} \), Fig. 2E). Thus, ndME<sup>CTD</sup> demethylates poorly at heterochromatic loci and preferentially targets euchromatin.

**Reduced Demethylation Efficiency at Long Heterochromatic Target Sites by ndME<sup>CTD</sup>**

Longer DME DMRs almost exclusively reside in heterochromatin (86.5% of 1–1.5 kb and 95.5% of ≥1.5 kb; Fig. 3A). We postulate that this is due to the dense methylation associated with heterochromatin, which may result in longer stretches of DNA demethylation during DME occupancy at these sites. Interestingly, the number of long DMRs is dramatically reduced in ndME<sup>CTD</sup>-complemented endosperm (Fig. 3 A and B). This reduction in the number of longer DMRs was not due to a lack of ndME<sup>CTD</sup> targeting to these sites, since the partial demethylation characteristic of ndME<sup>CTD</sup> activity occurred in all targets regardless of their length (SI Appendix, Fig. S3C). However, when we analyzed the length of the ndME<sup>CTD</sup> demethylated regions, we found that it produced much shorter DMRs in the heterochromatin targets (SI Appendix, Fig. S4). For example, there are 250 DME DMRs longer than 1.5 kb (median length, 1.9 kb). Among these, 165 are also DMRs of ndME<sup>CTD</sup> but are much shorter (median length, 400 bp) (Fig. 3 C and D). Thus, removal of the DME N-terminal region significantly reduced the extent of demethylation in these long targets.

The histone chaperone FACT complex is required for demethylation of approximately one-half of DME targets in Arabidopsis, particularly those in heterochromatin (23). DME colocalizes with SPT16 (the larger FACT subunit) in an in vivo biomolecular fluorescence complementation assay, suggesting that DME might recruit the FACT complex to these heterochromatic loci (23). Of the 250 long DME DMRs, 87% of them require FACT, raising the possibility that ndME<sup>CTD</sup> might be defective in recruiting FACT. To test this hypothesis, we examined how ndME<sup>CTD</sup> demethylates FACT-dependent vs. FACT-independent loci. In wild-type endosperm, both target groups are demethylated to a similar degree (Fig. 3 E, blue and green traces with similar shape and peak location). In ndME<sup>CTD</sup>-complemented
endosperm, FACT-independent loci are only slightly less demethylated compared with wild-type endosperm (Fig. 3E, orange trace moderately shifted to the left of blue and green traces). In contrast, demethylation at FACT-dependent loci is more severely impeded (Fig. 3E, magenta trace). These results support a model in which DME recruits FACT via its N-terminal region to make heterochromatic targets accessible.

Prevalent Ectopic Gene Body Demethylation by nDMECTD. We identified a set of new DMRs unique to nDMECTD, which we term “ectopic” targets (Fig. 2B). The CG methylation difference between dme-2 and wild-type endosperm for these DMRs is minimal but not absent (Fig. 2C, the red trace peaks close to 0, with a positive shoulder). We plotted the methylation status of nDMECTD-unique loci in wild-type endosperm to assess how they are demethylated by DME. This resulted in positive peaks for shorter (red trace) or longer (blue trace) TEs and intergenic sequences (green trace), showing that these nDMECTD-unique loci are also demethylated by DME but not to a sufficient degree to reach the DMR cutoff. In contrast, most nDMECTD-unique loci within coding sequences are not demethylated by DME (Fig. 4A, orange trace). This indicates that genic sequences are the primary ectopic targets of nDMECTD. This is also reflected by an increase in nDMECTD DMR frequency (Fig. 4B, red and orange traces) and a decrease in average CG methylation (SI Appendix, Fig. S5) within coding genes.

Gene body CG methylation, or gbM, is an evolutionarily conserved feature in mammals and angiosperms, but the origin and function of gbM has not been elucidated (3, 4). Approximately 15% (≈5,000) of Arabidopsis genes contain gbM (31). In our endosperm methylomes, we found 2,202 and 3,213 genes associated with DME and nDMECTD DMRs, respectively, which were largely mutually exclusive and thus constituted almost all of the genes with gbM (Fig. 4C). Among these, 2,260 genes were ectopically targeted by nDMECTD (Fig. 4C and Dataset S1). Compared with DME-targeted coding genes, these nDMECTD-unique genes have higher expression levels (Fig. 4D) and greater CG methylation (Fig. 4E). They include genes across most actively used cellular processes (SI Appendix, Table S6), consistent with the current theory that moderate gbM positively correlates with constitutively transcribed genes (3). Thus, nDMECTD has a greater tendency than DME to target higher CG methylated coding sequences compared with lower methylated genes.

