GARDENING THE GENOME: DNA METHYLATION IN *ARABIDOPSIS THALIANA*

Simon W.-L. Chan, Ian R. Henderson* and Steven E. Jacobsen*

Abstract | DNA methylation has two essential roles in plants and animals — defending the genome against transposons and regulating gene expression. Recent experiments in *Arabidopsis thaliana* have begun to address crucial questions about how DNA methylation is established and maintained. One cardinal insight has been the discovery that DNA methylation can be guided by small RNAs produced through RNA-interference pathways. Plants and mammals use a similar suite of DNA methyltransferases to propagate DNA methylation, but plants have also developed a glycosylase-based mechanism for removing DNA methylation, and there are hints that similar processes function in other organisms.

Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, California 90095, USA.
Correspondence to S.E.J.
e-mail: jacobsen@ucla.edu
**These authors contributed equally to this work*
doi:10.1038/nrg1601

Cytosine methylation is a eukaryotic gene-silencing mechanism that protects the genome by inactivating selfish DNA elements, including transposons[1,2]. The clearest demonstration of this has come from the model plant *Arabidopsis thaliana*, in which it was shown that transposons are mobilized when DNA methylation is reduced[3,4]. DNA methylation and/or conserved DNA methyltransferase enzymes are found in most eukaryotes, despite their loss from the model organisms such as *Saccharomyces cerevisiae* and *Caenorhabditis elegans*[5–12]. Eukaryotic DNA methyltransferases have structural homologues in the prokaryotic restriction–modification systems, which selectively degrade non-host DNA[13]. Therefore, it seems likely that the last universal common ancestor of all organisms used DNA methylation for genome defence.

Both plants and animals have co-opted DNA methylation to regulate selected endogenous genes[14]. Plants use DNA methylation for genomic imprinting and to modulate the expression of repeated gene families[15–18]. In mammals, DNA methylation controls genomic imprinting, X-chromosome inactivation and the silencing of tumour-suppressor genes[14]. Mammalian DNA methylation is mostly restricted to symmetrical CG sequences, although other sequences are methylated in mouse embryonic stem cells[14,19]. By contrast, plant DNA methylation occurs at CG, CNG (where N is any nucleotide) and CHH (an asymmetric site, where H is A, C or T) sequences, each of which has different genetic requirements for the maintenance of its methylation. *Arabidopsis thaliana* contains the full complement of the DNA-methyltransferase types that are found in mammals (FIG. 1), and recent experiments have elucidated key mechanisms of DNA methylation.

In this review, we first discuss which sequences are targeted for DNA methylation in *A. thaliana*. We then describe how DNA methylation is initiated and propagated. Finally, we review the surprising finding that DNA glycosylases can remove DNA methylation in plants.

Functions of DNA methylation in *A. thaliana*

DNA methylation in *A. thaliana* has two roles: it protects the genome from selfish DNA elements and regulates gene expression. These functions are reflected in the endogenous genome sequences of *A. thaliana* that are methylated (see TABLE 1 for a partial list of these sequences). Consistent with the primary role of DNA methylation in genome defence, transposons and other dispersed repeat sequences comprise most of the methylated DNA. These sequences cluster around centromeres but are also found in euchromatin[3,4,20–22]. Sequences

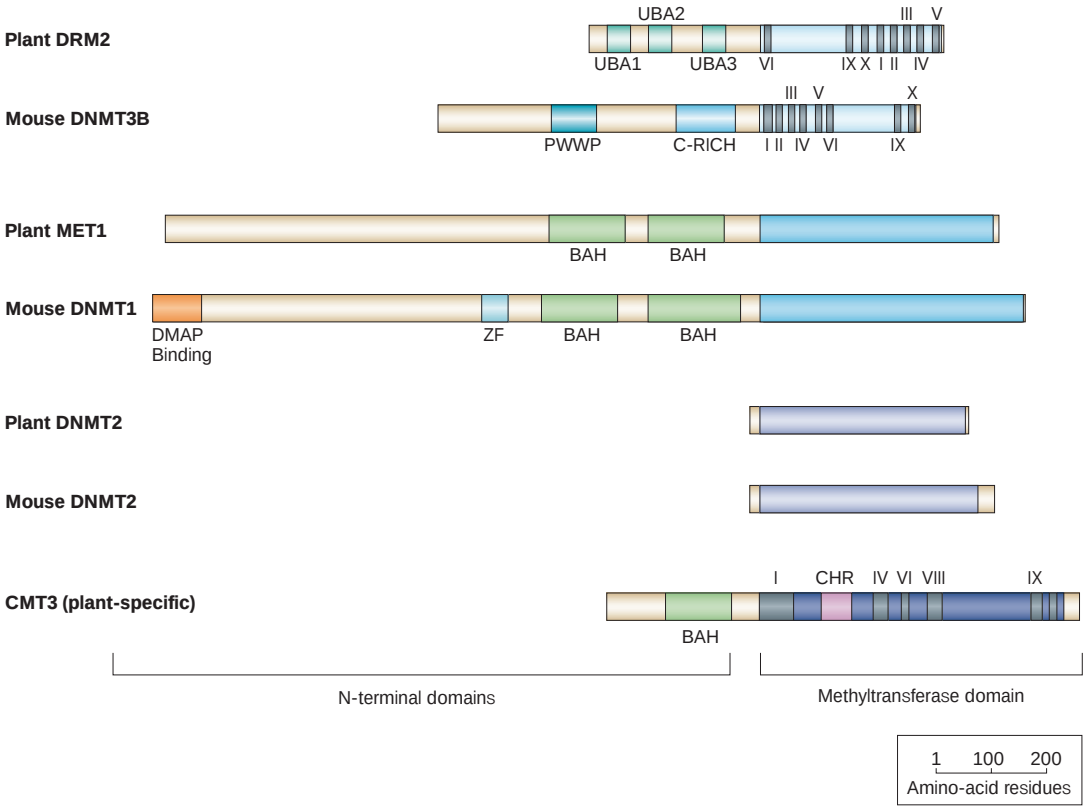

Figure 1 | **Four classes of DNA methyltransferase in *Arabidopsis thaliana*.** The *Arabidopsis thaliana* genome encodes four classes of DNA cytosine methyltransferases. The first class, represented by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), is orthologous to the mammalian DNA (cytosine-5-)-methyltransferase 3 (DNMT3) (such as the mouse DNMT3B shown in the diagram) and both function to control *de novo* methylation. DRM2 differs from DNMT3 proteins in that it has a unique N terminus and that its catalytic domains are rearranged in the linear sequence. METHYLTRANSFERASE 1 (MET1) is orthologous to the mammalian DNMT1 enzyme and both function to maintain CG methylation. Both MET1 and DNMT1 possess large N termini that contain bromo adjacent homology (BAH) domains. The DNMT2 class of methyltransferase is conserved in many eukaryote genomes but its function is unknown. Plant genomes are distinguished from mammalian genomes by the CHROMOMETHYLASE (CMT) class of methyltransferases, which function to control the maintenance of non-CG methylation. CHR, chromodomain; C-rich, cysteine rich; DMAP binding, DNMT1-associated protein binding; PWWP, Pro-Trp-Trp-Pro domain; UBA, ubiquitin associated domain; ZF, zinc finger.

containing local tandem or inverted repeats are also frequently methylated, and methylated repeats that are dispersed in the genome often contain local repeats. For example, DNA transposons have inverted repeats at their termini, retrotransposons are flanked by tandemly repeated sequences, and some tandem repeats, such as *MEA-ISR* (an intergenic subtelomeric repeat sequence downstream of the *MEDEA* (*MEA*) gene), are found at several loci in the genome[23,24]. Locally repeated sequences can be targeted for DNA methylation when they are transformed into *A. thaliana*, and can also result in the methylation of homologous sequences in *trans*[25–28]. These observations raise the as yet unanswered question of how plants distinguish repeated sequences from unique DNA sequences, which are seldom subject to *de novo* methylation.

