RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants

Marjori A. Matzke, Tatsuo Kanno, and Antonius J.M. Matzke

Institute of Plant and Microbial Biology, Academia Sinica, Taipei 115, Taiwan; email: marjorimatzke@gate.sinica.edu.tw, kanno0@gate.sinica.edu.tw, antoniusmatzke@gate.sinica.edu.tw

Annu. Rev. Plant Biol. 2015. 66:243–67

First published online as a Review in Advance on December 10, 2014

The *Annual Review of Plant Biology* is online at plant.annualreviews.org

This article's doi: 10.1146/annurev-arplant-043014-114633

Keywords

epigenetics, RNA-directed DNA methylation, RNA polymerase, small RNAs

Abstract

RNA-directed DNA methylation (RdDM) is an epigenetic process in plants that involves both short and long noncoding RNAs. The generation of these RNAs and the induction of RdDM rely on complex transcriptional machineries comprising two plant-specific, RNA polymerase II (Pol II)–related RNA polymerases known as Pol IV and Pol V, as well as a host of auxiliary factors that include both novel and refashioned proteins. We present current views on the mechanism of RdDM with a focus on evolutionary innovations that occurred during the transition from a Pol II transcriptional pathway, which produces mRNA precursors and numerous noncoding RNAs, to the Pol IV and Pol V pathways, which are specialized for RdDM and gene silencing. We describe recently recognized deviations from the canonical RdDM pathway, discuss unresolved issues, and speculate on the biological significance of RdDM for flowering plants, which have a highly developed Pol V pathway.

Contents

INTRODUCTION

Transgenes and RNA pathogens have been instrumental in the discovery of various types of homology-dependent gene-silencing phenomena in plants (92, 99). Posttranscriptional gene silencing (PTGS, also referred to as cosuppression) (103, 132), transcriptional gene silencing (TGS) (94), and RNA-mediated virus resistance (83) were all originally detected in plants engineered with various transgene constructs. Similarly, RNA-directed DNA methylation (RdDM) was first observed in transgenic tobacco plants infected with viroids, which are plant pathogens consisting solely of a non-protein-coding RNA several hundred base pairs in length (135).

Although the nature of the RNA species triggering RdDM was initially unknown, subsequent work implicated both small interfering RNAs (siRNAs) and long noncoding RNAs (lncRNAs). The siRNAs, which are derived from longer double-stranded RNA (dsRNA) precursors (98, 122), provide sequence-specific guides for methylation, whereas lncRNAs coordinate assembly of the silencing effector complex at target loci (139). The biogenesis and use of these RNAs to elicit RdDM require core proteins of the RNA interference (RNAi) machinery—DICER-LIKE (DCL), RNA-DEPENDENT RNA POLYMERASE (RDR), and ARGONAUTE (AGO) (21, 144; see also sidebar, Core RNAi Proteins)—as well as specialized transcriptional machineries comprising multiple plant-specific proteins. Foremost among these plant-specific factors and central to the mechanism of RdDM are two RNA polymerase II (Pol II)–related RNA polymerases known as Pol IV and Pol V. These polymerases are functionally diversified and act at different steps of the RdDM pathway: Pol IV is needed to produce the siRNA trigger for methylation, whereas Pol V acts downstream to enable de novo DNA methylation at the siRNA-targeted site, presumably by synthesizing lncRNAs (scaffold RNAs) that base pair to AGO-bound siRNAs and recruit other silencing components. Although the Pol IV branch of the RdDM pathway appears to be a common

CORE RNAi PROTEINS

RNAi refers to gene silencing triggered by dsRNA that is processed into siRNAs, which provide sequence-specific guides for inducing mRNA degradation (PTGS) or epigenetic modifications (TGS). Core RNAi proteins are involved in synthesizing the dsRNA precursor (RDRs), processing dsRNA to siRNAs (DCLs), and binding the siRNA guide strand in silencing effector complexes (AGOs) (13).

RDRs copy single-stranded RNA into dsRNA. *Arabidopsis* has six RDRs. RDR2 copies Pol IV–dependent transcripts into dsRNA in the canonical RdDM pathway, whereas RDR6 copies Pol II–dependent transcripts into dsRNA in the noncanonical RdDM pathway.

Dicer proteins, called DCLs in plants, are endoribonucleases that process larger dsRNA precursors into uniformly sized siRNAs. *Arabidopsis* has four DCLs: DCL1 generates 21-nt microRNAs, DCL2 generates 22-nt siRNAs from viruses, DCL3 generates 24-nt siRNAs involved in canonical RdDM, and DCL4 generates 21-nt siRNAs and *trans*-acting siRNAs.

In the silencing effector complex, AGO proteins bind siRNAs through their PAZ and MID domains and may have RNA slicer activity through their PIWI domains (97). There are ten AGO proteins in *Arabidopsis*, four of which—AGO4, AGO6, AGO8, and AGO9—are in the AGO4 clade, which incorporates 24-nt siRNAs. Primarily AGO4 and AGO6 act in canonical RdDM (13).

feature of land plants (22), the Pol V part has attained its highest level of complexity in flowering plants.

The proposed mechanism and biological roles of RdDM in plants have been discussed in several recent reviews (44, 79, 93, 111, 119, 137, 151). After summarizing this information, we focus on evolutionary innovations that occurred during the transition from the Pol II transcriptional pathway (which produces mRNA precursors and many noncoding RNAs) to the Pol IV and Pol V pathways (which generate the siRNAs and lncRNAs necessary for RdDM and epigenetic silencing). In addition to noncoding RNAs, RdDM requires refashioned and novel proteins whose precise roles in the mechanism of RdDM are not well understood. These specialized, plant-specific proteins provide some of the most compelling arguments that RdDM is distinct from other mechanisms of small RNA–mediated chromatin modifications in other eukaryotic organisms (18, 19, 20, 88). We highlight deviations from the canonical RdDM pathway, which illuminate a more prominent role for Pol II in RdDM; discuss several unsettled questions; and speculate on the significance of RdDM for flowering plants, which possess a complex Pol V pathway.

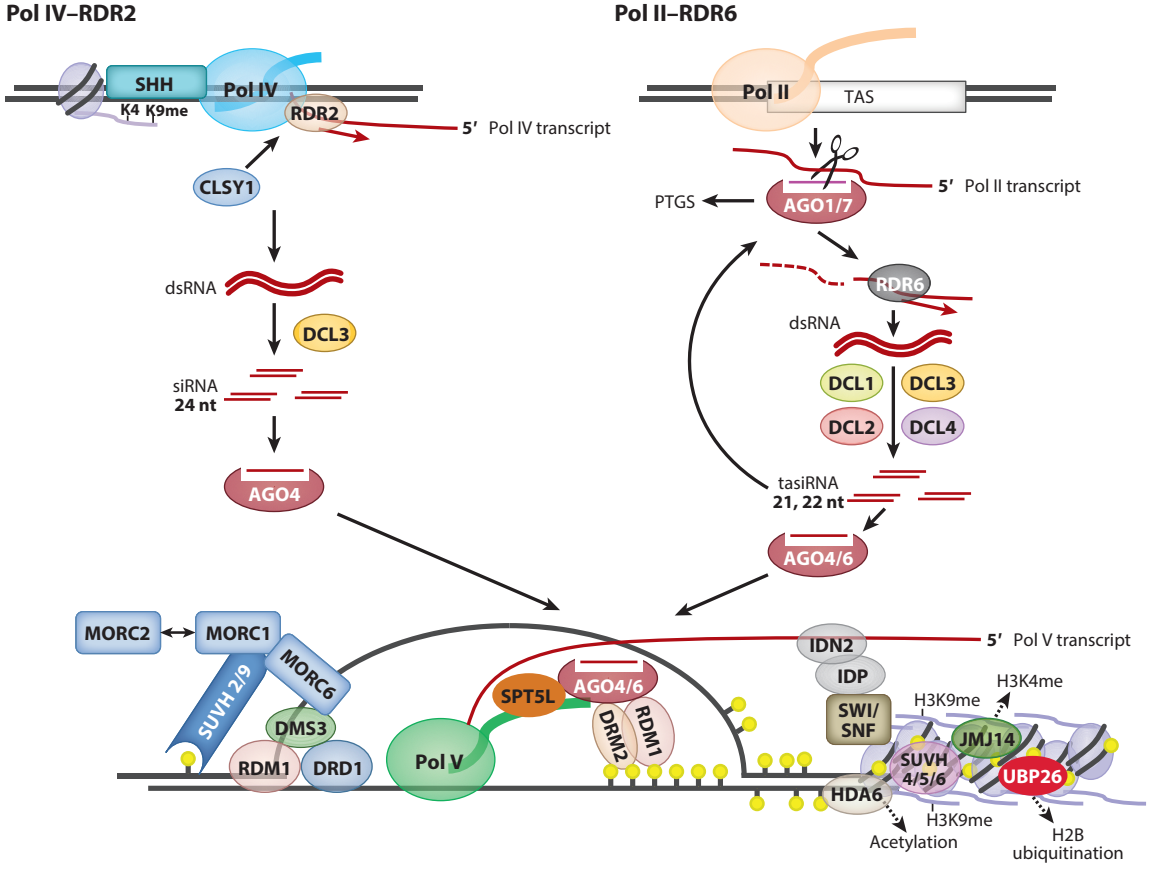
MECHANISM AND TARGETS OF RNA-DIRECTED DNA METHYLATION

Canonical RNA-Directed DNA Methylation

Current views on RdDM originate primarily from research conducted on *Arabidopsis thaliana*, which is amenable to genetic and biochemical analyses that have identified many components of the pathway and defined their mechanistic roles. According to contemporary models, the canonical RdDM pathway involves two sequential steps (44, 79, 93, 111, 119, 137, 151): (*a*) biogenesis of 24-nucleotide (nt) siRNAs, which requires Pol IV, RDR2, and DCL3, and (*b*) de novo methylation, which requires Pol V–dependent scaffold RNAs, AGO4-bound 24-nt siRNAs, and the de novo DNA methyltransferase DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), as well as several dedicated proteins whose precise roles are not

entirely clear (**Figure 1**). These specialized actors include members of the DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1) subfamily of Snf2 chromatin remodelers, which are necessary for both Pol IV and Pol V transcription. In addition, Pol V function requires unusual proteins that do not have strict counterparts in other organisms, such as RNA-DIRECTED DNA METHYLATION 1 (RDM1), and altered proteins, such as DEFECTIVE IN MERISTEM SILENCING 3 (DMS3), SUPPRESSOR OF TY INSERTION 5–LIKE (SPT5L), and microrchidia (MORC) proteins (**Figure 1**). Features of these proteins that are relevant for their roles in RdDM are discussed in more detail below. Several splicing factors are required for RdDM at different steps of the silencing pathway, although how they are recruited to target loci and how they act in the RdDM mechanism remain unclear (54).