Evolutionary History of the DME/ROS1 Glycosylase Family. We show that the catalytic region of DME is able to demethylate DNA in vivo and also has targeting ability. We also demonstrate that the N terminus likely plays a role in fine-tuning DME targeting to heterochromatin and restricting it from gene bodies. To provide clues as to the evolutionary origin of these bipartite domains, we carried out comparative analyses across plant lineages. Using various DME homologs as query sequences, we revealed a diversity of N-terminal domains associated with the DME catalytic core across various clades (Fig. 5). These indicate that a shorter protein, comprising only the C terminus of Arabidopsis DME, may represent the ancestral form of the 5mC glycosylase found in all plant lineages. Variations of the N-terminal domains are found in land plants and charophytes (Streptophyta), which have a divergent circularly permuted CXXC domain between the FCL and RRMF domains. In contrast, 1 or more copies of the classical CXXC can be found in chlorophyte and stramenopile algae at distinct positions. Chlorophyte
and stramenopile homologs also have additional chromatin readers, such as Tudor and PHD; DNA-binding domains, such as the AT-hook; and the Hsp70-interacting DnaJ domains (27). These accessory domains suggest a mode of regulating DNA glycosylase activity according to methylation status (via CXXC) or chromatin states (via PHD and Tudor). The DemeN and basic stretch of the DME N-terminal region are restricted to the angiosperm lineage and appear to be a late acquisition during land plant evolution (Fig. 5). Thus, the acquisition of this region coincides with the origins of double fertilization in plants and the emergence of plant gene imprinting.

**Discussion**

DME regulates gene imprinting and influences transgenerational epigenetic inheritance in *Arabidopsis* (32). DME demethylates the central and vegetative gamete companion cell genomes at thousands of loci, but the mechanism of DME targeting remains elusive. This is due to the restriction of its expression to the ephemeral nuclei embedded within gametophytes, which largely prohibits biochemical interrogation by currently available tools. In contrast, genetic analysis coupled with endosperm transcriptome and methylome profiling has been instrumental in revealing DME’s molecular function (9, 12–14, 23, 33). Here we used genetic complementation and endosperm DNA methylation profiling to show that the catalytic core of DME is sufficient to rescue the *dme* seed abortion phenotype and pollen germination defects (Fig. 1). We present evidence that nDME<sup>CTD</sup> can demethylate most canonical DME target sites, implying that the catalytic region contains targeting information. We propose the DME protein has a bipartite structure (Fig. 6) and demonstrate a requirement for the N-terminal region in assisting heterochromatin demethylation, possibly via FACT. Although nDME<sup>CTD</sup> complements *dme*-associated developmental defects (Fig. 1B and SI Appendix, Tables S1 and S3), it does not fully rescue the *dme* endosperm DNA hypermethylation phenotype. Instead, we observed a reduced degree of demethylation by nDME<sup>CTD</sup> in all the endogenous DME target loci, regardless of length (Fig. 2A and SI Appendix, Fig. S3C). To investigate the cause of this partial demethylation, we carried out data analyses to rule out technical differences in how experiments were performed. First, the nDME<sup>CTD</sup> methylomes were generated from selfing Col transgenic lines, whereas the control wild-type and *dme*-2 endosperm methylomes were derived from crosses of Col and Ler parents (13). To assess the extent to which the difference between Col×Ler and Col×Col F1 endosperm affects methylomes and/or how sample collection and preparation methods might influence the DNA methylation profile, we compared the methylation differences between wild-type (Col×Ler) and *nDME<sup>CTD</sup>-complemented* (Col×Col) endosperm at DME target sites and non-DME target sites. The kernel density plot shows a general trend toward slight hypomethylation at non-DME target sites in *nDME<sup>CTD</sup>-complemented* endosperm compared with...
wild-type endosperm was detected (SI Appendix, Fig. S6A), indicating that the difference between Col-0 and Col-0 endosperm, or differing sample preparations, does not contribute to the demethylation of DME target loci in nDMECTD-complemented endosperm.