In some relatively rare cases, DNA methylation can control endogenous gene expression as well as the silencing of transposons. The ribosomal RNA (rRNA) genes are found in tandem arrays containing hundreds of copies that are stochastically silenced. The silenced subset of rRNA genes is methylated, and treatment with 5-azacytidine, a drug that inhibits DNA methyltransferases, can reverse this silencing[18]. The *FWA* transcription factor is an example of a developmental gene that is regulated by DNA methylation. *FWA* is expressed only in the extra-embryonic ENDOSPERM and is silenced in all other plant tissues by the methylation of tandem repeats at the start of its transcribed region. Expression in the endosperm occurs when DNA methylation in this tissue is removed (see section on demethylation by DNA glycosylases)[17,28]. In other cases, DNA methylation controls the overall level of expression from a family of repeated genes. For example, the pathogen-resistance gene *BALL* (*BAL*) resides in a complex gene cluster and is silenced by DNA methylation under laboratory growth conditions[15]. The expression of tryptophan biosynthesis genes in the *PHOSPHORIBOSYLANTHRANILATE ISOMERASE* (*PAI*) family can be modulated by DNA methylation. In some *A. thaliana* ACCESSIONS, a *PAI1*-*PAI4* inverted repeat triggers the silencing of the homologous *PAI2* gene[16].

ENDOSPERM
The triploid seed tissue, which often provides nutrition to the developing embryo. It is formed by the fertilization of the embryo sac central cell (diploid) by a sperm nucleus (haploid) from the pollen.

ACCESSION
A homozygous line of *Arabidopsis thaliana* collected from a natural population at a specific location.

Table 1 | Examples of endogenous DNA methylation in *Arabidopsis thaliana*

Site	Type of repeat	Biology and function	Reference
Retrotransposons	Dispersed and tandem	Silencing and genome defence	22
DNA transposons	Dispersed and inverted	Silencing and genome defence	22
CEN	Tandem	Centromere function?	119
rDNA	Tandem	rRNA transcriptional regulation	18
Specific methylated genes			
BAL cluster	Many	Gene regulation	15
PAI	Dispersed and inverted	Gene regulation	16
FWA	Tandem	Developmental gene expression and imprinting	17,28
MEA-ISR	Dispersed and tandem	Unknown	24
CG clusters	None	Unknown	34
PHAVOLUTA/PHABULOSA	None	Development	120
DRM2	None	Unknown	121

BAL, *BALL*; CEN, centromere; *DRM2*, *DOMAINS REARRANGED METHYLTRANSFERASE 2*; *FWA*, encodes a homeodomain-containing transcription factor; *MEA-ISR*, intergenic subtelomeric repeat sequence downstream of the *MEDEA* gene; *PAI*, *PHOSPHORIBOSYLANTHRANILATE ISOMERASE* gene family.

In plants, the DNA methylation of promoter regions usually inhibits transcription, but methylation in coding regions does not generally affect gene expression[29–31]. In genes that are exceptions to this rule, such as *SUPERMAN* (*SUP*) and *AGAMOUS* (*AG*), DNA methylation in the transcribed portion of the gene probably causes transcriptional shut-down because there are important controlling elements in these regions[32,33]. *FWA* is an unusual case, because the methylated direct repeats are in the 5′ end of the transcribed region, yet DNA methylation inhibits gene expression[28]. In this case, the proximity of DNA methylation to the promoter might allow it to inhibit transcription. A recent microarray study has uncovered new sites of CG methylation that reside predominantly in the 3′ end of genes, although the biological function and genesis of methylation in these regions is unknown[34]. The situation in *A. thaliana* contrasts with that in *Neurospora crassa*, where the DNA methylation of coding sequences can attenuate transcriptional elongation[35]. However, in mammalian cells, exons and introns are routinely methylated, but only the methylation status of the promoter seems to affect gene expression[14].

Establishing DNA methylation

One of the key questions regarding the function of DNA methylation is how specific sequences are targeted during the establishment of gene silencing. In many experimental plant systems, DNA methylation can be guided by RNA — in the first described example, a transgenic RNA VIROID caused the DNA methylation of a sequence homologous to that viroid[27,31,36–38]. Small RNAs involved in these processes are generated by the well-studied RNAi pathway, in which dsRNA is cleaved into 21–25 nucleotide small interfering RNAs (siRNAs) by the ribonuclease Dicer[39]. The fact that single-stranded viroids and RNA viruses replicate by means of a dsRNA intermediate might allow them to be targeted by RNAi[37,38]. Virus-derived siRNAs are then thought to guide DNA methylation to homologous sequences. Examples of virus-induced gene silencing that feature *trans*-DNA methylation exemplify the biological role of RNAi and DNA methylation as genome-defence mechanisms. Mutations in the *RNA-DEPENDENT RNA POLYMERASE 6* (*RDR6*), *SUPPRESSOR OF GENE SILENCING 3/SILENCING DEFECTIVE 2* (*SGS3/SDE2*) and *SILENCING DEFECTIVE 3* (*SDE3*) genes impair virus-induced gene silencing and also cause defects in the RNAi-like silencing that is induced by highly transcribed sense transgenes (called sense post-transcriptional gene silencing or S-PTGS)[40–42]. S-PTGS causes the DNA methylation of homologous loci, providing another link between siRNA-mediated gene silencing and DNA methylation[30]. Transcriptional silencing by DNA methylation is one of several downstream mechanisms that siRNAs can use to downregulate gene expression[39]. Chromatin targeting by siRNAs is conserved in diverse eukaryotes, as shown by siRNA-directed transcriptional silencing in the fission yeast *Schizosaccharomyces pombe* and human cells, and by programmed genome elimination in the ciliated protozoan *Tetrahymena thermophila*[43–49].

The phenomenon of RNA-directed DNA methylation raised interest in those plant DNA methyltransferases that might be involved in establishing DNA methylation. Proteins of the DOMAINS REARRANGED METHYLTRANSFERASE (DRM) family are orthologous to mammalian *de novo* methyltransferases of the DNA (cytosine-5-)-methyltransferase 3 (DNMT3) family, although the catalytic domains of these two families are arranged differently in the linear amino-acid sequence[50] (FIG. 1). *Arabidopsis thaliana drm1 drm2* double mutants lack every type of *de novo* DNA methylation tested, including methylation of transformed tandem repeats, transcribed inverted repeats and DNA methylation induced by the transgenic *SUPERMAN* inverted repeat called *clk–st* [25,51,52]. *DRM1* is expressed at a much lower level than *DRM2*, and the *drm2* mutant recapitulates all the tested phenotypes of *drm1 drm2* mutants[52]. In the case of the *FWA* tandem repeats, *de novo* DNA methylation of a transformed copy requires a complete siRNA-generating pathway, including *RNA-DEPENDENT*

VIROID
An infectious agent of plants that consists of ssRNA but that lacks the protein component that is typical of viruses.

RNA POLYMERASE 2 (*RDR2*), *DICER-LIKE 3* (*DCL3*), *RNA POLYMERASE D1* (*RPD1*) and *ARGONAUTE 4* (*AGO4*)[25,53,54]. This observation was unexpected because there is no obvious mechanism for generating an initial dsRNA (that is, a Dicer substrate) from a direct repeat — it is not clear how the RNAi proteins are recruited specifically to direct repeats as opposed to unique sequences. This is in contrast to inverted repeat sequences, where unidirectional transcription and foldback creates dsRNA that can be cleaved by Dicer. The fact that mutations in RNA-silencing genes completely phenocopy the *de novo* methyltransferase *drm1 drm2* double mutant indicates that the guidance of DNA methylation by siRNA might extend beyond inverted repeats or viruses.

An absolute test of whether all *de novo* DNA methylation in *A. thaliana* is guided by siRNA has been hampered by the fact that plants encode multiple paralogues of each RNAi protein (there are 6 RNA-dependent RNA polymerases, 4 Dicers and 10 Argonautes)[54–56]. A further complication is that *dicer-like 1* (*dcl1*) and *argonaute 1* (*ago1*) mutants have severe developmental defects because the encoded proteins are important for microRNA-mediated regulation of embryonic development[55,57]. *De novo* DNA methylation induced by the transcription of inverted repeats does not require AGO4, but is completely dependent on the DRM methyltransferases[26,51]. One explanation for this result is that redundant siRNA-generating pathways can guide DRM methyltransferases to inverted repeats and homologous loci. The existence of different size classes of siRNAs that have distinct biological functions supports this interpretation. The DCL3 pathway is mostly associated with 24-nucleotide siRNAs that direct DNA methylation, whereas 21-nucleotide siRNAs made by an unknown Dicer are involved in mRNA degradation but can also target DNA methylation caused by viruses and S-PTGS[53,54]. Although the role of DRM proteins is clear cut, other DNA methyltransferases might function in *de novo* DNA methylation, as indicated by the fact that *chromomethylase 3* (*cmt3*) mutants cannot initiate DNA methylation that is induced by the inverted repeat *PAI1-PAI4* (see below for further discussion)[58]. The effects of *drm* mutants on this system have not been tested, and it might be that initial DNA methylation by DRM1 and DRM2 requires subsequent action by CMT3 for its full establishment.