RdDM modifies cytosines in all sequence contexts (CG, CHG, and CHH, where H is A, T, or C) within the region of RNA-DNA sequence homology, which can be as short as 30 base pairs (109, 110). siRNAs directed to transcriptional regulatory regions are able to trigger promoter methylation and TGS (98, 122). Methylation at symmetrical CG and CHG nucleotide groups is maintained during subsequent rounds of DNA replication through the action of maintenance methyltransferases, which recognize methylation on one DNA strand but not the other. CG and CHG methylation is maintained by the conserved Dnmt1-type enzyme METHYLTRANSFERASE 1 (MET1) and the plant-specific CHROMOMETHYLASE 3 (CMT3), respectively (44, 79, 93, 111, 119, 151). CMT3 works together with the members of the



SU(VAR)3–9 HOMOLOG 4 (SUVH4) family of histone methyltransferases to maintain methylation of histone H3 at the lysine 9 residue (H3K9), a mark of repressed chromatin (30, 36) (**Figure 1**).

Genetic screens have identified several histone-modifying enzymes that remove active marks to reinforce the silent state in conjunction with DNA methylation. These enzymes include HISTONE DEACETYLASE 6 (HDA6), which deacetylates multiple lysines on multiple histones (3, 9, 33); JUMONJI 14 (JMJ14), which demethylates lysine 4 of histone H3 (H3K4) (28, 121); and UBIQUITIN-SPECIFIC PROTEASE 26 (UBP26), which deubiquitinates histone H2B (125) (**Figure 1**). The dsRNA-binding protein INVOLVED IN DE NOVO METHYLATION 2 (IDN2) (4) acts in a complex with two partially redundant paralogs (IDP1 and IDP2) (40, 143, 149) and the SWI/SNF chromatin-remodeling complex (157) to adjust nucleosome spacing to facilitate DNA methylation (**Figure 1**).

Targets of RNA-Directed DNA Methylation

RdDM is targeted to transposons and repeats distributed throughout the genome, but its presence is particularly notable at smaller and younger transposons and repeats in euchromatic chromosome arms (56, 81, 130, 147, 156). Although DNA methylation is an important factor controlling transposons (58), these elements are not generally mobilized in mutants defective in RdDM in *Arabidopsis*, suggesting that stable silencing of transposons is not the major function of RdDM. Reliable transposon silencing depends instead on the chromatin remodeler DECREASED DNA METHYLATION 1 (DDM1), MET1, and CMT3, as indicated by the movement of transposons in mutants impaired in these factors (58). *Arabidopsis* mutants deficient in most RdDM factors also do not show obvious developmental phenotypes when grown under standard conditions (93). However, in other species that have larger genomes containing more repetitive DNA, such as

Figure 1

Canonical and noncanonical RNA-directed DNA methylation (RdDM) pathways. (*Upper left*) Pol IV–RDR2–dependent small interfering RNA (siRNA) biogenesis (canonical pathway). SHH1 binds to histone H3 unmethylated at lysine 4 (K4) and methylated at K9 (K9me) and recruits Pol IV to transcribe the target region. RDR2 physically interacts with Pol IV (45) and converts Pol IV transcripts into double-stranded RNA (dsRNA) with the assistance of the chromatin remodeler CLSY1. The dsRNAs are processed into 24-nucleotide (nt) siRNAs by DCL3, and the guide strand is incorporated into AGO4 or AGO6, which then enters the Pol V–mediated pathway of de novo DNA methylation. (*Upper right*) Pol II–RDR6–dependent siRNA biogenesis (noncanonical pathway). Pol II transcribes *TAS* noncoding RNAs, which undergo microRNA-guided slicing by either AGO1 or AGO7. An RNA cleavage product is copied by RDR6 into dsRNA, which is processed into 21–24-nt siRNAs by various DCL activities. The 21–22-nt size class, called *trans*-acting siRNAs (tasiRNAs), are loaded onto AGO1 or AGO7 to induce posttranscriptional gene silencing (PTGS) of complementary target mRNAs in the cytoplasm, while some tasiRNAs are incorporated into AGO4/6 to guide Pol V–mediated de novo DNA methylation of *TAS* genes. A related pathway of Pol II–RDR6 RdDM occurs at newly inserted transposons. (*Bottom*) Pol V–dependent DNA methylation and heterochromatin formation. Pol V, recruited to some loci by the methyl-DNA-binding proteins SUVH2 and SUVH9, transcribes the target locus to produce a scaffold RNA, which base pairs with the siRNA guide bound to AGO4/6. AGO4 interacts with the Ago-hook regions in SPT5L and the C-terminal domain of NRPE1 as well as DRM2, which catalyzes de novo methylation of DNA (*yellow lollipops*) at the siRNA-targeted site. RDM1 may link AGO4 and DRM2, and may also function with DRD1 and DMS3 in the DDR complex to ease Pol V transcription through chromatin and open the DNA duplex. Heteromers of MORC1 (or MORC2) and MORC6 associate with SUVH2/9 and may also interact loosely or transiently with DMS3 to form a cohesion-like complex that generates and stabilizes the unwound state. Pol V transcripts are also used for heterochromatin formation. The IDN2/IDP complex binds Pol V scaffold RNAs and recruits the SWI/SNF complex to favorably position nucleosomes for methylation. Active histone marks, such as H3K4me, acetylation, and ubiquitination, are removed from nucleosomes by JMJ14, HDA6, and UBP26, respectively (*dotted arrows*). SUVH4–6 catalyze repressive H3K9me to reinforce the silent state. All small RNAs are stabilized by methylation at their 3′ ends by HUA ENHANCER 1 (HEN1) (not shown) (153). Pol IV and Pol V likely depend on DMS4 for import into the nucleus (48, 65). Figure based in part on the transcription fork model developed by Pikaard et al. (111). *Arabidopsis* Genome Initiative (AGI) numbers for all factors are given in **Supplemental Table 1**.

Supplemental Material

maize and rice, mutations in genes encoding RdDM factors can affect development and transposon mobility, suggesting a larger role for RdDM in regulating gene expression and maintaining genome stability in these species (112, 136).

Pol IV is recruited to a subset of target loci through SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1), a SAWADEE domain–containing protein that recognizes unmethylated H3K4 and methylated H3K9 through its Tudor domain (78, 150). Recruitment of Pol V at loci containing preexisting DNA methylation is enhanced by SUVH2 and SUVH9, which bind methylated CG dinucleotides through their SET and RING finger–associated (SRA) domains (60, 73, 84) (**Figure 1**). Although several thousand sites of Pol IV and Pol V occupancy have been identified in chromatin immunoprecipitation sequencing experiments, consensus promoter motifs for these polymerases have not been discerned (78, 138). However, one study demonstrated that Pol IV uses a different promoter than Pol II to transcribe the same target sequence (117). Some targets of RdDM overlap with those of REPRESSOR OF SILENCING 1 (ROS1), a DNA glycosylase-lyase activity that functions in active demethylation of DNA (152). These common targets illustrate the potential reversibility of RdDM, which may dynamically modulate the expression of adjacent genes (51, 70) and be important for stress responses (93, 96).

Noncanonical RNA-Directed DNA Methylation

Recent findings indicate that canonical RdDM involving Pol IV–RDR2–DCL3 is just one variant of a more complex interconnected network of pathways that link RdDM with PTGS. At some *TAS* loci (68, 142), which encode *trans*-acting siRNAs, and newly integrated transposons (90, 105, 108, 113), a Pol II–RDR6 pathway that transitions from PTGS to TGS has been observed. This pathway relies on Pol II transcripts that are copied by RDR6 and processed by DCL2 and DCL4 into 21–22-nt siRNAs. These siRNAs are thought to interact with AGO4 or AGO6 to elicit RdDM in a Pol V–dependent manner and to bind to AGO1 to guide cleavage of target transcripts in a classical PTGS process (**Figure 1**). A plant-specific protein, NEEDED FOR RDR2-INDEPENDENT DNA METHYLATION (NERD), which contains several chromatin-associated domains as well as glycine-tryptophan (GW) repeats that bind AGO proteins (113), has been implicated in RDR6 RdDM. NERD associates with histone H3 and AGO2 and is needed for siRNA accumulation at NERD target loci, which tend to be newly acquired transposon insertions (113).

THE EVOLUTION OF POL IV AND POL V PATHWAYS

All eukaryotes have three different, multisubunit DNA-dependent RNA polymerases (RNAPs) that transcribe nuclear genes and are essential for viability: Pol I transcribes large ribosomal RNAs; Pol II transcribes mRNA precursors and numerous noncoding RNAs, which remain largely uncharacterized; and Pol III transcribes 5S ribosomal RNAs and transfer RNAs. Each RNAP has a pair of unique largest and second-largest subunits, which form the catalytic center, and up to ten smaller subunits, of which five (subunits 5, 6, 8, 10, and 12) are shared by all three RNAPs (24, 87). The plant-specific Pol IV and Pol V, which are dispensable for viability, are most closely related to Pol II (87, 115, 131). Similarly to the other three RNAPs, Pol IV and Pol V have unique largest subunits, termed NUCLEAR RNA POLYMERASE D1 (NRPD1) and NRPE1, respectively; in *Arabidopsis*, they share the same second-largest subunit, NRPD2/E2 (53, 66, 107, 114) (**Supplemental Table 1**; follow the **Supplemental Material link** from the Annual Reviews home page at **http://www.annualreviews.org**). Further diversification of Pol IV and Pol V from Pol II and from each other has occurred through duplications or alterations of subunits 3, 4, 5, 7, and 9 (115, 131) (**Figure 2**).

Supplemental Material

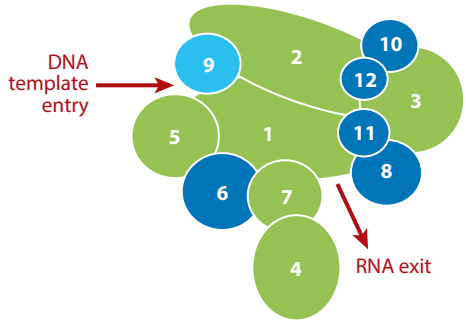


Figure 2

Subunit composition of Pol IV and Pol V. Subunits are numbered and arranged according to information available for yeast Pol II (141). Number 1 refers to the largest subunit. The subunits shared by Pol II, Pol IV, and Pol V are shown in dark blue, and the subunits unique to Pol IV and/or Pol V are shown in green (115, 131). A ninth subunit, shown in light blue, that is common among the three RNA polymerases carries out a special Pol V regulatory function (37, 128). The largest and second-largest subunits interact to form the catalytic core for DNA-dependent RNA polymerization (131). All four subunits comprising the jaw region (1, 2, 5, and 9), which grips the DNA template during transcription (25), have been identified in genetic screens for Pol V–specific factors (37), suggesting that specializations in this region adapt Pol V to transcribe templates with particular structural or epigenetic features. Regions of DNA template entry and RNA exit are placed according to those in Reference 131.

Pol IV and Pol V are evolving rapidly, as indicated by the 10- and 20-times-higher amino acid substitution rates of their second-largest and largest subunits, respectively, compared with the cognate subunits in Pol II (87). The multistep scenario originally proposed by Luo & Hall (87) for the evolution of the largest and second-largest subunits of Pol IV and Pol V has been largely confirmed and extended by more recent phylogenetic analyses. Initially, in the ancestor of Charales, the closest green algal relative of land plants, the gene encoding NRPB1, the largest subunit of Pol II, was duplicated and one copy diversified to give rise to NRPD1. After the divergence of land plants from Charales, the second-largest subunit, NRPD2/E2, arose, presumably through a reverse transcription event because the intron positions in NRPB2 do not correspond with those in NRPD2/E2. Finally, in the ancestor of flowering plants, NRPD1 was duplicated and one copy diversified to form NRPE1 (87). NRPD1 and NRPE1 have several common features that reveal their relatedness to each other and their divergence from NRPB1 to support specialized functions in RdDM.