We performed the following experiment to rule out the possibility that when propagating dme-2 homozygous lines, loss of DME activity might cause epigenome alterations that render certain DME target loci no longer recognizable by nDMECTD. Among the 3 independent nDMECTD-complemented lines, 1 line (nDMECTD-3) was propagated from a transformed DME/dme-2 heterozygous plant in which the sporophyte was never dme-2/dme-2 in the absence of the nDMECTD transgene. Two other lines were generated by directly transforming dme-2/dme-2 homozygous plants, with dme-2 ovules derived from dme-2/dme-2 megaspore mother cells before nDMECTD complementation. Examination of the demethylation profiles of these 3 independent lines at DME canonical target loci revealed that the 3 lines had very similar demethylation profiles (SI Appendix, Fig. S6B), indicating that nDMECTD partial demethylation is likely not a result of unexpected epigenome alterations induced by dme-2/dme-2 plants.

The nDMECTD transgene, although controlled by a native DME promoter and fully complemented dme-2 seed abortion (Fig. 1B and C), might lack critical regulatory elements needed to drive stable protein production to the levels comparable to the endogenous DME. Thus, an endosperm methylome complemented by a full-length DME cDNA construct within the otherwise similar T-DNA backbone as the nDMECTD is needed to rule out any unforeseen artifacts introduced by the transgene. Throughout the course of this study, a full-length DME cDNA (DMEFL; ref. 15) was included side by side as a control for complementation assays and for methylation profiling. However, for reasons not completely understood, we observed that although it complemented dme-2 seed abortion (3:1 ratio of viable to aborted F2 seeds, 1,518:556, χ² = 3.62, P = 0.057; SI Appendix, Fig. S6C and Table S7) as reported previously (15, 34), the DMEFL consistently suffered from negative interference in the dme-2 mutant background that resulted in very low demethylation activity in vivo, which most likely does not reflect the activity of the full-length DME protein (SI Appendix, Fig. S6D). To avoid confounding the interpretation, we excluded the DMEFL results and focused only on methylome comparisons between wild-type and nDMECTD-complemented endosperm. Although we cannot completely rule out a negative effect of the dme-2 genetic background on the nDMECTD transgene, 2 factors suggest that this negative effect, if any, would be minimal. First, nDMECTD generates roughly the same number of DMRs as the endogenous DME (Fig. 2B). Second, within the shared DMRs between nDMECTD and DME, the degree of demethylation is quite comparable (Fig. 2D and SI Appendix, Fig. S7A). When directly comparing the difference between DME and nDMECTD (DME minus nDMECTD) in these completely overlapping DMRs that cover >1.2 million bases, the density plot peaks slightly to the negative side with a broader shoulder (SI Appendix, Fig. S7B).
indicating that nDME<sup>CTD</sup> is capable of demethylation to almost the DME level in these tissues. Thus, we suspect that nDME<sup>CTD</sup>-complemented endosperm methylome data can reasonably reflect the catalytic activity of the nuclear-localized DME<sup>CTD</sup>. However, we also caution that in the absence of a proper full-length DME cDNA control, we cannot completely rule out that partial demethylation of DME<sup>CTD</sup> might be caused by other factors.

Finally, it is also possible that the N-terminal region may assist the glycosylase by binding to DNA templates (via the AT-hook-like motif) to promote demethylation, and that without it, DME<sup>CTD</sup> has a reduced affinity to target DNA and exhibits lower demethylation activity. Previous studies of ROS1 support such a model, showing that the basic stretch/AT-hook region of ROS1 binds strongly to DNA templates in vitro in a non-sequence-specific manner, and that removal of the ROS1 basic stretch region impairs the sliding capacity of ROS1 on the DNA template (35), significantly reducing ROS1 5mC excision activity (25).

DME preferentially targets smaller euchromatic transposons that flank coding genes, and also targets gene-poor heterochromatin regions for demethylation (13). Since heterochromatin regions are compacted, demethylation in these regions requires substantial chromatin remodeling, including the temporary eviction of nucleosomes for DME to gain access to DNA. The FACT complex has been shown to be required for DME-mediated demethylation, primarily at heterochromatin targets (23), and we also noted that these sites have increased nucleosome occupancy (SI Appendix, Fig. S8). It is tempting to speculate that the DME N-terminal region is required to recruit FACT activity for demethylation (Dataset S2), SPT16 has been shown to colocalize with DME in vivo (23), suggesting a direct or indirect interaction between DME and FACT.

nDME<sup>CTD</sup> also displayed a reduced capacity for demethylating FACT-independent loci. Thus, it is possible that the N-terminal region is needed to recruit other chromatin remodelers at FACT-independent targets, that is, if nucleosomes are natural barriers for DME demethylation in euchromatin as well as heterochromatin (SI Appendix, Fig. S8). We envision a working model (Fig. 6) in which the catalytic region directs DME to FACT-independent loci, while the N-terminal region is required to interact with the local chromatin environment, stabilizing binding to the
chromosomal template and/or aiding demethylation of flanking sequences by remodeling nucleosomes.