A key mechanistic question in RNA-directed DNA methylation is how siRNAs target the DRM enzymes to specific sequences. In one model, genomic DNA unwinds, allowing base pairing with the siRNA. In a second model, transcription of a gene produces a nascent RNA that base pairs with an siRNA, tethering a complex — which includes chromatin-modifying enzymes and possibly the DRM methyltransferases — to a given locus. Indeed, there is some evidence for this 'surveillance transcription' model from experiments in *S. pombe*[59,60]. An RNA transcription-based model is supported by the intriguing finding that the plant-specific DNA-dependent RNA polymerase IV (pol IV) is important for siRNA-directed DNA methylation[61,62]. The large subunit of pol IV is encoded by *SDE4*, now renamed *RPD1*, which was previously shown to function in a genetic pathway with the RNAi genes *RDR2*, *DCL3* and *AGO4* to generate siRNAs that direct the DRM methyltransferases[25,53].

Maintenance of CG DNA methylation

Cytosine methylation in mammalian genomes occurs predominantly in the context of CG sequences and is maintained by the DNMT1 methyltransferase[14]. CG methylation is also the most common modification within plant genomes and occurs most frequently in heterochromatic sequences, particularly near the centromeres[63,64]. Antisense suppression of the *A. thaliana* DNMT1-like gene *METHYLTRANSFERASE 1* (*MET1*) causes a global reduction in cytosine methylation, particularly at CG sites[65,66]. Mutations in *MET1* have also been isolated in screens for the global loss of DNA methylation and for the release of transcriptional gene silencing[67,68]. Weak *met1* mutations initially have no morphological phenotype, but following inbreeding, *met1* lines accumulate numerous EPIMUTATIONS[28,68,69]. Some epimutations are due to the loss or gain of endogenous gene silencing[28,69]. Developmental mutations also arise in backgrounds that have reduced CG methylation through the mobilization and insertion of transposable elements[3,4,70]. The weak *met1* alleles behave as recessive mutations and therefore must be made homozygous before the effects on DNA methylation can be seen[68]. By contrast, null *met1* heterozygotes have dominant phenotypes associated with the loss of DNA methylation and gene silencing[67]. This phenotype is due to a requirement for MET1 during the GAMETOPHYTIC STAGE of the *A. thaliana* life cycle[67]. Plant gametogenesis involves post-meiotic DNA replication of haploid nuclei before fertilization. In the absence of *MET1*, loci that are normally silenced lose their methylation at the gametophyte stage and become transcriptionally activated, leading to the occurrence of epimutations in the progeny of *met1* heterozygotes[67]. Therefore, the maintenance of CG methylation, and potentially other aspects of epigenetic gene regulation, are key processes in the gametophytic stage of the plant life cycle.

Additional loci that are required for CG methylation and gene silencing were characterized in the same genetic screens in which *met1* mutations were isolated. The *DECREASE IN DNA METHYLATION 1* (*DDM1*) gene encodes a protein related to SWI2/SNF2-like chromatin remodelling enzymes[71]. Indeed, *in vitro* studies have shown that recombinant DDM1 has ATPase activity and is able to remodel nucleosomes[72]. Mutations in *DDM1* cause a loss of both CG DNA methylation and methylation at lysine residue K9 in histone H3 (H3-K9) at heterochromatic loci[22,73–75]. The mechanism of DDM1 action is unknown, but the related mouse protein lymphoid specific helicase (LSH) also controls DNA methylation patterning, indicating a conserved function[76]. Although H3-K9 methylation is lost from silenced loci in *ddm1* mutants, their global levels of this modification are normal[73]. *MET1* and *DDM1* are both required for full levels of CG methylation, but their mutant phenotypes are different. Null *ddm1* mutants do not show the gametophytic effects on gene silencing

EPIMUTATION
A heritable change in gene expression but not gene sequence. This usually takes place by an abnormal increase or decrease in the methylation status of a gene. This can then be heritable for many generations.

GAMETOPHYTIC STAGE
The haploid phase of the plant life-cycle, in which a post-meiotic cell undergoes 2–3 mitoses. In flowering plants, the embryo sac comprises the female structure and the male form is the pollen grain.

that are observed in null *met1* mutants[67]. Furthermore, the *ddm1-2 met1-1* double mutation has additive consequences on development[68].

The *HISTONE DEACETYLASE 6* (*HDA6*) gene is also required for the maintenance of CG methylation at some loci[77–79]. HDA6 is related to the Rpd3 class of histone deacetylases of *S. cerevisiae*, which function in transcriptional silencing[80]. Mutations in *HDA6* were found as strong suppressors of the maintenance of CG methylation directed by RNA from a *NOPALINE SYNTHASE* promoter inverted repeat[77]. In another screen that also isolated an *hda6* mutation, the transgenic locus that was silenced contained an inverted repeat[78]. Compared with the *MET1* and *DDM1* genes, *HDA6* has more specific effects on endogenous silenced loci, as loss of *HDA6* has no effect on the methylation of the centromeric repeats, although it does cause reactivation of the *Athila-TSI* retroelement[77,79]. The rDNA in *hda6* mutants is cytologically decondensed and this is associated with hyperactylated histones, reduced DNA methylation and increased H3-K4 methylation[79]. In addition to HDA6, a second histone deacetylase, HDT1 is required for maintenance of rRNA gene promoter methylation[18].

CG methylation and H3-K9 methylation frequently occur together at transcriptionally silenced, heterochromatic loci[22]. CG methylation controls H3-K9 methylation in *A. thaliana*[64,74], as most H3-K9 methylation is lost at centromeric and repeated heterochromatic loci in null *met1-3* mutants[81]. Conversely, histone methylation can influence cytosine methylation in *A. thaliana*, indicating that a feedback loop between these modifications reinforces the silenced state[58,82]. In addition to controlling CG methylation, *MET1* is also required to maintain non-CG methylation[65,66,82–84]. However, all non-CG methylation is eliminated in *drm1 drm2 cmt3* triple mutants in the presence of functional *MET1*[24]. This indicates that the effects of *MET1* on non-CG methylation are indirect and that pre-existing CG methylation, or a chromatin mark associated with it, might be able to attract non-CG methylation. Methyl binding domain (MBD) proteins are good candidates for the recognition of the CG methylation mark and the promotion of other types of histone or DNA modifications[85–88]. These proteins bind methylated DNA and might recruit other factors that mediate epigenetic modifications. For example, mouse MBD-protein complexes can lead to further repressive chromatin modifications such as histone methylation and deacetylation[89–91].

Small RNA signalling pathways are able to establish CG methylation equally as well as non-CG methylation[25,26]. However, RNA does not appear to be required for the maintenance of CG methylation. Removal of an RNA-trigger sequence that directs CG methylation does not prevent the maintenance of methylation, unless MET1 function is compromised[92,93]. Furthermore, mutations in the RNAi machinery that affect DNA methylation do not prevent the maintenance of CG methylation[25,26]. A *met1* mutation does cause the specific loss of CG methylation in the presence of an siRNA signal, and one interpretation of this result is that MET1 is an RNA-guided *de novo* methyltransferase[94]. In an alternative model, this observation can be explained by the sequence specificity of DRM2, which prefers non-CG sites *in vitro*, although it does methylate the CG dinucleotide less efficiently[95]. Therefore, in a *met1* mutant in which the maintenance of CG methylation is defective, the RNA-guided activity of DRM2 might propagate CG methylation at a lower level. This model is supported by the fact that MET1 function cannot establish methylation in any context, as a *drm1 drm2* mutant completely blocks *de novo* methylation driven by diverse triggers[25,51,52].