NRPD1 and NRPE1

The common ancestry of NRPD1 and NRPE1 and features distinguishing them from NRPB1 are evident both in the N-terminal segments, which contain the catalytic domains, and in the C-terminal domains (CTDs), which serve as platforms for binding various factors. The N-terminal portions comprise conserved domains A–H, which organize the active site and mediate interactions with proteins (44) (**Figure 3*a***). Differences in these regions in NRPD1 and NRPE1 relative to the canonical largest-subunit structure include numerous amino acid substitutions and deletions, including a deletion of approximately 190 amino acids between domains F and G (87) (**Supplemental Figure 1**). In addition, there are several divergent amino acid residues around the aspartic acid (D) triad forming the metal A binding site at the catalytic center, which otherwise has an invariant sequence (NADFDGD) in other RNAPs (**Figure 3*a***). The functional significance of these changes

Supplemental Material

in the catalytic domains of NRPD1 and NRPE1 is unknown, but they may allow Pol IV and Pol V to accommodate different templates or carry out unconventional catalytic activities (75).

In contrast to the recognizable similarity of most domains in their N-terminal portions, NPRD1 and NRPE1 differ completely from NRPB1 in their CTDs. The CTD of NRPB1 consists exclusively of a heptapeptide repeat (**Figure 3*a***), which provides a platform for enzymes that process

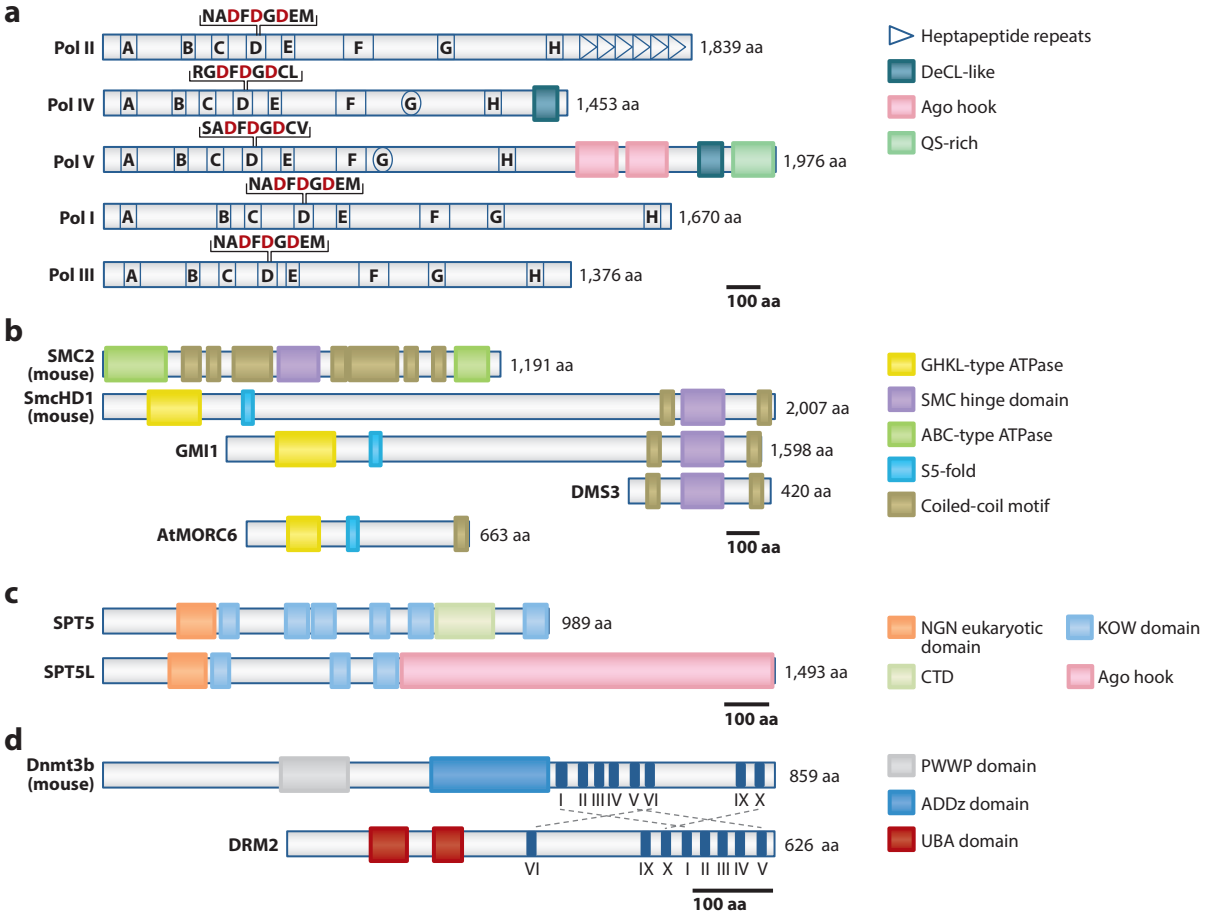


Figure 3

Domain structures of RNA-directed DNA methylation (RdDM) factors and related proteins. (*a*) RNA polymerase largest subunits. Conserved domains A–H are important for organizing the active center and mediating interactions with other subunits (44). The circled G domain in NRPD1 and NRPE1 indicates the approximate position of a 190-base-pair deletion relative to NRPB1, which is shown in detail in **Supplemental Figure 1** together with additional sequence variations among the three largest subunits. The amino acid sequence of the loop region containing the metal A binding site with the conserved aspartic acid (D) triad (*red letters*) is bracketed. There are 26 copies of the heptapeptide Tyr-Ser-Pro-Thr-Ser-Pro-Ser in the C-terminal domain (CTD) of NRPB1 in *Saccharomyces cerevisiae*, 29 in that of *Arabidopsis*, and 52 in that of humans (34). (*b*) Domain structures of SMC2, SmcHD1, GMI1, DMS3, and AtMORC6. (*c*) Domain structures of SPT5 and SPT5L. (*d*) Comparison between mammalian Dnmt3b and plant DRM2. Domains in the C-terminal region are indicated in roman numerals. The PWWP domain has a conserved Pro-Trp-Trp-Pro motif. Additional abbreviations not defined in the main text: aa, amino acids; ADDz, ATRX, Dnmt3, and Dnmt3L PHD-like zinc finger domain; ATRX, alpha thalassemia/mental retardation syndrome X; Dnmt3, DNA methyltransferase 3; Dnmt3L, DNA methyltransferase 3–like; PHD, plant homeodomain–like zinc finger domain; Q, glutamine; S, serine; UBA, ubiquitin associated.

Supplemental Material

the nascent RNA and catalyze cotranscriptionally various epigenetic modifications (34). By contrast, the CTD of NRPD1 lacks heptapeptide repeats and contains instead a sequence similar to a group of small plant-specific proteins called defective chloroplasts and leaves (DeCLs) (**Figure 3*a***). The functions of authentic DeCL proteins, which are located in mitochondria, chloroplasts, and nuclei, are not entirely clear, but they are thought to participate in ribosomal RNA processing (7, 74). NRPD1 thus appears to be the result of a fusion between sequences encoding the N-terminal portion of NRPB1 and a DeCL protein. The function of the DeCL motif in NRPD1 is not known, but, similarly to authentic DeCL proteins, it may be involved in RNA processing or interactions.

The derivation of NRPE1 from NRPD1 is supported by the presence of a DeCL motif in the CTDs of both proteins. However, the CTD of NRPE1 has further expanded to include a so-called Argonaute (Ago)–hook region comprising nearly 20 GW repeats (5, 35) (**Figure 3*a***). The Ago hook allows Pol V to cotranscriptionally associate with AGO proteins that bind the siRNA guide for methylation (5) (**Figure 1**). The origin of the GW repeats in the CTD of NRPE1 is not clear, but the presence of a degenerate tandem repeat (44) is a feature shared by the Ago-hook region of SPT5L, another component of the Pol V pathway that is discussed further below.

Phylogenetic distribution. As mentioned above, the Pol IV pathway appears to be present in all land plants investigated so far, and this is consistent with the existence of putative orthologs of NRPD1 in a range of land plants (**Figure 4**). As expected from previous analyses (87, 131), NRPE1 is present in all flowering plants examined so far, including the basal angiosperm *Amborella trichopoda* (**Figure 4**), for which the whole genome sequence has recently become available (2). However, contrary to an earlier scenario suggesting that NRPE1 originated in the ancestor of flowering plants (87), putative NRPE1 orthologs appear to be present in the lycophyte *Selaginella moellendorffii* and the moss *Physcomitrella patens* (**Figure 4**). Intriguingly, the putative NRPE1 protein in *Physcomitrella* contains an Ago-hook region in the CTD but lacks a recognizable DeCL domain, which is also absent in the putative NRPD1 ortholog in this species (**Figure 4**). Moreover,

	NRPD1	NRPE1	RDM1	DRD1	DMS3	AtMORC6	SPT5L	DRM2
Arabidopsis thaliana								
Oryza sativa								
Amborella trichopoda								No UBA
Picea abies		No Ago hook No DeCL						No UBA
Selaginella moellendorffii								No UBA
Physcomitrella patens	No DeCL	No DeCL						No UBA
Chlamydomonas reinhardtii								

Figure 4

Phylogenetic distribution of specialized Pol V pathway components. Orange and white boxes denote the existence and absence, respectively, of a putative ortholog for each RNA-directed DNA methylation factor in each species. Light orange boxes in the NRPD1, NRPE1, and DMS3 columns indicate the presence of a similar protein that either lacks characteristic motifs (NRPD1 and NRPE1) or is not the correct size (DMS3). Domains were identified by searching PROSITE (**http://prosite.expasy.org**), InterPro (**http://www.bi.ac.uk/interpro**), and SMART (**http://smart.embl-heidelberg.de**). Accession numbers for these putative orthologs are given in **Supplemental Table 1**.

Supplemental Material

Supplemental Material

the amino acids around the metal A binding site of the catalytic center of the putative NRPE1 ortholog in *Physcomitrella* are identical to the invariant NADFDGD sequence found in Pol I, Pol II, and Pol III instead of the (S/G/A)ADFDGD sequence that is present in NRPE1 from other plants (**Supplemental Figure 2**). By contrast, in *Selaginella*, the putative orthologs of NRPD1 and NRPE1 appear to be more similar to those in flowering plants, containing a DeCL motif in the CTDs as well as an Ago-hook region in the CTD of NRPE1 (**Figure 4**). However, the amino acids around the metal A binding sites in *Selaginella* NRPD1 and NRPE1 deviate from the conserved sequences (**Supplemental Figure 2**).

Any functional consequences of the DeCL motif deficit in *Physcomitrella* NRPE1 and NRPD1 will remain unknown until the function of the DeCL motifs in Pol IV and Pol V is determined. The presence of putative NRPE1 orthologs and the absence of the specialized Pol V pathway components RDM1, DMS3, and SPT5L (**Figure 4**), which are discussed further below, in *Physcomitrella* and *Selaginella* suggests the presence of a rudimentary Pol V–mediated RdDM pathway in these lower plants. In accordance with this notion, *Physcomitrella* has DCL3-generated 22–24-nt siRNAs that are involved in retrotransposon silencing, although they are unlikely to directly trigger DNA methylation (22). *Physcomitrella* may represent either a separate pathway of Pol V evolution or an arrested early stage before the acquisition of the full complement of NRPE1 domains and Pol V auxiliary factors that are present in flowering plants.