gbM is evolutionarily conserved, and approximately 15% of Arabidopsis genes contain gbM (4). CG-methylated genes are often constitutively expressed housekeeping genes (5), raising the possibility that these genes have evolved to reside in the nuclear activity of DME that is more accessible by nDMECTD than by DME. It is also possible that DME is actively repelled by certain open chromatin histone marks, and such repulsion is missing in nDMECTD. This scenario would be analogous to the mammalian de novo DNA methyltransferase DNMT3, where its binding to an allosteric activator, unmethylated histone H3, is strongly inhibited by H3K4 methylation (36). Equally probable is that additional factor(s) (e.g., methyl-binding proteins) might be associated with these higher CG-methylated genes (Fig. 4E) that restrict DME access.

Tracing the evolutionary history of DME-like genes (Fig. 5), we found that a bacterial version of the HhH-FCL pair underwent a horizontal gene transfer to the ancestor of plants, followed by a gene duplication. One copy was fused to an RMRF domain and further acquired an insert in the glycosylase domain, giving the ancestral form of DME in plants. This was likely then transferred to the stramenopiles from a secondary chlorophyte endosymbiont of this lineage. Finally, at the base of the streptophyte lineage likely reflects adjustment to the unique methylation and chromatin environment of the larger Streptophyta and land plant genomes.

Our study shows that the catalytic core of DME was present in ancient plant ancestors and is alone capable of targeting and demethylating. The N-terminal domains, confined to flowering plant DME proteins, likely evolved to facilitate access to diverse chromatin states, in turn mediating gene imprinting and the transgenerational silencing of transposons (13).

**Experimental Procedures**

Detailed descriptions of the experimental methods are provided in *SI Appendix, Methods*. Sequencing data have been deposited in the Gene Expression Omnibus database (accession no. GSE130559).

**Plant Materials and Seed Phenotype Analysis.** Heterozygous DME/dme-2 lines in the Col-0 background were subjected to Agrobacterium-mediated transformation. Seeds were sterilized with 30% bleach solution; plated on a 0.5x MS nutrient medium with 1.5% sucrose, 0.8% agar, and 40 μg/mL hygromycin; and stratified at 4 °C for 2 d. Germinated seedlings were transferred to soil and grown in a growth room under a 16-h light/8-h dark cycle at 23 °C. Siliques from T1 transgenic plants were dissected at 14–16 d after self-pollination using a stereoscopic microscope (Stetero Discovery.V12; Carl Zeiss). The numbers of viable and aborted seeds in transgenic lines were statistically analyzed with the χ2 test. The probabilities of deviation from a 1:1 or 3:1 segregation ratio for viable and aborted seeds were also calculated.

**RNA Extraction, cDNA Synthesis, and qRT-PCR Analysis.** Total RNA was extracted using Trizol reagent (Invitrogen) and treated with TURBO DNase (Ambion) according to the manufacturers' instructions. For cDNA synthesis, 5 μg of total RNA was reverse-transcribed using SuperScript III reverse transcriptase and oligo(dT) primer (Invitrogen). cDNA was treated with RNase H (Invitrogen) at 37 °C for 20 min and then diluted 10-fold with H2O. For each 15-μl qPCR reaction, 1 μL of diluted cDNA was used. qPCR analyses were run on an Applied Biosystems 7500 Fast Real-Time PCR System using Roche FastStart Universal SYBR Green Master Mix. The qRT-PCR primers are listed in *SI Appendix, Table S8*. Ct values were normalized against ACT2 (At3g17870) mRNA or UBC (At5g25760) mRNA. The abundance of mRNAs was expressed as relative to controls, with the control value set to 1. The error bars represent the SD of 4 biological replicates.

**Protein Domain Analysis and Phylogenetic Inference.** We used a domain-centric computational strategy to study DME and its related proteins. Specifically, we identified DME homologs using iterative profile searches with PSI-BLAST (37) from the protein nonredundant database at the National Center for Biotechnology Information. Multiple sequence alignments were built using PROMAL S (38), followed by careful manual adjustments. Consensus secondary structures were predicted using the PSIPRED (39) Jpred program (40). Conserved domains were further characterized based on comparisons with available domain models from Pfam (41) and sequence/structural features. PhyML (42) was used to determine the maximum likelihood tree using the Jones–Taylor–Thornton model for amino acid substitution with a discrete gamma model (4 categories with gamma shape parameter 1.096). The tree was rendered using MEGA Tree Explorer (43).