Maintenance of CNG methylation

Plant genomes contain extensive methylation at CNG sites and genetic screens have demonstrated that methylation in this context is controlled by a pathway that is distinct from the one that regulates CG methylation[96]. The *SUP* floral developmental gene becomes densely hypermethylated and silenced in several backgrounds that have genome-wide hypomethylation, although DNA methylation probably does not control its expression in normal development[69,97]. Silenced *SUP* alleles (termed *clark kent* or *clk*) are recessive and cause the formation of additional stamens and defective female floral organs. Mutant *met1* plants frequently acquire *clk* epimutations, indicating that MET1 activity is not required for *SUP* hypermethylation[69,97]. Mutations in the gene encoding the plant-specific methyltransferase CMT3 were isolated in a suppressor screen that used a stable, non-reverting line containing a transgenic *SUP* inverted repeat (called *clk-st*)[82]. CMT3 also has an N-terminal CHROMODOMAIN[98]. In addition to controlling CNG methylation at *SUP*, *CMT3* methylates several additional endogenous sequences such as the *Athila* and *Ta3* transposons[82]. Mutations in *CMT3* were also isolated as suppressors of the *PAI*-silencing system[99]. The *PAI2* gene encodes an intermediate enzyme in the tryptophan biosynthetic pathway. Loss of *PAI* expression results in pleiotropic phenotypes, including the accumulation of fluorescent tryptophan intermediates, reduced size and fertility. In the WS accession of *A. thaliana*, an unlinked *PAI1-PAI4* inverted repeat causes methylation of itself and of *PAI2* (REFS 16,100). A screen for suppressors of this silencing identified mutations in *CMT3* and showed that the main effect was the loss of non-CG methylation and of gene silencing[99]. The CMT class of methyltransferases has also been characterized through reverse genetic approaches in maize[101]. Together, these studies indicate that CMT genes encode the major CNG methyltransferases in plants.

A second class of suppressor mutations isolated from both the *SUP* and *PAI* screens was in the *KRYPTONITE* (*KYP*) gene (also known as *SU(VAR)3-9 HOMOLOG 4* (*SUVH4*)), which encodes a protein related to the H3-K9 methyltransferases SU(VAR)3-9 (mammal) and CLR4 (yeast)[58,102]. KYP possesses a catalytic SET methyltransferase domain and has H3-K9 methyltransferase activity *in vitro*, and is required *in vivo* for this mark[58,102–104]. KYP differs from SU(VAR)3-9 in that it lacks an N-terminal chromodomain and instead has a YDG domain that has been implicated in histone binding[105]. The *kyp* mutation also leads to a loss of CNG and

CHROMODOMAIN
A protein domain shared by several regulators of chromatin structure. Different classes of chromodomains have been implicated in binding histones, RNA and DNA.

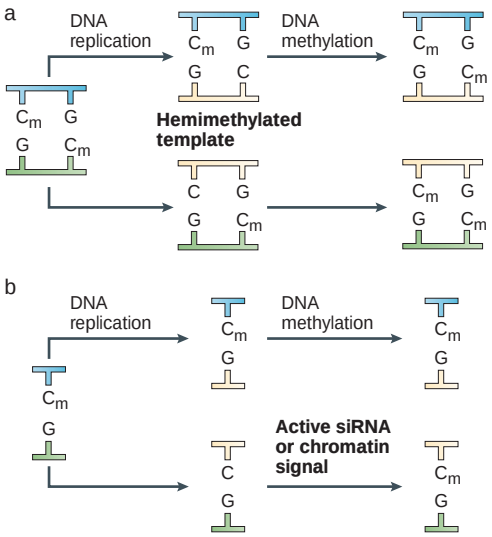

Figure 2 | **Mechanisms for maintenance of CG and asymmetric methylation. a** | Following DNA replication of methylated cytosine residues in a symmetric context (CG), both daughter strands are hemimethylated. This provides a template for MET1/DNMT1 class of methyltransferases to maintain CG methylation. **b** | This is not the case for methylated cytosines in non-symmetric-sequence contexts (CNG and CHH), so either a chromatin or RNA signal is necessary to maintain methylation at these sites.

asymmetric methylation at the *SUP* locus (although less than *cmt3*) and has similar effects on endogenous loci such as the *Ta3* transposon[58,102]; therefore, in *A. thaliana*, a histone methyltransferase controls CNG methylation. DNA methylation is also genetically downstream of H3-K9 methylation in *N. crassa*[106]. Interestingly, the *kyp* mutation reduces methylation at both the *clk-st SUP* inverted repeat and the endogenous gene, whereas it only reduces methylation at the *PAI2* singlet gene and not the *PAI1-PAI4* inverted repeat[58,102]. This indicates that the control of methylation at the *PAI* trigger might differ from that at other sequences.

The *kyp* and *cmt3* mutations show epistasis, indicating that these genes function in a linear pathway to control CNG methylation[58]. This raises the question as to how KYP targets CMT3-cytosine-methyltransferase activity. The *in vitro* binding of CMT3 to LIKE HETEROCHROMATIN PROTEIN 1 (LPH1), which itself binds methyl-H3-K9, provides an attractive model for this targeting[82]. However, the failure of *lph1* mutants to suppress the *clk-st* and *PAI* silencing systems weakens this possibility[58,103]. Further studies have demonstrated that the chromodomain of CMT3 binds histone H3 tails, but only if they are simultaneously methylated at positions K9 and K27 (REF. 103). *In vivo* H3-K27 methylation cytologically co-localizes with CMT3 target regions such as centromeric heterochromatin[103]. In animals, H3-K27 methylation is associated with several heterochromatic loci, and the addition of this mark is catalysed by the POLYCOMB GROUP (PcG) proteins[107,108]. However, in plants, both *kyp* and PcG gene mutants fail to abolish the heterochromatic staining of methyl-H3-K27 (REF. 103). Together, these data indicate that a maintenance loop might exist whereby histone methylation mediated by KYP and an unknown K27 methylase recruits CMT3 and targets CNG methylation to heterochromatic loci. As discussed above, non-CG methylation might also be downstream of MET1 and CG methylation at many loci and this might be achieved through histone methylation[81].

Although CMT3 maintains the bulk of CNG methylation at hypermethylated *SUP* and *PAI*, the DRM methyltransferases can also propagate CNG and there are locus-specific differences in the requirement for each class of methyltransferase[24–26,82,99]. In the centromere and at the pericentromeric transposon *Ta3* , CMT3 controls CNG methylation[24]. At sites of euchromatic DNA methylation that are characterized by inverted repeats (for example, *clk-st* and *nopaline-synthase* inverted repeat) or dispersed repeats (for example *AtSN1*), CNG and CHH (see below for further discussion) are controlled redundantly by CMT3, DRM1 and DRM2 (REFS 26,82,99). Non-CG methylation at the tandem repeats *FWA* and *MEA-ISR* is almost exclusively dependent on DRM1 and DRM2 (REF. 25). As a further sign of redundant function, *cmt3* and *drm1 drm2* mutant plants seem to be morphologically similar to the wild type, but *cmt3 drm1 drm2* triple mutants have a pleiotropic developmental phenotype[24]. This observation indicates that non-CG methylation controls developmentally important genes, but unlike the variable epigenetic phenotypes seen in *met1* and *ddm1* mutant plants, the extent of the morphological defects is relatively homogeneous in a population of *cmt3 drm1 drm2* triple mutants[24,70]. It is not known whether the developmental defects caused by loss of non-CG methylation can be inherited independently of the *cmt3* and *drm* mutations.

RNAi and maintenance of non-CG methylation

Plant genomes are distinguished by their significant levels of non-CG methylation. Within the *A. thaliana* genome, methylated loci vary in the extent to which they contain methylation in different sequence contexts. Indeed, silenced loci with higher levels of non-CG methylation can be silenced by this modification rather than by CG methylation[16,97,109]. Methyl cytosine in an asymmetric context differs fundamentally from methyl cytosine in a CG context. Asymmetric methylation lacks a methylated cytosine on the opposite strand, so the cell cannot use a hemimethylated template as a guide for replication, and probably requires an active signal for propagation during each cell division (FIG. 2). An excellent candidate for such an active signal is the persistent generation of siRNA corresponding to the methylated sequence.

RNA-guided DNA methylation initiated by RNA viruses and inverted-repeat transgenes has been observed in both CG and non-CG sequence contexts. A firm genetic link between siRNAs and the maintenance of non-CG methylation at a transcriptionally silenced gene emerged when *ago4* was identified as a

POLYCOMB GROUP (PcG). Genes in this group were identified as mutations in *Drosophila melanogaster*, which caused homeotic transformations. PcG proteins modify chromatin and maintain transcriptional decisions required for correct development.