Gymnosperms. Whether the Pol V pathway is present in gymnosperms is unclear, because a search of the Norway spruce (*Picea abies*) genome database identified only a truncated NRPE1 protein (**Supplemental Table 1**). However, this apparent absence may reflect incomplete annotation of the spruce genome. Whether the putative partial NRPE1 ortholog in spruce, which lacks an obvious DeCL motif and Ago-hook region that are present in the CTDs of the angiosperm NRPE1 (**Figure 4**), can function in the RdDM pathway is unknown. Although gymnosperms possess a Pol IV pathway, 24-nt siRNAs are produced at low levels compared with those of flowering plants and are highly specific to reproductive tissues (106). More analysis of DNA methylation in gymnosperms is required to assess the extent to which RdDM contributes to genome methylation.

Specialized Pol IV and Pol V Pathway Components

In addition to significant modifications in the largest subunits, NRPD1 and NRPE1, the evolution of the Pol IV and Pol V pathways has entailed the emergence of a new subfamily of chromatin remodelers, the DRD1 subfamily, as well as altered and novel proteins. The Pol V branch of the pathway, which is responsible for facilitating the de novo methylation step (**Figure 1**), is particularly complex and relies on several unusual factors, including DMS3, RDM1, and SPT5L, that appear to be present only in flowering plants.

The DRD1 subfamily of Snf2 chromatin-remodeling proteins. DRD1 is the founding member of a subfamily of Snf2 chromatin-remodeling proteins that is present only in land plants (6) (**Figure 4**). Snf2 proteins, which are involved in replication, transcription, DNA repair, and recombination, use the energy of ATP to locally modulate chromatin by displacing nucleosomes and disrupting DNA-histone contacts (41). The *Arabidopsis* genome encodes 41 Snf2 proteins distributed among six groups and 18 subfamilies (6). DRD1, part of the Rad54-like group (41), is the only plant-specific Snf2 subfamily in *Arabidopsis*.

The DRD1 subfamily contains six members in *Arabidopsis*, and nearly all have been implicated in either Pol IV– or Pol V–dependent steps of the RdDM pathway (**Supplemental Table 1**), presumably acting to remodel chromatin in advance of Pol IV or Pol V transcription. DRD1

was identified in several classical genetic screens for compromised Pol V function in RdDM (43, 64, 67) and shown to be required for production of Pol V scaffold transcripts (139). By contrast, CLASSY 1 (CLSY1) is needed for Pol IV activity, as indicated by its identification in a genetic screen that also retrieved NRPD1 and RDR2 (123) and in a mass spectrometry analysis of affinity-purified NRPD1 (80). The latter study also identified further DRD1 subfamily members, namely CLSY2, CLSY3, and CLSY4 (**Supplemental Table 1**), as NRPD1-interacting proteins (80). An ortholog in maize, required to maintain repression 1 (RMR1), which is most similar to CLSY3 and CLSY4 (6), was identified in a screen for mutants defective in paramutation (46). Of the *Arabidopsis* DRD1 subfamily members, only chromatin remodeling 34 (CHR34), which is most closely related to DRD1, remains functionally uncharacterized (**Supplemental Table 1**).

Supplemental Material

The DRD1 subfamily has a notable evolutionary history, progressing from "apparent nonexistence in non-plant species and lower plants, such as *Volvox carteri* and *Chlamydomonas reinhardtii*, to the largest and most diverse subfamily in current-day higher plants" (6, p. 7) (**Figure 4**). An analysis of the large DRD1 subfamily in tomato revealed a complex series of evolutionary events leading to substantial functional diversification (6). The reasons behind the rapid expansion and functional variation of the DRD1 subfamily remain obscure. Their prominent role in siRNA-mediated epigenetic processes implies that this form of gene regulation has adaptive value in plants, which is consistent with proposed roles of RdDM in stress responses (6, 96) and reproduction (93). Additional work is required to unravel the full range of functions that require DRD1 subfamily members.

The mechanism or mechanisms by which DRD1 subfamily proteins operate during the various steps of the RdDM pathway are not known. The DRD1 protein is currently assigned a role in the DDR complex (**Figure 1**), a biochemically defined complex containing DRD1, DMS3, and RDM1 that may assist Pol V transcription through chromatin (77, 139, 140) (**Figure 1**). More refined roles for DRD1 in the mechanism of Pol V–mediated de novo methylation might emerge from deeper studies. Other members of the Rad54-like Snf2 subfamily could offer clues to the mechanism of action of DRD1 and related proteins. The Rad54 protein functions in homologous recombination and DNA repair, presumably by stimulating Rad51-mediated invasion of single-stranded DNA into the target duplex (41). Another Rad54-like protein, J-binding protein 2 (JBP2), facilitates telomeric silencing in trypanosomes by facilitating a process involving DNA glycosylation (29). A recurrent theme is that members of the Rad54-like subfamily of Snf2 proteins are involved in repair or covalent modification of DNA.

DMS3: an SMC hinge domain–containing protein. DMS3 is an unusual structural maintenance of chromosomes (SMC) hinge domain–containing protein that has been identified in several classical genetic screens for mutants defective in RdDM (4, 64). Authentic SMC proteins contain a central hinge domain flanked by two long coiled coils, which terminate in ATP-binding sites (**Figure 3*b***). Unlike Snf2 ATPases, which act locally to modify chromatin structure, SMC proteins, which contain an ATP-binding cassette (ABC) type of ATPase, are involved in modulating higher-order chromosome organization and dynamics. There are six SMC proteins in eukaryotes, which act as heterodimers, joined at the central hinge region that is present in these molecules (17). SMC1 and SMC3 are core subunits of the cohesion complex that reversibly embraces sister chromatids. SMC2 and SMC4 form the core of the condensin complex, which is important for chromosome condensation. The SMC5 and SMC6 heterodimer is involved in DNA repair and checkpoint responses (17).

DMS3 is essentially a solo hinge protein that is much smaller than authentic SMC proteins (**Figure 3*b***). Clear homologs are found only in flowering plants (**Figure 4**). Interestingly, the hinge region of DMS3 is less similar to the hinge region of authentic SMC proteins and more similar to

the hinge region of another *Arabidopsis* protein, γ-irradiation and mitomycin C induced 1 (GMI1) (12), and a related mammalian protein, structural maintenance of chromosomes hinge domain–containing 1 (SmcHD1) (10). However, the latter two proteins are considerably larger than DMS3, and in addition to the SMC hinge contain a terminal ATPase domain of the GHKL (gyrase, HSP90, histidine kinase, MutL) type (**Figure 3*b***). GHKL ATPases of the MORC subfamily have also been identified together with DMS3 in classical genetic screens to identity mutants defective in RdDM and TGS in *Arabidopsis*.

MORC ATPases. The founding member of the MORC family was identified in a mouse mutant impaired in spermatogenesis (57). MORCs represent a novel family of eukaryotic chromatin proteins that contain a GHKL domain and an S5 domain, which together constitute an active ATPase module (59, 76). Plant MORCs, which are distantly related to mammalian MORCs (76), have a C-terminal coiled-coil domain that forms a putative basic leucine zipper motif, which may mediate protein-protein interactions (76) (**Figure 3*b***). Until recently, little was known about the roles of MORC ATPases in chromatin structure and dynamics. Based on the functions of prokaryotic GHKL ATPases, such as gyrases and topoisomerases, Iyer et al. (59) proposed that MORC ATPases in eukaryotes may influence chromatin by manipulating DNA superstructure in response to epigenetic cues such as DNA methylation and histone modifications.

MORC proteins have been implicated recently in RdDM and heterochromatin silencing, but their precise roles are not clear. Of the seven MORC proteins in *Arabidopsis* (**Supplemental Table 1**), AtMORC6, which is most distantly related to the other six (62, 76), was identified in two independent classical genetic screens for mutants disabled in transgene TGS and RdDM (14, 86). Both AtMORC1 and AtMORC6 were identified in a separate screen for heterochromatin factors (102). AtMORC1 was also retrieved and named compromised resistance to turnip crinkle virus (CRT1) in an unrelated classical genetic screen for mediators of defense signaling (62). AtMORC1/CRT1 and its closest paralog, AtMORC2, were shown to be required for various types of disease resistance (61–63). Additional work is needed to understand the relationship between the roles of MORC proteins in epigenetic processes and responses to pathogens.

Supplemental Material

Regarding the role of MORC proteins in epigenetic modifications, several studies have reported modest decreases in DNA methylation and repressive histone modifications in *morc6* mutants (14, 85, 86). A separate investigation, however, did not detect changes in either DNA methylation or histone methylation in *morc1* and *morc6* mutants but did discern alterations in higher-order chromatin conformation at pericentromeric regions, suggesting that MORC1 and MORC6 acted downstream of DNA and histone modifications to reinforce TGS (102). AtMORC6 has been reported to interact with the SMC hinge domain–containing protein DMS3, suggesting that the two proteins cooperate as a cohesion-like complex to modulate chromatin configuration (86, 111) (**Figure 1**). However, other biochemical studies did not find a stable interaction between AtMORC6 and DMS3 but did detect associations between MORC proteins and SUVH2/9 (**Figure 1**) (84). AtMORC1 and AtMORC2 have been reported to act nonredundantly to form heteromers with AtMORC6 in modulating higher-order chromatin structure (101) (**Figure 1**). Different combinations of MORC proteins may have locus-specific effects on chromatin structure or act at different steps of RdDM or immune pathways.

GHKL ATPases, SMC hinge domains, DNA repair, and RNA-directed DNA methylation. Although MORCs are present in all land plants, clear orthologs of the SMC hinge domain–containing protein DMS3 are found only in flowering plants, and their evolutionary origin is unclear (**Figure 4**). DMS3 and AtMORC6, which is in a distinct clade relative to the other *Arabidopsis* MORCs (76), display highly correlated coexpression in *Arabidopsis* and were shown in

one study to interact physically to stimulate ATPase activity (85). It is therefore worth considering the mechanistic implications of a functional association between these two proteins, particularly in view of mammalian SmcHD1 and *Arabidopsis* GMI1, which contain both the SMC hinge and the GHKL ATPase domain in a single polypeptide (**Figure 3*b***). Conceivably, DMS3 and a MORC-like protein might have resulted from transposon-mediated fracturing of a *GMI1* paralog into two separate genes (12). GMI1 is needed for somatic homologous recombination and DNA repair in *Arabidopsis* (12). Mammalian SmcHD1 has both epigenetic and DNA repair functions and has been proposed to promote homologous DNA repair through its ability to alter chromatin states (23). It thus appears that proteins comprising both an SMC hinge and GHKL ATPase are important for DNA repair, including homology-directed repair. Plant MORC proteins also have DNA/RNA-binding capacity and endonuclease activity in vitro and hence may be involved in DNA recombination and repair (76). Human MORC2 is involved in DNA repair–associated chromatin remodeling to permit DNA repair (82). Thus, in addition to DRD1, the putative Snf2 chromatin remodeler in the Rad54-like subfamily, DMS3 and MORCs represent other Pol V pathway components that are related to proteins functioning in DNA repair.

RDM1: a small protein with a unique protein fold. RDM1 is the most enigmatic protein in the Pol V pathway with respect to its precise roles in the mechanism of RdDM and its evolutionary history. RDM1 is a small plant-specific protein of 163 amino acids that has a unique protein fold and a domain of unknown function (DUF1950). Its crystal structure, which was determined before any functional information was available, revealed a new fold in protein conformation space (1). The structural analysis also suggested that RDM1 forms homodimers and that each monomer contains a hydrophobic pocket. Sequence homology searches have failed to identify any related protein in organisms other than flowering plants (**Figure 4**). The evolutionary origin of RDM1 is thus a mystery, although a tentative relationship to a DNA polymerase subunit has been noted (118). As indicated by the nearly complete loss of DNA methylation in *rdm1* mutants (42, 118, 126), RDM1 plays a crucial role in the RdDM pathway that remains to be fully understood.

RDM1 has been placed in two positions in the Pol V pathway: acting as a bridging protein between AGO4 and DRM2 (42) and acting as a component of the DDR complex that facilitates Pol V transcription (77) (**Figure 1**). It is not clear how to reconcile these two proposed roles. An analysis of several point mutations that prevent dimerization suggested that RDM1 acts as a monomer in the DDR complex and as a homodimer during the de novo methylation step (118). This may be relevant in view of the reported interaction of RDM1 and DRM2 (42), which also forms a dimer that is critical for its catalytic activity (155). The role of the hydrophobic pocket in RDM1 is obscure, although it was observed to bind methylated single-stranded DNA (42) (**Figure 1**). RDM1 homodimers, which contain two opposing hydrophobic pockets (1), could conceivably bind methylated double-stranded nucleic acids, although this possibility awaits experimental confirmation.

SPT5L. SPT5L [also termed KOW TRANSCRIPTION FACTOR 1 (KTF1)] is a plant-specific derivative of the transcription factor Spt5, known as N-utilization substance G (NusG) in bacteria. Other than core RNAP subunits, the Spt5/NusG family of proteins are the only transcriptional regulators that are conserved in all three domains of life—eukaryotes, bacteria, and archaea—and they have essential functions in transcriptional elongation and other stages of transcription (47). In eukaryotes, Spt5 proteins coordinate transcriptional elongation with chromatin modifications and pre-mRNA processing (47). Spt5 is a large, highly conserved protein containing a NusG N-terminal (NGN) domain—a CTD that binds the zinc finger protein Spt4 to form a functional

complex—and multiple Kyrpides, Ouzounis, and Woese (KOW) domains, which may mediate protein-protein and protein–nucleic acid interactions (47) (**Figure 3*c***).

Supplemental Material

The *Arabidopsis* genome encodes two authentic Spt5 proteins, the pollen-specific SPT5-1 and the constitutively expressed SPT5-2 (32), as well as the related SPT5L (**Supplemental Table 1**). Similar to authentic Spt5 proteins, SPT5L contains in its N-terminal half an NGN domain and several KOW domains. The C-terminal half, which diverges completely from Spt5 proteins, contains an Ago-hook region comprising many GW repeats (**Figure 3*c***). Thus, compared with Spt5 proteins, SPT5L has gained the ability to interact with AGO proteins and participate in siRNA-mediated chromatin modifications that are imposed during transcription.

A role for SPT5L in RdDM was discovered following a computational analysis conducted to reveal Ago-hook proteins (8) and in a classical genetic screen designed to retrieve RdDM-defective mutants (49). Both of these studies demonstrated that SPT5L interacts with AGO4. SPT5L was also identified in a mass spectrometry analysis of affinity-purified NRPE1 (55). These results are consistent with the participation of SPT5L in the Pol V pathway as an AGO4 and NRPE1-associated protein (55, 116) (**Figure 1**).

Full-length homologs of SPT5L are present only in flowering plants (**Figure 4**). Among these proteins, the N-terminal half is far more conserved than the C-terminal half, which contains the Ago-hook region. In different species, this tract varies not only in nucleotide sequence and the number of GW repeats but also in length. The functional consequences of these differences are unknown, but they may affect the strength of interactions with AGO proteins. The Ago-hook region of SPT5L in *Arabidopsis*, many other dicots, and some monocots contains a degenerate tandem repeat (**Supplemental Figure 3**), which may lead to slippage during DNA replication and result in the relatively high degree of length and sequence variability in this region. A degenerate tandem repeat of 16 amino acids is also present in the Ago-hook region of NRPE1 (44), and this region is also highly variable among NRPE1 orthologs. SPT5L and NRPE1 illustrate the propensity of transcriptional proteins to acquire, through a mechanism that may involve exon capture or alternative splicing (5), rapidly evolving Ago-hook regions that are critical for their function in RdDM.

Summary of Changes

Major changes in the evolutionary transition from the Pol II transcriptional pathway to the current Pol IV/Pol V pathway in flowering plants occurred through duplication and modification of Pol II subunits, including a gene fusion involving NRPB1 and a DeCL protein (NRPD1), acquisition of Ago-hook regions by proteins involved in transcription (NRPE1 and SPT5L), co-option and modification of proteins involved in DNA repair (DRD1 subfamily members, DMS3, and MORCs), and the emergence of a novel protein that may be derived from a nucleic acid polymerase subunit (RDM1). Although their precise roles in the RdDM mechanism remain to be fully clarified, these refashioned and novel proteins are likely to carry out their specialized roles by facilitating Pol IV and Pol V transcription through chromatin, establishing a chromatin substrate favorable for de novo methylation during Pol V transcription, and enabling interaction of the Pol V complex with AGO/siRNAs and DRM2 to effect de novo methylation of DNA (**Figure 1**).

DRM2

DRM2, a member of the conserved Dnmt3 family of de novo methyltransferases, has special features that may be relevant for its role in RdDM (15, 43, 72, 104). Most strikingly, the catalytic domains in the C-terminal portion are rearranged relative to mammalian counterparts such that

motifs VI–X precede I–V (16) (**Figure 3*d***). The functional consequences of this rearrangement remain uncertain, but it does not seem to affect the ability of DRM2 to adopt the overall fold of conventional DNA methyltransferases (155). DRM2 forms a dimer that is essential for catalytic activity in vivo (155) and has been reported to interact with RDM1 (42), AGO4 (155), and lncRNAs (11) (**Figure 1**). One can speculate that the domain rearrangement is in some way important for DRM2 to participate in one or more of these proposed interactions in the Pol V pathway.

The rearrangement in DRM2 occurs at the same position in many plant species, suggesting that it occurred early during plant evolution (155). DRM2 homologs are present in land plants (148), although not all of these contain the N-terminal ubiquitin-associated (UBA) domains found in most flowering plant DRM2 proteins (**Figure 4**). The UBA domains are not required for catalytic activity (155), but they may recognize certain chromatin modifications during RdDM (50). *Chlamydomonas* lacks a DRM2 homolog and has only low levels of nuclear DNA methylation, which probably results from a mechanism different from that in flowering plants (39).

UNRESOLVED ISSUES

How Pervasive Is Pol II Involvement in RNA-Directed DNA Methylation?

During RNAi-mediated heterochromatin formation in fission yeast, Pol II transcribes both the precursor of siRNAs and scaffold transcripts that base pair with the siRNA guide and recruit the silencing effector complex (69). Although Pol IV and Pol V are specialized to execute these steps in plants, Pol II is capable of performing both of these functions in pathways leading to RdDM at certain loci. In addition to transcribing precursors of 21–22-nt siRNAs in RDR6-RdDM pathways (90, 105, 108, 113), Pol II transcribes hairpin RNAs that are processed to siRNAs from endogenous inverted DNA repeats (31, 52). At intergenic, low-copy-number loci, Pol II synthesizes scaffold transcripts and interacts with AGO4/siRNAs, presumably through several GW repeats in the second-largest subunit, NPRB2, leading to application of repressive epigenetic modifications that induce TGS (154). At some heterochromatic loci, Pol II transcripts or transcription can recruit Pol IV and/or Pol V to elicit TGS (154). Pol II transcription has been reported to promote Pol IV–dependent siRNA amplification and RdDM at a low-copy-number transgene locus (145). Further work is required to understand the full manner in which Pol II either acts independently or collaborates with Pol IV and Pol V to induce RdDM at various sites in the genome. Conceivably, Pol II transcription could be associated with the application of histone modifications—as occurs in fission yeast, which lacks DNA methylation—whereas Pol V transcription would allow establishment of covalent modification of DNA.

How Many RNA Triggers Are There?

Although the current consensus holds that DCL3-generated 24-nt siRNAs are the trigger for RdDM, the finding of RDR6-RdDM and other observations suggest that this view is incomplete. Recent whole-genome bisulfite sequencing on a large collection of epigenetic mutants (126) has reinforced previous data showing considerable redundancy among DCLs, and presumably different size classes of siRNA, in RdDM (52). Whole-genome bisulfite sequencing has demonstrated that the level of methylation in *dcl3* mutants is only weakly reduced, whereas methylation in *dcl2 dcl4 dcl3* triple mutants is more strongly decreased, indicating that DCL2 and DCL4 can compensate for lack of DCL3 activity at most RdDM target sites (126). In addition, mutants defective in RDR1 and RDR6 show strong losses of methylation even though they are involved in producing 21–22-nt siRNAs (126). These results are consistent with the idea that that RDR6-DCL2/4-dependent

21–22-nt siRNAs act to establish DNA methylation that is then further maintained and enhanced by Pol IV–RDR2–DCL3–dependent 24-nt siRNAs (90, 93, 105, 108, 113).

Dalakouras & Wassenegger (26) suggested that longer dsRNAs trigger RdDM. In their model, AGO4-bound siRNAs are not needed to guide RdDM but are necessary to cleave Pol II–, Pol IV–, or Pol V–generated scaffold RNAs to provide substrates for RDR2/6, producing dsRNAs that directly induce methylation. Their arguments rely in part on nonuniform accumulation of small RNAs from processed hairpin RNAs. However, these may be explained by siRNA accumulation biases based on sequence composition (117).

siRNA-RNA or siRNA-DNA?

Early debates considered whether the strict limitation of RdDM to the region of RNA-DNA sequence homology reflected base pairing of siRNAs with target DNA or with nascent RNA transcribed at the target site (91). The detection of Pol V transcripts at target loci (139) that resemble the Pol II scaffold transcripts required for siRNA-mediated heterochromatin formation in fission yeast appeared to settle the matter in favor of interactions between siRNAs and Pol V–generated scaffold RNAs. In support of this model, a recent study noted that DRM2 tends to methylate the same DNA strand as the siRNA instead of the complementary strand that can base pair to siRNAs (155). However, there is no direct evidence that Pol V transcripts per se are required as opposed to the act of Pol V transcription itself, which may provide a permissive chromatin environment for RdDM to occur independently of the lncRNA product (71). The exquisite targeting of RdDM (27, 110, 117) is most consistent with direct siRNA-DNA base pairing (134). The unique transcriptional machinery required for RdDM, including factors related to proteins involved in homologous recombination and DNA repair, also suggests a mechanism that differs from that operating in fission yeast.