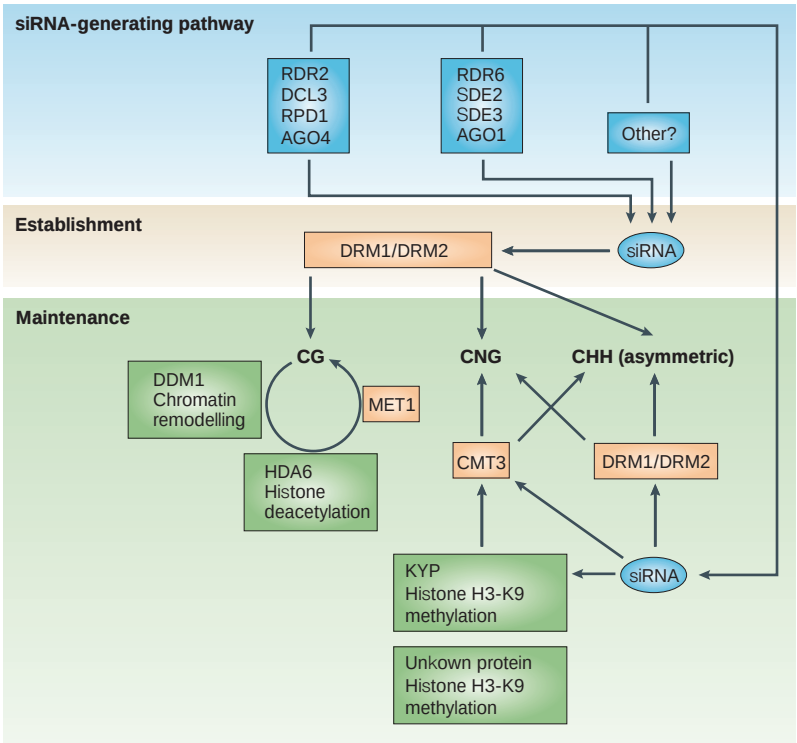

Figure 3 | **Establishment and maintenance of DNA methylation in *Arabidopsis thaliana*.** Different pathways control the maintenance and establishment of cytosine methylation in different sequence contexts. The maintenance of CG methylation requires a pathway that involves the DNA methyltransferase MET1, DECREASE IN DNA METHYLATION 1 (DDM1), chromatin remodelling and HDA6 histone-deacetylase activity. Maintenance of non-CG methylation (CNG and CHH) requires the methyltransferases CHROMOMETHYLASE 3 (CMT3) and DOMAINS REARRANGED METHYLTRANSFERASE 1 and 2 (DRM1/DRM2), which function with varying degrees of redundancy. CMT3 can be directed by histone H3-K9 methylation catalysed by KRYPTONITE (KYP) and an unknown H3-K27 methyltransferase. siRNA targeting can also direct CMT3 and DRM1/DRM2 to maintain non-CG methylation. Establishing methylation in all contexts is entirely dependent on DRM1/DRM2 methyltransferase activity. DRM1/DRM2 can be targeted to establish *de novo* methylation from siRNA generated from multiple branches of the RNAi pathway. AGO1, ARGONAUTE 1; AGO4, ARGONAUTE 4; DCL3, DICER-LIKE 3; RDR2, RNA-DEPENDENT RNA POLYMERASE 2; RDR6, RNA-DEPENDENT RNA POLYMERASE 6; RPD1, RNA POLYMERASE D1; SDE3, SILENCING DEFECTIVE 3; SGS3, SUPPRESSOR OF GENE SILENCING 3.

weak suppressor of hypermethylated *SUP*[110]. The *AGO4* gene was required for full levels of DNA and histone methylation at *SUP*, but *ago4* did not suppress silencing as completely as *cmt3* or *kyp*. Subsequently, it was shown that non-CG methylation at the tandemly repeated *FWA* and *MEA-ISR* loci requires the RPD1/RDR2/DCL3/AGO4 pathway[25]. So, at direct repeats, *rpd1*, *rdr2*, *dcl3* and *ago4* mutants phenocopy *drm1 drm2* double mutants, both to establish DNA methylation in all sequence contexts and to maintain CNG and asymmetric methylation. This indicates that siRNAs guide DRM2 to maintain non-CG DNA methylation. Studies in *S. pombe* have indicated that a feedback loop exists between chromatin modifications and siRNA production by RNA-dependent RNA polymerase[59,60,111]. A similar system might exist in *A. thaliana*, because the *drm1 drm2* double mutant has lower levels of *AtSN1* siRNA and because *ddm1* and *met1* mutants have reduced levels of siRNAs derived from some heterochromatic loci[22,26].

Although *AGO4* is required to maintain non-CG DNA methylation at tandem repeats, it has much weaker effects when an inverted-repeat trigger is used[25,26]. In one such system, the removal of an inverted-repeat trigger caused more extensive elimination of non-CG methylation than the presence of the *ago4-1* mutation[26]. This implies the existence of an *AGO4*-independent pathway that directs methylation in response to these inverted-repeat triggers. An attractive candidate for this alternative pathway is a redundant RNAi pathway that also directs DNA methylation. It will be interesting to explore how the mechanism of methylation distinguishes inverted and direct repeats.

The recent discovery of the *DEFECTIVE IN RNA DIRECTED DNA METHYLATION 1* (*DRD1*) gene — which encodes a plant-specific SWI2/SNF2 ATPase related to the widely conserved DNA-damage protein Rad54 — implies that chromatin remodelling has a role in the maintenance of siRNA-directed non-CG methylation[109]. A transcribed inverted repeat of the soybean α′-promoter provides an experimental system in which silencing is controlled by non-CG methylation[109]. A genetic screen that used this system revealed that RNA-directed non-CG methylation requires DRD1. Interestingly, *DRD1* is not required for non-CG methylation at the centromeric repeat or in the rDNA, indicating that its function might be restricted to propagating non-CG methylation at particular loci. The fact that *DRD1* was not isolated from either the *SUP* or *PAI* screens further underlines the fact that individual loci have specific silencing requirements.

Propagation of pre-existing non-CG DNA methylation by an active signal has caused problems in nomenclature that are essentially semantic, with some investigators referring to maintenance of non-CG methylation as 'persistent *de novo* methylation'. However, there is clearly an important distinction between *de novo* DNA methylation at a completely unmethylated locus (which is absolutely dependent on DRM methyltransferases) and mitotic replication of non-CG DNA methylation at a locus that was previously methylated (which is often partially independent of DRM methyltransferases)[24,25,51,52]. For this reason, we prefer the terms '*de novo* methylation' for the initial establishment of DNA methylation and 'maintenance of non-CG methylation' for the replication of pre-established patterns, even though these processes might share some mechanistic similarities. The establishment and maintenance of DNA methylation by DRM, MET1, and CMT3 enzymes and other factors are shown in FIG. 3.

Demethylation by DNA glycosylases

The establishment and maintenance phases of cytosine methylation are well studied in both mammals and plants. By contrast, the existence of a demethylation pathway has been an enduring controversy. Recent work in plants has identified a family of DNA glycosylases as proteins that can remove DNA methylation and alleviate silencing.

REPRESSOR OF SILENCING 1 (*ROS1*) was characterized in a screen that used a *COLD REGULATED 78*

(*COR78*) promoter — *LUCIFERASE* transgene — as a reporter[112]. The transgene promoter is homologous to the endogenous *COR78* gene and generates siRNA to these sequences[112]. In wild-type plants, these siRNAs are not sufficient to direct DNA methylation and transcriptional gene silencing to either the transgene or to the endogene. However, DNA methylation and silencing of both sequences occurs in a *ros1* mutant background[112]. This indicates that *ROS1* normally antagonizes the ability of the siRNA derived from the transgene to direct DNA methylation and gene silencing. *ROS1* encodes a nuclear protein that has similarity to DNA glycosylases of the HhH family[112]. DNA glycosylases can initiate base excision repair by cleaving the DNA backbone at the base removal site[113]. One role of ROS1-like glycosylases is in DNA repair, and indeed *ros1* mutant plants are hypersensitive to DNA-damage agents[112]. *In vitro* analysis with recombinant ROS1 also demonstrated that ROS1 was able to introduce strand breaks to an *Msp*I-methylated DNA template[112]. So, *ROS1* might function as a demethylase by removing methylated cytosine residues from DNA (FIG. 4). In this model, siRNA-targeted DNA methylation is continuously removed by the activity of ROS1 in DNA repair. This leaves an important question of how ROS1 is targeted to specific loci. Interestingly, the *ros1* mutant acquires developmental phenotypes following inbreeding, indicating that it regulates endogenous loci[112]. Nevertheless, *ROS1* does not appear to control methylation at silenced loci such as the centromere, rDNA or the *Athila* and *Ta3* transposons[112].