Although AGO proteins are normally thought to bind small RNAs that base pair to complementary target RNAs, these proteins are also able to bind small DNAs, as indicated by the recent finding of DNA-guided DNA interference in prokaryotes. In this process, AGO proteins loaded with small, plasmid-derived DNAs 15–25 nt in length guide cleavage of complementary plasmid DNA in a host defense response (127). It is thus conceivable that during RdDM, siRNA-bound AGO interacts directly with DNA. lncRNAs can interact with the ribosomal DNA promoter in mammalian cells to create an RNA:DNA triplex that is recognized by the de novo methyltransferase Dnmt3 (120).

Alternate Templates and Reactions?

Pol IV and Pol V can transcribe double-stranded DNA in vitro, but this activity is relatively weak compared with that of Pol II (45). The sequence divergence in the catalytic centers and clefts suggests that Pol IV and Pol V may be adapted to alternative templates and might either catalyze reactions other than nucleotidyl transfer or use a mechanism different from that of other RNAPs (75). Pol IV and Pol V have been proposed to transcribe heterochromatic templates that are resistant to transcription by other RNAPs (44, 53, 75). Similar to Pol II transcription of viroid RNA genomes (38), Pol IV and Pol V may also transcribe unconventional templates such as extrachromosomal dsRNA (133).

In addition to their ability to catalyze rapid nucleotide addition, RNAPs, including Pol IV and Pol V, possess hydrolytic endonuclease and exonuclease activity that relies on the same catalytic center required for the nucleotidyl transferase reaction (75). Thus, in principle, Pol IV and Pol V could act in RNA processing as unconventional nucleases (75). In this context, it is worth recalling

that both NRPD1 and NRPE1 contain in their CTDs a DeCL motif that is presumably derived from authentic DeCL proteins, which are believed to be involved in processing of ribosomal RNA (7). Although the role of the DeCL motif in NRPD1 and NRPE1 remains obscure, the conservation of this region in these proteins suggests an essential contribution to Pol IV and Pol V function.

Summary

Until the mechanism of RdDM at its entire range of targets is understood, it is wise to keep an open mind with respect to the nature of the trigger RNA and its nucleic acid pairing partner, the relative roles of Pol II and Pol V transcripts and/or transcription in RdDM at different loci, and the possibility that Pol IV and Pol V may have alternate templates and activities in siRNA-mediated gene silencing or other pathways.

SPECULATION ON THE IMPORTANCE OF RNA-DIRECTED DNA METHYLATION FOR FLOWERING PLANTS

The RdDM pathway has been implicated in transposon silencing, pathogen defense, stress responses, reproduction, and interallelic and intercellular communication (93). These traditional roles assume that RdDM acts in the expected manner to repress the transcription of transposons and transposon-associated genes. However, transposons are effectively silenced in other organisms without the need for additional RNAPs, elaborate transcriptional machineries, and methylation outside of the conventional CG context. Moreover, the fact that transposons are not unleashed in RdDM-defective mutants suggests that RdDM plays only a minor role in transposon control. RdDM, and in particular the Pol V branch of the pathway, has reached the pinnacle of complexity in flowering plants. It is therefore tempting to speculate that RdDM serves an as-yet-undefined function that has been especially important for this specific plant group.

Based on comparisons between gymnosperms and flowering plants, we consider here the possibility that RdDM, as a mechanism related to those used to silence transposons, emerged as a means to accelerate the diploidization of polyploid genomes. In this scenario, the true consequences of RdDM would become obvious only over an evolutionary timescale. Whole-genome duplications (polyploidy) have played a major role during flowering plant evolution (124). Newly formed polyploid genomes undergo a process of diploidization, which involves rapid as well as more gradual changes that include both sequence loss and diversification of duplicated genes (89). RdDM could conceivably contribute to these genome-reshaping events if the Pol V pathway is stimulated by the shock of genome doubling (95, 129) to more intensively target duplicated genes and transposons. Subsequent loss of these sequences could potentially occur through siRNA-mediated heterochromatin formation and DNA elimination, a process that removes transposon-like sequences from the somatic genome in ciliated protozoa (20). Sequence diversification of methylated, duplicated genes could be driven by the inherent mutagenicity of 5-methylcytosine, which spontaneously converts to thymine. The dense methylation potentially induced by RdDM in coding regions of multicopy genes (146) elevates the mutation rate by ensuring that virtually every cytosine is a possible candidate for conversion.

In contrast to the genomes of flowering plants, the large Norway spruce genome shows no evidence of whole-genome duplications and is laden with transposons, suggesting inefficient gene-silencing and purging mechanisms (100). Moreover, as discussed above, Norway spruce appears to lack a fully functional Pol V pathway. Drawing a connection between the presence of the Pol V pathway and the prevalence of polyploidy in flowering plants, we hypothesize that

Pol V–mediated RdDM enhances the diploidization process and in that way contributes to the remarkable evolutionary success of flowering plant lineages. The extent and consequences of Pol V activity following whole-genome duplication can be tested by generating polyploids in *pol v* mutant backgrounds and assessing genome-wide DNA methylation and sequence variation in subsequent generations.

SUMMARY POINTS

1. RNA-directed DNA methylation (RdDM) is a complex epigenetic pathway requiring two plant-specific, Pol II–related RNA polymerases termed Pol IV and Pol V as well as a number of specialized factors.
2. RdDM is targeted to specific regions of the genome by Pol IV–dependent small interfering RNAs (siRNAs) that interact with scaffold RNAs transcribed by Pol V at target loci.
3. In addition to the canonical RdDM pathway involving Pol IV–RDR2–DCL3–generated 24-nucleotide siRNAs, a noncanonical pathway involving Pol II–RDR6–DCL2/4–dependent 21–22-nucleotide siRNAs links posttranscriptional and transcriptional gene-silencing pathways.
4. Functional innovations driving the evolutionary transition from Pol II to Pol IV/Pol V transcriptional pathways include an apparent gene fusion, co-option and modification of genes involved in DNA repair, addition of Ago-hook domains to existing transcriptional proteins, and the emergence of a novel protein.
5. Specialized components of the Pol V pathway appear to be present only in flowering plants and are important for forging interactions with AGO proteins that bind siRNAs and generating chromatin states favorable for Pol V transcription and de novo DNA methylation.
6. Pol II is increasingly recognized as an important participant in both siRNA-dependent and siRNA-independent pathways of DNA methylation.
7. In view of the striking amino acid sequence divergence in their catalytic centers relative to Pol II, Pol IV and Pol V may accommodate unusual templates and carry out alternative catalytic activities.
8. In addition to contributions to transposon silencing, Pol V–mediated RdDM may have unique roles in flowering plants that are evident only over an evolutionary timescale.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Work in the Matzke lab is supported by Academia Sinica and the Taiwan Ministry of Science and Technology (NSC Project number MOST 103-2311-B-001-004-MY3). We thank Wen-Dar Lin for help in preparing **Supplemental Figure 1**.

Supplemental Material

LITERATURE CITED

1. Allard ST, Bingman CA, Johnson KA, Wesenberg GE, Bitto E, et al. 2005. Structure at 1.6 Å resolution of the protein from gene locus At3g22680 from *Arabidopsis thaliana*. *Acta Crystallogr. F* 61:647–50
2. *Amborella* Genome Proj. 2013. The *Amborella* genome and the evolution of flowering plants. *Science* 342:1241089
3. Aufsatz W, Mette MF, van der Winden J, Matzke M, Matzke AJM. 2002. HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. *EMBO J.* 21:6832–41
4. Ausin I, Mockler TC, Chory J, Jacobsen SE. 2009. IDN1 and IDN2 are required for *de novo* DNA methylation in *Arabidopsis thaliana*. *Nat. Struct. Mol. Biol.* 16:1325–27
5. Azevedo J, Cooke R, Lagrange T. 2011. Taking RISCs with Ago hookers. *Curr. Opin. Plant Biol.* 14:594–600
6. Bargsten JW, Folta A, Mlynárová L, Nap JP. 2013. Snf2 family gene distribution in higher plant genomes reveals DRD1 expansion and diversification in the tomato genome. *PLOS ONE* 8:e81147
7. Bellaoui M, Gruissem W. 2004. Altered expression of the *Arabidopsis* ortholog of DCL affects normal plant development. *Planta* 219:819–26
8. Bies-Etheve N, Pontier D, Lahmy S, Picart C, Vega D, et al. 2009. RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep.* 10:649–54
9. Blevins T, Pontvianne F, Cocklin R, Podicheti R, Chandrasekhara C, et al. 2014. A two-step process for epigenetic inheritance in *Arabidopsis*. *Mol. Cell* 54:30–42
10. Blewitt ME, Gendrel AV, Pang Z, Sparrow DB, Whitelaw N, et al. 2008. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.* 40:663–69
11. Böhmdorfer G, Rowley MJ, Kuciński J, Zhu Y, Amies I, Wierzbicki AT. 2014. RNA-directed DNA methylation requires stepwise binding of silencing factors to long non-coding RNA. *Plant J.* 79:181–91
12. Böhmdorfer G, Schleiffer A, Brunmeir R, Ferscha S, Nizhynska V, et al. 2011. GMI1, a structural-maintenance-of-chromosomes-hinge domain-containing protein, is involved in somatic homologous recombination in *Arabidopsis*. *Plant J.* 67:420–33
13. Bologna NG, Voinnet O. 2014. The diversity, biogenesis, and activities of endogenous silencing small RNAs in *Arabidopsis*. *Annu. Rev. Plant Biol.* 65:473–503
14. Brabbs TR, He Z, Hogg K, Kamenski A, Li Y, et al. 2013. The stochastic silencing phenotype of *Arabidopsis morc6* mutants reveals a role in efficient RNA-directed DNA methylation. *Plant J.* 75:836–46
15. Cao X, Aufsatz W, Zilberman D, Mette MF, Huang MS, et al. 2003. Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr. Biol.* 13:2212–17
16. Cao X, Springer NM, Muszynski MG, Phillips RL, Kaeppler S, Jacobsen SE. 2000. Conserved plant genes with similarity to mammalian *de novo* DNA methyltransferases. *PNAS* 97:4979–84
17. Carter SD, Sjögren C. 2012. The SMC complexes, DNA and chromosome topology: right or knot? *Crit. Rev. Biochem. Mol. Biol.* 47:1–16
18. Castel SE, Martienssen RA. 2013. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat. Rev. Genet.* 14:100–12
19. Cecere G, Grishok A. 2014. A nuclear perspective on RNAi pathways in metazoans. *Biochim. Biophys. Acta* 1839:223–33
20. Chalker DL, Yao MC. 2011. DNA elimination in ciliates: transposon domestication and genome surveillance. *Annu. Rev. Genet.* 45:227–46
21. Chan SW, Zilberman D, Xie Z, Johansen LK, Carrington JC, Jacobsen SE. 2004. RNA silencing genes control de novo methylation. *Science* 303:1336
22. Cho SH, Addo-Quaye C, Coruh C, Arif MA, Ma Z, et al. 2008. *Physcomitrella patens* DCL3 is required for 22–24 nt siRNA accumulation, suppression of retrotransposon-derived transcripts, and normal development. *PLOS Genet.* 4:e10000314
23. Coker H, Brockdorff N. 2014. SMCHD1 accumulates at DNA damage sites and facilitates the repair of DNA double-strand breaks. *J. Cell Sci.* 127:1869–74

24. Cramer P, Armache KJ, Baumli S, Benkert S, Brueckner F, et al. 2008. Structure of eukaryotic RNA polymerases. *Annu. Rev. Biophys.* 37:337–52
25. Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B, et al. 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288:640–49
26. Dalakouras A, Wassenegger M. 2013. Revisiting RNA-directed DNA methylation. *RNA Biol.* 10:453–55
27. Daxinger L, Kanno T, Bucher E, van der Winden J, Naumann U, et al. 2009. A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. *EMBO J.* 28:48–57
28. Deleris A, Greenberg MV, Ausin I, Law RW, Moissiard G, et al. 2010. Involvement of a Jumonji-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA methylation. *EMBO Rep.* 11:950–55
29. DiPaolo C, Kieft R, Cross M, Sabatini R. 2005. Regulation of trypanosome DNA glycosylation by a SWI2/SNF2-like protein. *Mol. Cell* 17:441–51
30. Du J, Johnson LM, Groth M, Feng S, Hale CJ, et al. 2014. Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Mol. Cell* 55:495–504
31. Dunoyer P, Brosnan CA, Schott G, Wang Y, Jay F, et al. 2010. An endogenous, systemic RNAi pathway in plants. *EMBO J.* 29:1699–712
32. Dürr J, Lolas IB, Sørensen BB, Schubert V, Houben A, et al. 2014. The transcript elongation factor SPT4/SPT5 is involved in auxin-related gene expression in *Arabidopsis*. *Nucleic Acids Res.* 42:4332–47
33. Earley K, Lawrence RJ, Pontes O, Reuther R, Enciso AJ, et al. 2006. Erasure of histone acetylation by *Arabidopsis* HDA6 mediates large-scale gene silencing in nucleolar dominance. *Genes Dev.* 20:1283–93
34. Eick D, Geyer M. 2013. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* 113:8456–90
35. El-Shami M, Pontier D, Lahmy S, Braun L, Picart C, et al. 2007. Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev.* 21:2539–44
36. Enke RA, Dong Z, Bender J. 2011. Small RNAs prevent transcription-coupled loss of histone H3 lysine 9 methylation in *Arabidopsis thaliana*. *PLOS Genet.* 7:e1002350
37. Eun C, Lorković ZJ, Sasaki T, Naumann U, Matzke AJM, Matzke M. 2012. Use of forward genetic screens to identify genes required for RNA-directed DNA methylation in *Arabidopsis thaliana*. *Cold Spring Harb. Symp. Quant. Biol.* 77:195–204
38. Fels A, Hu K, Riesner D. 2001. Transcription of potato spindle tuber viroid by RNA polymerase II starts predominantly at two specific sites. *Nucleic Acids Res.* 29:4589–97
39. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, et al. 2010. Conservation and divergence of methylation patterning in plants and animals. *PNAS* 107:8689–94
40. Finke A, Kuhlmann M, Mette MF. 2012. IDN2 has a role downstream of siRNA formation in RNA-directed DNA methylation. *Epigenetics* 7:950–60
41. Flaus A, Martin DM, Barton GJ, Owen-Hughes T. 2006. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucl. Acids Res.* 34:2887–905
42. Gao Z, Liu HL, Daxinger L, Pontes O, He X, et al. 2010. An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature* 465:106–9
43. Greenberg MV, Ausin I, Chan SW, Cokus SJ, Cuperus JT, et al. 2011. Identification of genes required for *de novo* DNA methylation in *Arabidopsis*. *Epigenetics* 6:344–54
44. Haag JR, Pikaard CS. 2011. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat. Rev. Mol. Cell Biol.* 12:483–92
45. Haag JR, Ream TS, Marasco M, Nicora CD, Norbeck AD, et al. 2012. In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. *Mol. Cell* 48:811–18
46. Hale CJ, Stonaker JL, Gross SM, Hollick JB. 2007. A novel Snf2 protein maintains trans-generational regulatory states established by paramutation in maize. *PLOS Biol.* 5:e275
47. Hartzog GA, Fu J. 2013. The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation. *Biochim. Biophys. Acta* 1829:105–15
48. He XJ, Hsu YF, Zhu S, Liu HL, Pontes O, et al. 2009. A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development. *Genes Dev.* 23:2717–22

49. He XJ, Hsu YF, Zhu S, Wierzbicki AT, Pontes O, et al. 2009. An effector of RNA-directed DNA methylation in *Arabidopsis* is an ARGONAUTE 4- and RNA-binding protein. *Cell* 137:498–508
50. Henderson IR, Deleris A, Wong W, Zhong X, Chin HG, et al. 2010. The *de novo* cytosine methyltransferase DRM2 requires intact UBA domains and a catalytically mutated paralog DRM3 during RNA-directed DNA methylation in *Arabidopsis thaliana*. *PLOS Genet.* 6:e100118
51. Henderson IR, Jacobsen SE. 2008. Tandem repeats upstream of the *Arabidopsis* endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev.* 22:1597–606
52. Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, et al. 2006. Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat. Genet.* 38:721–25
53. Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. 2005. RNA polymerase IV directs silencing of endogenous DNA. *Science* 308:118–20
54. Huang CF, Zhu JK. 2014. RNA splicing factors and RNA-directed DNA methylation. *Biology* 26:243–54
55. Huang L, Jones AM, Searle I, Patel K, Vogler H, et al. 2009. An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. *Nat. Struct. Mol. Biol.* 16:91–93
56. Huettel B, Kanno T, Daxinger L, Aufsatz W, Matzke AJM, Matzke M. 2006. Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*. *EMBO J.* 25:2828–36
57. Inoue N, Hess KD, Moreadith RW, Richardson LL, Handel MA, et al. 1999. New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. *Hum. Mol. Genet.* 8:1201–7
58. Ito H, Kakutani T. 2014. Control of transposable elements in *Arabidopsis thaliana*. *Chromosome Res.* 22:217–23
59. Iyer LM, Abhiman S, Aravind L. 2008. MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. *Biol. Direct* 3:8
60. Johnson LM, Du J, Hale CJ, Bischof S, Feng S, et al. 2014. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* 507:124–28
61. Kang HG, Hyong WC, von Einem S, Manosalva P, Ehlers K, et al. 2012. CRT1 is a nuclear-translocated MORC endonuclease that participates in multiple levels of plant immunity. *Nat. Commun.* 3:1297
62. Kang HG, Kuhl JC, Kachroo P, Klessig DF. 2008. CRT1, an *Arabidopsis* ATPase that interacts with diverse resistance proteins and modulates disease resistance to turnip crinkle virus. *Cell Host Microbe* 3:48–57
63. Kang HG, Oh CS, Sato M, Katagiri F, Glazebrook J, et al. 2010. Endosome-associated CRT1 functions early in resistance gene-mediated defense signaling in *Arabidopsis* and tobacco. *Plant Cell* 22:918–36
64. Kanno T, Bucher E, Daxinger L, Huettel B, Böhmdorfer G, et al. 2008. A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nat. Genet.* 40:670–75
65. Kanno T, Bucher E, Daxinger L, Huettel B, Kreil DP, et al. 2010. RNA-directed DNA methylation and plant development require an IWR1-type transcription factor. *EMBO Rep.* 11:65–71
66. Kanno T, Huettel B, Mette MF, Aufsatz W, Jaligot E, et al. 2005. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* 37:761–65
67. Kanno T, Mette MF, Kreil DP, Aufsatz W, Matzke M, Matzke AJM. 2004. Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Curr. Biol.* 14:801–5
68. Kanno T, Yoshikawa M, Habu Y. 2013. Locus-specific requirements of DDR complexes for gene-body methylation of TAS genes in *Arabidopsis thaliana*. *Plant Mol. Biol. Rep.* 31:1048–52
69. Keller C, Bühler M. 2013. Chromatin-associated ncRNA activities. *Chromosome Res.* 21:627–41
70. Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJ, et al. 2007. Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *Plant J.* 49:38–45
71. Kornienko AE, Guenzl PM, Barlow DP, Pauler FM. 2013. Gene regulation by the act of long non-coding RNA transcription. *BMC Biol.* 11:59
72. Kuhlmann M, Finke A, Mascher M, Mette MF. 2014. DNA methylation maintenance consolidates RNA-directed DNA methylation and transcriptional gene silencing over generations in *Arabidopsis thaliana*. *Plant J.* 80:269–81
73. Kuhlmann M, Mette MF. 2012. Developmentally non-redundant SET domain proteins SUVH2 and SUVH9 are required for transcriptional gene silencing in *Arabidopsis thaliana*. *Plant Mol. Biol.* 79:623–33

74. Lahmy S, Guilleminot J, Cheng CM, Bechtold N, Albert S, et al. 2004. DOMINO1, a member of a small plant-specific gene family, encodes a protein essential for nuclear and nucleolar functions. *Plant J.* 39:809–20
75. Landick R. 2009. Functional divergence in the growing family of RNA polymerases. *Structure* 17:323–25
76. Langen G, von Einem S, Koch A, Imani J, Pai SB, et al. 2014. The Compromised Recognition of Turnip Crinkle Virus1 subfamily of microrchidia ATPases regulates disease resistance in barley to biotrophic and necrotrophic pathogens. *Plant Physiol.* 164:866–78
77. Law JA, Ausin I, Johnson LM, Vashisht AA, Zhu JK, et al. 2010. A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in *Arabidopsis*. *Curr. Biol.* 20:951–56
78. Law JA, Du J, Hale CJ, Feng S, Krajewski K, et al. 2013. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* 498:385–89
79. Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11:204–20
80. Law JA, Vashisht AA, Wohlschlegel JA, Jacobsen SE. 2011. SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. *PLOS Genet.* 7:e1002195
81. Lee TF, Gurazada SG, Zhai J, Li S, Simon SA, et al. 2012. RNA polymerase V-dependent small RNAs in *Arabidopsis* originate from small, intergenic loci including most SINE repeats. *Epigenetics* 7:781–95
82. Li DQ, Nair SS, Ohshiro K, Kumar A, Nair VS, et al. 2012. MORC2 signaling integrates phosphorylation-dependent, ATPase-coupled chromatin remodeling during the DNA damage response. *Cell Rep.* 2:1657–69
83. Lindbo JA, Silva-Rosales L, Proebsting WM, Dougherty WG. 1993. Induction of a highly specific antiviral state in transgenic plants: implications for regulation of gene expression and virus resistance. *Plant Cell* 5:1749–59
84. Liu ZW, Shao CR, Zhang CJ, Zhou JX, Zhang SW, et al. 2014. The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. *PLOS Genet.* 10:e1003948
85. Lorković ZJ. 2012. MORC proteins and epigenetic regulation. *Plant Signal. Behav.* 7:1561–65
86. Lorković ZJ, Naumann U, Matzke AJM, Matzke M. 2012. Involvement of a GHKL ATPase in RNA-directed DNA methylation in *Arabidopsis thaliana*. *Curr. Biol.* 22:933–38
87. Luo J, Hall BD. 2007. A multistep process gave rise to RNA polymerase IV of land plants. *J. Mol. Evol.* 64:101–12
88. Luteijn MJ, Ketting RF. 2013. PIWI-interacting RNAs: from generation to transgenerational epigenetics. *Nat. Rev. Genet.* 14:523–34
89. Ma XF, Gustafson JP. 2005. Genome evolution of allopolyploids: a process of cytological and genetic diploidization. *Cytogenet. Genome Res.* 109:236–49
90. Marí-Ordóñez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O. 2013. Reconstructing *de novo* silencing of an active plant retrotransposon. *Nat. Genet.* 45:1029–39
91. Matzke M, Kanno T, Daxinger L, Huettel B, Matzke AJM. 2009. RNA-mediated chromatin-based silencing in plants. *Curr. Opin. Cell Biol.* 21:367–76
92. Matzke M, Matzke AJM. 2004. Planting the seeds of a new paradigm. *PLOS Biol.* 2:e133
93. Matzke M, Mosher RA. 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* 15:394–408
94. Matzke M, Primig M, Trnovsky J, Matzke AJM. 1989. Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO J.* 8:643–49
95. McClintock B. 1984. The significance of responses of the genome to challenge. *Science* 226:792–801
96. McCue AD, Nuthikattu S, Reeder SH, Slotkin RK. 2012. Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLOS Genet.* 8:e1002474
97. Meister G. 2013. Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.* 14:447–59
98. Mette MF, Aufsatz W, van der Winden J, Matzke M, Matzke AJM. 2000. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* 19:5194–5201
99. Meyer P. 2013. Transgenes and their contributions to epigenetic research. *Int. J. Dev. Biol.* 57:509–15