ROS1 is part of a small gene family in *A. thaliana*. Although *ROS1* is expressed broadly throughout development, a second glycosylase gene, *DEMETER* (*DME*), has a very specific expression pattern and function during female gametogenesis[17,112,114]. The *DME* gene was isolated as a mutation that showed parent-of-origin effects on seed viability[114]. Only maternal *DME* is required for viable seeds and this reflects its expression specifically in the central cell of the female gametophyte[114]. *DME* is required in the central cell and endosperm for the imprinted expression of the maternal genes *FWA* and *MEA*[17,114]. *MEA* encodes a polycomb protein that is required for seed viability and is expressed from the maternal allele[115]. Methylation of the *FWA* gene at its tandem repeats silences its expression in vegetative tissue[28]. However, *FWA* has maternally imprinted expression in the endosperm during seed development[17]. This activation is achieved by maternal-specific expression of *DME* in the female gametophyte, accompanied by the demethylation of *FWA*[17]. Because the endosperm is a terminally differentiating tissue, methylation does not need to be re-established at *FWA* in the next generation. This 'one-way' control of imprinting differs fundamentally from the methylation–demethylation cycles involved in mammalian imprinting[17]. At the genetic level, *DME* functions antagonistically to *MET1* in the control of seed development[116]. However, DME has not yet been shown to function as a demethylase at the biochemical level and might also function in processes that are independent of DNA methylation.

Conclusions

The distribution of DNA methylation in *A. thaliana* is consistent with its dual roles in genome defence and in gene regulation. DNA methylation is initiated by the DRM enzymes, which seem to be guided to particular loci by siRNAs. The maintenance of DNA methylation depends partly on the nature of repeated sequences at the target locus. Distinct molecular pathways involving chromatin-modifying enzymes and DNA methyltransferases cooperate to maintain DNA methylation in different sequence contexts.

Future experiments that focus on *A. thaliana* DNA methyltransferases are likely to yield insights that apply to DNA methylation in many organisms. A compelling mechanistic question is how siRNAs that are homologous to a given locus lead to chromatin modifications such as H3-K9 methylation and DNA methylation by the KYP and DRM proteins. The feedback loop that seems to maintain gene silencing in *S. pombe* might be too simplistic to apply to gene silencing in *A. thaliana* owing to the lack of DNA methylation in fission yeast[59,60,111]. For this reason, it will be crucially important to understand how the methyltransferases CMT3 and MET1 cooperate with histone modifying enzymes (such as H3-K9 and H3-K27 methyltransferases) and chromatin-remodelling proteins to maintain gene silencing.

Genome-wide studies of *A. thaliana* DNA methylation that are currently in progress are likely to generate new biological questions and reveal the true extent to which a genome-defence mechanism is used for developmental gene control[22,34,117,118].

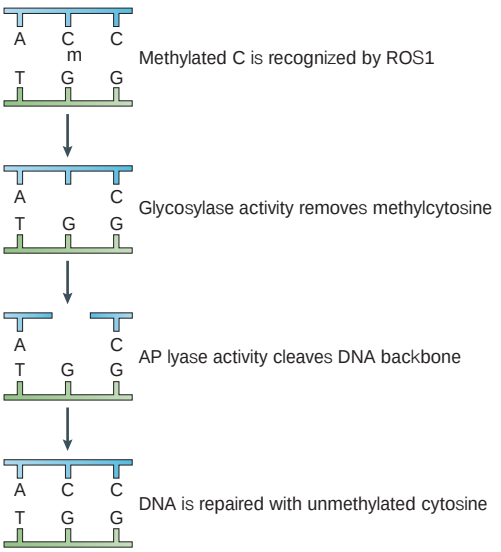

Figure 4 | **Mechanism of DNA demethylation by a DNA glycosylase/lyase.** Biochemical and genetic data indicate a model for the role of DNA glycosylase/lyase molecules, such as REPRESSOR OF SILENCING 1 (ROS1), in cytosine demethylation. ROS1 recognition of a methylated cytosine leads to glycosylase removal of the base from the DNA backbone. AP (apurinic) lyase activity then cleaves the DNA backbone at the site of cytosine removal, which is subsequently repaired.

1. Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**, 335–340 (1997).
2. Martienssen, R. A. & Colot, V. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* **293**, 1070–1074 (2001).
3. Singer, T., Yordan, C. & Martienssen, R. A. Robertson's *Mutator* transposons in *A. thaliana* are regulated by the chromatin-remodeling gene *DECREASE IN DNA METHYLATION* (*DDM1*). *Genes Dev.* **15**, 591–602 (2001).
4. Miura, A. *et al.* Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature* **411**, 212–214 (2001).
5. ten Lohuis, M. R. & Miller, D. J. Light-regulated transcription of genes encoding peridinin chlorophyll *a* proteins and the major intrinsic light-harvesting complex proteins in the dinoflagellate *Amphidinium carterae* hulburt (Dinophycae). *Plant Physiol.* **117**, 189–196 (1998).
6. Colot, V. & Rossignol, J. L. Eukaryotic DNA methylation as an evolutionary device. *Bioessays* **21**, 402–411 (1999).
7. Fronk, J. & Magiera, R. DNA methylation during differentiation of a lower eukaryote, *Physarum polycephalum*. *Biochem. J.* **304** (Part 1), 101–104 (1994).
8. Gutierrez, J. C., Callejas, S., Borniquel, S. & Martin-Gonzalez, A. DNA methylation in ciliates: implications in differentiation processes. *Int. Microbiol.* **3**, 139–146 (2000).
9. Fisher, O., Siman-Tov, R. & Ankri, S. Characterization of cytosine methylated regions and 5-cytosine DNA methyltransferase (Ehmeth) in the protozoan parasite *Entamoeba histolytica*. *Nucleic Acids Res.* **32**, 287–297 (2004).
10. Lyko, F., Whittaker, A. J., Orr-Weaver, T. L. & Jaenisch, R. The putative *Drosophila* methyltransferase gene *dDnmt2* is contained in a transposon-like element and is expressed specifically in ovaries. *Mech. Dev.* **95**, 215–217 (2000).
11. Kreppel, L. *et al.* dictyBase: a new *Dictyostelium discoideum* genome database. *Nucleic Acids Res.* **32** (Database issue), D332–D333 (2004).
12. Gutierrez, A. & Sommer, R. J. Evolution of *dnmt-2* and *mbd-2*-like genes in the free-living nematodes *Pristionchus pacificus*, *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *Nucleic Acids Res.* **32**, 6388–6396 (2004).
13. Cheng, X. Structure and function of DNA methyltransferases. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 293–318 (1995).
14. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
15. Stokes, T. L., Kunkel, B. N. & Richards, E. J. Epigenetic variation in *Arabidopsis* disease resistance. *Genes Dev.* **16**, 171–182 (2002).
16. Bender, J. & Fink, G. R. Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of *Arabidopsis*. *Cell* **83**, 725–734 (1995).
17. Kinoshita, T. *et al.* One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation. *Science* **303**, 521–523 (2004).
Demonstrates that *FWA* is a maternally imprinted gene in the endosperm. Monoallelic expression is controlled by *DME* and involves demethylation, establishing a new model of 'one-way' imprinting in plants.
18. Lawrence, R. J. *et al.* A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance. *Mol. Cell* **13**, 599–609 (2004).
19. Ramsahoye, B. H. *et al.* Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl Acad. Sci. USA* **97**, 5237–5242 (2000).
20. Hirochika, H., Okamoto, H. & Kakutani, T. Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation. *Plant Cell* **12**, 357–369 (2000).
21. The *Arabidopsis* Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
22. Lippman, Z. *et al.* Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**, 471–476 (2004).
Large-scale analysis of the heterochromatic knob on chromosome 4 that provides new insight into the relationships between DNA methylation, histone modifications and small RNAs.
23. Craig, N. L., Craigie, R. C., Gellert, M. & Lambowitz, A. M. *Mobile DNA II* 3–12 (American Society for Microbiology, Washington DC, 2002).
24. Cao, X. & Jacobsen, S. E. Locus-specific control of asymmetric and CpNpG methylation by the *DRM* and *CMT3* methyltransferase genes. *Proc. Natl Acad. Sci. USA* **99**, 16491–16498 (2002).
25. Chan, S. W. *et al.* RNA silencing genes control *de novo* DNA methylation. *Science* **303**, 1336 (2004).
Shows that, surprisingly, RNAi proteins are needed to establish DNA methylation at a directly repeated sequence.
26. Zilberman, D. *et al.* Role of *Arabidopsis* ARGONAUTE 4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr. Biol.* **14**, 1214–1220 (2004).
27. Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A. & Matzke, A. J. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* **19**, 5194–5201 (2000).
28. Soppe, W. J. *et al.* The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* **6**, 791–802. (2000).
29. Jones, L. *et al.* RNA–DNA interactions and DNA methylation in post-transcriptional gene silencing. *Plant Cell* **11**, 2291–2301 (1999).
30. Park, Y. D. *et al.* Gene silencing mediated by promoter homology occurs at the level of transcription and results in meiotically heritable alterations in methylation and gene activity. *Plant J.* **9**, 183–194 (1996).
31. Stam, M., Viterbo, A., Mol, J. N. & Kooter, J. M. Position-dependent methylation and transcriptional silencing of transgenes in inverted T-DNA repeats: implications for posttranscriptional silencing of homologous host genes in plants. *Mol. Cell Biol.* **18**, 6165–6177 (1998).
32. Sieburth, L. E. & Meyerowitz, E. M. Molecular dissection of the *AGAMOUS* control region shows that *cis* elements for spatial regulation are located intragenically. *Plant Cell* **9**, 355–365 (1997).
33. Ito, T., Sakai, H. & Meyerowitz, E. M. Whorl-specific expression of the *SUPERMAN* gene of *Arabidopsis* is mediated by *cis* elements in the transcribed region. *Curr. Biol.* **13**, 1524–1530 (2003).
34. Tran, R. K. *et al.* DNA methylation profiling identifies CG methylation clusters in *Arabidopsis* genes. *Curr. Biol.* **15**, 154–159 (2005).
35. Rountree, M. R. & Selker, E. U. DNA methylation inhibits elongation but not initiation of transcription in *Neurospora crassa*. *Genes Dev.* **11**, 2383–2395 (1997).
36. Wassenegger, M., Heimes, S., Riedel, L. & Sanger, H. L. RNA-directed *de novo* methylation of genomic sequences in plants. *Cell* **76**, 567–576 (1994).
The first demonstration that RNA can guide DNA methylation in any organism.
37. Angell, S. M. & Baulcombe, D. C. Consistent gene silencing in transgenic plants expressing a replicating potato virus X RNA. *EMBO J.* **16**, 3675–3684 (1997).
38. Dalmay, T., Hamilton, A., Mueller, E. & Baulcombe, D. C. Potato virus X amplicons in *Arabidopsis* mediate genetic and epigenetic gene silencing. *Plant Cell* **12**, 369–379 (2000).
39. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343–349 (2004).
40. Mourrain, P. *et al. Arabidopsis SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* **101**, 533–542 (2000).
41. Dalmay, T., Horsefield, R., Braunstein, T. H. & Baulcombe, D. C. *SDE3* encodes an RNA helicase required for post-transcriptional gene silencing in *Arabidopsis*. *EMBO J.* **20**, 2069–2078 (2001).
42. Dalmay, T., Hamilton, A., Rudd, S., Angell, S. & Baulcombe, D. C. An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* **101**, 543–553 (2000).
43. Mochizuki, K., Fine, N., Fujisawa, T. & Gorovsky, M. Analysis of a *piwi*-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. *Cell* **110**, 689 (2002).
44. Yao, M. C., Fuller, P. & Xi, X. Programmed DNA deletion as an RNA-guided system of genome defense. *Science* **300**, 1581–1584 (2003).
45. Taverna, S., Coyne, R. & Allis, C. Methylation of histone H3 at lysine 9 targets programmed DNA elimination in *Tetrahymena*. *Cell* **110**, 701 (2002).
46. Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
Pioneering study revealing that small RNAs in the RNAi pathway can transcriptionally silence genes and are required for centromeric heterochromatin in *S. pombe*.
47. Kawasaki, H. & Taira, K. Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature* **431**, 211–217 (2004).
48. Morris, K. V., Chan, S. W., Jacobsen, S. E. & Looney, D. J. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**, 1289–1292 (2004).
49. Matzke, M. A. & Birchler, J. A. RNAi-mediated pathways in the nucleus. *Nature Rev. Genet.* **6**, 24–35 (2005).
50. Cao, X. *et al.* Conserved plant genes with similarity to mammalian *de novo* DNA methyltransferases. *Proc. Natl Acad. Sci. USA* **97**, 4979–4984 (2000).
51. Cao, X. *et al.* Role of the *DRM* and *CMT3* methyltransferases in RNA-directed DNA methylation. *Curr. Biol* **13**, 2212–2217 (2003).
52. Cao, X. & Jacobsen, S. E. Role of the *Arabidopsis* DRM methyltransferases in *de novo* DNA methylation and gene silencing. *Curr. Biol.* **12**, 1138–1144 (2002).
Characterization of the role of the plant *de novo* DRM methyltransferases in establishing gene silencing.
53. Hamilton, A., Voinnet, O., Chappell, L. & Baulcombe, D. Two classes of short interfering RNA in RNA silencing. *EMBO J.* **21**, 4671–4679 (2002).
Provides a new insight into the diversity of small RNA pathways in plants, revealing that there are two main size classes of siRNA associated with different modes of silencing.
54. Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, e104 (2004).
Genetic demonstration that distinct RNAi proteins are specialized for different functions in development, gene silencing and viral resistance.
55. Schauer, S. E., Jacobsen, S. E., Meinke, D. W. & Ray, A. *DICER-LIKE 1*: blind men and elephants in *Arabidopsis* development. *Trends Plant Sci.* **7**, 487–491 (2002).
56. Carmell, M. A., Xuan, Z., Zhang, M. Q. & Hannon, G. J. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* **16**, 2733–2742 (2002).
57. Vaucheret, H., Vazquez, F., Crete, P. & Bartel, D. P. The action of *ARGONAUTE 1* in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Dev.* **18**, 1187–1197 (2004).
58. Malagnac, F., Bartee, L. & Bender, J. An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation. *EMBO J.* **21**, 6842–6852 (2002).
Together with reference 102, this paper reveals that histone methylation functions to control CNG DNA methylation in *A. thaliana*. This paper also demonstrates genetically that the *A. thaliana* homologue of *HETEROCHROMATIN PROTEIN 1* is not involved in this pathway.
59. Sugiyama, T., Cam, H., Verdel, A., Moazed, D. & Grewal, S. I. RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc. Natl Acad. Sci. USA* **102**, 152–157 (2005).
60. Motamedi, M. R. *et al.* Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**, 789–802 (2004).
61. Herr, A. J., Jensen, M. B., Dalmay, T. & Baulcombe, D. C. RNA polymerase IV directs silencing of endogenous DNA. *Science* 24 February 2005 (doi:10.1126/science.1106910).
62. Onodera, Y. *et al.* Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**, 613–622 (2005).
References 61 and 62 provide an exciting role for RNA pol IV in transcriptional silencing and support a model in which surveillance transcription is used to direct DNA methylation.
63. Fransz, P., De Jong, J. H., Lysak, M., Castiglione, M. R. & Schubert, I. Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl Acad. Sci. USA* **99**, 14584–14589 (2002).
64. Soppe, W. J. *et al.* DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*. *EMBO J.* **21**, 6549–6559 (2002).
65. Ronemus, M. J., Galbiati, M., Ticknor, C., Chen, J. & Dellaporta, S. L. Demethylation-induced developmental pleiotropy in *Arabidopsis*. *Science* **273**, 654–657 (1996).
66. Finnegan, E. J., Peacock, W. J. & Dennis, E. S. Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. *Proc. Natl Acad. Sci. USA* **93**, 8449–8454 (1996).
67. Saze, H., Scheid, O. M. & Paszkowski, J. Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nature Genet.* **34**, 65–69 (2003).
Null *met1* mutants display immediate effects on gene silencing owing to a loss of DNA methylation in the gametophytic generation. Reference 81 extends this study to reveal that CG methylation also controls non-CG methylation and histone methylation.
68. Kankel, M. W. *et al. Arabidopsis* MET1 Cytosine methyltransferase mutants. *Genetics* **163**, 1109–1122 (2003).
69. Jacobsen, S. E., Sakai, H., Finnegan, E. J., Cao, X. & Meyerowitz, E. M. Ectopic hypermethylation of flower-specific genes in *Arabidopsis*. *Curr. Biol.* **10**, 179–186 (2000).