100. Michael TP. 2014. Plant genome size variation: bloating and purging DNA. *Brief. Funct. Genomics* 13:308–17
101. Moissiard G, Bischof S, Husmann D, Pastor WA, Hale CJ, et al. 2014. Transcriptional gene silencing by *Arabidopsis* microrchidia homologues involves the formation of heteromers. *PNAS* 111:7474–79
102. Moissiard G, Cokus SJ, Cary J, Feng S, Billi AC, et al. 2012. MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* 336:1448–51
103. Napoli C, Lemieux C, Jorgensen R. 1990. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes *in trans*. *Plant Cell* 2:279–89
104. Naumann U, Daxinger L, Kanno T, Eun C, Long Q, et al. 2011. Genetic evidence that DNA methyltransferase DRM2 has a direct catalytic role in RNA-directed DNA methylation in *Arabidopsis thaliana*. *Genetics* 187:977–79
105. Nuthikattu S, McCue AD, Panda K, Fultz D, DeFraia C, et al. 2013. The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21–22 nucleotide small interfering RNAs. *Plant Physiol.* 162:116–31
106. Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, et al. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* 497:579–84
107. Onodera Y, Haag JR, Ream T, Costa-Nunes P, Pontes O, Pikaard CS. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* 120:613–22
108. Panda K, Slotkin RK. 2013. Proposed mechanism for the initiation of transposable element silencing by the RDR6-directed DNA methylation pathway. *Plant Signal. Behav.* 8:e25206
109. Pélissier T, Thalmeir S, Kempe D, Sänger HL, Wassenegger M. 1999. Heavy *de novo* methylation at symmetrical and non-symmetrical sites is a hallmark of RNA-directed DNA methylation. *Nucleic Acids Res.* 27:1625–34
110. Pélissier T, Wassenegger M. 2000. A DNA target of 30 bp is sufficient for RNA-directed DNA methylation. *RNA* 6:55–65
111. Pikaard CS, Haag JR, Pontes OM, Blevins T, Cocklin R. 2012. A transcription fork model for Pol IV and Pol V-dependent RNA-directed DNA methylation. *Cold Spring Harb. Symp. Quant. Biol.* 77:205–12
112. Pikaard CS, Tucker S. 2009. RNA-silencing enzymes Pol IV and Pol V in maize: more than one flavor? *PLOS Genet.* 5:e1000736
113. Pontier D, Picart C, Roudier F, Garcia D, Lahmy S, et al. 2012. NERD, a plant-specific GW protein, defines an additional RNAi-dependent chromatin-based pathway in *Arabidopsis*. *Mol. Cell* 48:121–32
114. Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, et al. 2005. Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev.* 19:2030–40
115. Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, et al. 2009. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol. Cell* 33:192–203
116. Rowley MJ, Avrutsky MI, Sifuentes CJ, Pereira L, Wierzbicki AT. 2011. Independent chromatin binding of ARGONAUTE4 and SPT5L/KTF1 mediates transcriptional gene silencing. *PLOS Genet.* 7:e1002120
117. Sasaki T, Lee TF, Liao WW, Naumann U, Liao JL, et al. 2014. Distinct and concurrent pathways of Pol II- and Pol IV-dependent siRNA biogenesis at a repetitive *trans*-silencer locus in *Arabidopsis thaliana*. *Plant J.* 79:127–38
118. Sasaki T, Lorković ZJ, Liang SC, Matzke AJM, Matzke M. 2014. The ability to form homodimers is essential for RDM1 to function in RNA-directed DNA methylation. *PLOS ONE* 9:e88190
119. Saze H, Tsugane K, Kanno T, Nishimura T. 2012. DNA methylation in plants: relationship to small RNAs and histone modifications, and functions in transposon inactivation. *Plant Cell Physiol.* 53:766–84
120. Schmitz KM, Mayer C, Postepska A, Grummt I. 2010. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* 24:2264–69
121. Searle IR, Pontes O, Melnyk CW, Smith LM, Baulcombe DC. 2010. JMJ14, a JmjC domain protein, is required for RNA silencing and cell-to-cell movement of an RNA silencing signal in *Arabidopsis*. *Genes Dev.* 24:986–91

122. Sijen T, Vijn I, Rebocho A, van Blokland R, Roelofs D, et al. 2001. Transcriptional and posttranscriptional gene silencing are mechanistically related. *Curr. Biol.* 11:436–40
123. Smith LM, Pontes O, Searle I, Yelina N, Yousafzai FK, et al. 2007. An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in *Arabidopsis*. *Plant Cell* 19:1507–21
124. Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, et al. 2009. Polyploidy and angiosperm diversification. *Am. J. Bot.* 96:336–48
125. Sridhar VV, Kapoor A, Zhang K, Zhu J, Zhou T, et al. 2007. Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. *Nature* 447:735–38
126. Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE. 2013. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* 152:352–64
127. Swarts DC, Jore MM, Westra ER, Zhu Y, Janssen JH, et al. 2014. DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* 507:258–61
128. Tan EH, Blevins T, Ream TS, Pikaard CS. 2012. Functional consequences of subunit diversity in RNA polymerases II and V. *Cell Rep.* 1:208–14
129. Tayalé A, Parisod C. 2013. Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenet. Genome Res.* 140:79–96
130. Tran RK, Zilberman D, de Bustos C, Ditt RF, Henikoff JG, et al. 2005. Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in *Arabidopsis*. *Genome Biol.* 6:R90
131. Tucker SL, Reece J, Ream TS, Pikaard CS. 2010. Evolutionary history of plant multisubunit RNA polymerases IV and V: subunit origins via genome-wide and segmental gene duplications, retrotransposition, and lineage-specific subfunctionalization. *Cold Spring Harb. Symp. Quant. Biol.* 75:285–97
132. van der Krol AR, Mur LA, Beld M, Mol JNM, Stuitje AR. 1990. Flavonoid genes in petunia: addition of a limited number of gene copies may lead to suppression of gene expression. *Plant Cell* 2:291–99
133. Vaucheret H. 2005. RNA polymerase IV and transcriptional silencing. *Nat. Genet.* 37:659–60
134. Wassenegger M. 2000. RNA-directed DNA methylation. *Plant Mol. Biol.* 43:203–20
135. Wassenegger M, Heimes S, Riedel L, Sänger HL. 1994. RNA-directed de novo methylation of genomic sequences in plants. *Cell* 76:567–76
136. Wei L, Gu L, Song X, Cui X, Lu Z, et al. 2014. Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. *PNAS* 111:3877–82
137. Wierzbicki AT. 2012. The role of long non-coding RNAs in transcriptional gene silencing. *Curr. Opin. Plant Biol.* 15:517–22
138. Wierzbicki AT, Cocklin R, Mayampurath A, Lister R, Rowley MJ, et al. 2012. Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the *Arabidopsis* epigenome. *Genes Dev.* 26:1825–36
139. Wierzbicki AT, Haag JR, Pikaard CS. 2008. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135:635–48
140. Wierzbicki AT, Ream TS, Haag JR, Pikaard CS. 2009. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat. Genet.* 4:630–34
141. Wild T, Cramer P. 2012. Biogenesis of multisubunit RNA polymerases. *Trends Biochem. Sci.* 37:99–105
142. Wu L, Mao L, Qi Y. 2012. Roles of DICER-LIKE and ARGONAUTE proteins in *TAS*-derived small interfering RNA-triggered DNA methylation. *Plant Physiol.* 160:990–99
143. Xie M, Ren G, Costa-Nunes P, Pontes O, Yu B. 2012. A subgroup of SGS3-like proteins act redundantly in RNA-directed DNA methylation. *Nucleic Acids Res.* 40:4422–31
144. Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, et al. 2004. Genetic and functional diversification of small RNA pathways in plants. *PLOS Biol.* 2:e104
145. You W, Lorković ZJ, Matzke AJM, Matzke M. 2013. Interplay among RNA polymerases II, IV and V in RNA-directed DNA methylation at a low copy transgene locus in *Arabidopsis thaliana*. *Plant Mol. Biol.* 82:85–96
146. You W, Tyczewska A, Spencer M, Daxinger L, Schmid MW, et al. 2012. Atypical DNA methylation of genes encoding cysteine-rich peptides in *Arabidopsis thaliana*. *BMC Plant Biol.* 12:51

147. Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, et al. 2013. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 28:193–205
148. Zemach A, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328:916–19
149. Zhang CJ, Ning YQ, Zhang SW, Chen Q, Shao CR, et al. 2012. IDN2 and its paralogs form a complex required for RNA-directed DNA methylation. *PLOS Genet.* 8:e1002693
150. Zhang H, Ma ZY, Zeng L, Tanaka K, Zhang CJ, et al. 2013. DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. *PNAS* 110:8290–95
151. Zhang H, Zhu JK. 2011. RNA-directed DNA methylation. *Curr. Opin. Plant Biol.* 14:142–47
152. Zhang H, Zhu JK. 2012. Active DNA demethylation in plants and animals. *Cold Spring Harb. Symp. Quant. Biol.* 77:161–73
153. Zhao Y, Mo B, Chen X. 2012. Mechanisms that impact microRNA stability in plants. *RNA Biol.* 9:1218–23
154. Zheng B, Wang Z, Li S, Yu B, Liu JY, Chen X. 2009. Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in *Arabidopsis*. *Genes Dev.* 23:2850–60
155. Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, et al. 2014. Molecular mechanism of action of plant DRM *de novo* DNA methyltransferases. *Cell* 157:1050–60
156. Zhong X, Hale CJ, Law JA, Johnson LM, Feng S, et al. 2012. DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat. Struct. Mol. Biol.* 19:870–75
157. Zhu Y, Rowley MJ, Böhmdorfer G, Wierzbicki AT. 2013. A SWI/SNF chromatin-remodeling complex acts in noncoding RNA-mediated transcriptional silencing. *Mol. Cell* 49:298–309