70. Kakutani, T., Jeddeloh, J. A., Flowers, S. K., Munakata, K. & Richards, E. J. Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. *Proc. Natl Acad. Sci. USA* **93**, 12406–12411 (1996).
71. Jeddeloh, J. A., Stokes, T. L. & Richards, E. J. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nature Genet.* **22**, 94–97 (1999).
72. Brzeski, J. & Jerzmanowski, A. Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors. *J. Biol. Chem.* **278**, 823–828 (2003).
73. Gendrel, A. V., Lippman, Z., Yordan, C., Colot, V. & Martienssen, R. A. Dependence of heterochromatic histone H3 methylation patterns on the *Arabidopsis* gene *DDM1*. *Science* **297**, 1871–1873 (2002).
74. Johnson, L., Cao, X. & Jacobsen, S. Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr. Biol.* **12**, 1360–1367 (2002).
75. Lippman, Z., May, B., Yordan, C., Singer, T. & Martienssen, R. Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol.* **1**, e67 (2003).
76. Dennis, K., Fan, T., Geiman, T., Yan, Q. & Muegge, K. Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev.* **15**, 2940–2944 (2001).
77. Aufsatz, W., Mette, M. F., Van Der Winden, J., Matzke, M. & Matzke, A. J. HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. *EMBO J.* **21**, 6832–41 (2002).
78. Murfett, J., Wang, X. J., Hagen, G. & Guilfoyle, T. J. Identification of *Arabidopsis* histone deacetylase HDA6 mutants that affect transgene expression. *Plant Cell* **13**, 1047–10461 (2001).
79. Probst, A. V. *et al. Arabidopsis* histone deacetylase HDA6 is required for maintenance of transcriptional gene silencing and determines nuclear organization of rDNA repeats. *Plant Cell* **16**, 1021–1034 (2004).
80. Kurdistani, S. K. & Grunstein, M. Histone acetylation and deacetylation in yeast. *Nature Rev. Mol. Cell Biol.* **4**, 276–284 (2003).
81. Tariq, M. *et al.* Erasure of CpG methylation in *Arabidopsis* alters patterns of histone H3 methylation in heterochromatin. *Proc. Natl Acad. Sci. USA* **100**, 8823–8827 (2003).
See reference 67.
82. Lindroth, A. M. *et al.* Requirement of CHROMOMETHYLASE 3 for maintenance of CpXpG methylation. *Science* **292**, 2077–2080 (2001).
Together with reference 99, this paper genetically characterizes CMT3 as the plant methyltransferase that functions to maintain CNG DNA methylation.
83. Finnegan, E. J. & Dennis, E. S. Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. *Nucleic Acids Res.* **21**, 2383–2388 (1993).
84. Kishimoto, N. *et al.* Site specificity of the *Arabidopsis* METI DNA methyltransferase demonstrated through hypermethylation of the superman locus. *Plant Mol. Biol.* **46**, 171–183 (2001).
85. Scebba, F. *et al. Arabidopsis* MBD proteins show different binding specificities and nuclear localization. *Plant Mol. Biol.* **53**, 715–731 (2003).
86. Ito, M., Koike, A., Koizumi, N. & Sano, H. Methylated DNA-binding proteins from *Arabidopsis*. *Plant Physiol.* **133**, 1747–1754 (2003).
87. Berg, A. *et al.* Ten members of the *Arabidopsis* gene family encoding methyl-CpG-binding domain proteins are transcriptionally active and at least one, *AtMBD11*, is crucial for normal development. *Nucleic Acids Res.* **31**, 5291–5304 (2003).
88. Zemach, A. & Grafi, G. Characterization of *Arabidopsis thaliana* methyl-CpG-binding domain (MBD) proteins. *Plant J.* **34**, 565–572 (2003).
89. Ng, H. H. *et al.* MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nature Genet.* **23**, 58–61 (1999).
90. Fuks, F., Hurd, P. J., Deplus, R. & Kouzarides, T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res.* **31**, 2305–2312 (2003).
91. Fuks, F. *et al.* The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J. Biol. Chem.* **278**, 4035–4040 (2003).
92. Aufsatz, W., Mette, M. F., Van Der Winden, J., Matzke, A. J. & Matzke, M. RNA-directed DNA methylation in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99**, 16499–16506 (2002).
93. Jones, L., Ratcliff, F. & Baulcombe, D. C. RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. *Curr. Biol.* **11**, 747–757 (2001).
94. Aufsatz, W., Mette, M. F., Matzke, A. J. & Matzke, M. The role of MET1 in RNA-directed *de novo* and maintenance methylation of CG dinucleotides. *Plant Mol. Biol.* **54**, 793–804 (2004).
95. Wada, Y., Ohya, H., Yamaguchi, Y., Koizumi, N. & Sano, H. Preferential *de novo* methylation of cytosine residues in non-CpG sequences by a domains rearranged DNA methyltransferase from tobacco plants. *J. Biol. Chem.* **278**, 42386–42393 (2003).
96. Gruenbaum, Y., Naveh-Many, T., Cedar, H. & Razin, A. Sequence specificity of methylation in higher plant DNA. *Nature* **292**, 860–862 (1981).
97. Jacobsen, S. E. & Meyerowitz, E. M. Hypermethylated *SUPERMAN* epigenetic alleles in *Arabidopsis*. *Science* **277**, 1100–1103 (1997).
98. Henikoff, S. & Comai, L. A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in *Arabidopsis*. *Genetics* **149**, 307–318 (1998).
99. Bartee, L., Malagnac, F. & Bender, J. *Arabidopsis cmt3* chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. *Genes Dev.* **15**, 1753–1758 (2001).
See reference 82.
100. Melquist, S. & Bender, J. Transcription from an upstream promoter controls methylation signaling from an inverted repeat of endogenous genes in *Arabidopsis*. *Genes Dev.* **17**, 2036–2047 (2003).
101. Papa, C. M., Springer, N. M., Muszynski, M. G., Meeley, R. & Kaeppler, S. M. Maize chromomethylase *Zea methyltransferase 2* is required for CpNpG methylation. *Plant Cell* **13**, 1919–1928 (2001).
102. Jackson, J. P., Lindroth, A. M., Cao, X. & Jacobsen, S. E. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**, 556–560 (2002).
See reference 58.
103. Lindroth, A. M. *et al.* Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE 3. *EMBO J.* **23**, 4286–4296 (2004).
104. Jackson, J. P. *et al.* Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in *Arabidopsis thaliana*. *Chromosoma* **112**, 308–315 (2004).
105. Citterio, E. *et al.* Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Mol. Cell Biol.* **24**, 2526–2535 (2004).
106. Tamaru, H. & Selker, E. U. A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* **414**, 277–283 (2001).
The first clear genetic demonstration that DNA methylation can be controlled by histone methylation.
107. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
108. Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131–135 (2003).
109. Kanno, T. *et al.* Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Curr. Biol.* **14**, 801–805 (2004).
Intriguing report describing a role for chromatin remodelling in RNA-directed methylation. DRD1 appears to be specialized for the control of methylation at non-CG sequences.
110. Zilberman, D., Cao, X. & Jacobsen, S. E. ARGONAUTE 4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**, 716–719 (2003).
An ARGONAUTE protein controls DNA methylation at a transcriptionally silenced gene, linking RNAi and heterochromatin in *A. thaliana*.
111. Schramke, V. & Allshire, R. Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. *Science* **301**, 1069–1074 (2003).
112. Gong, Z. *et al.* ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. *Cell* **111**, 803–814 (2002).
An exciting report that characterizes a role for DNA glyscosylase/lyase protein ROS1 in antagonizing DNA methylation and gene silencing. The first genetic characterization of the much sought after DNA demethylation pathway.
113. Krokan, H. E., Standal, R. & Slupphaug, G. DNA glycosylases in the base excision repair of DNA. *Biochem. J.* **325** (Part 1), 1–16 (1997).
114. Choi, Y. *et al.* DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell* **110**, 33–42 (2002).
DEMETER is a DNA glycosylase/lyase protein, related to ROS1, that has a specialized function in control of monoallelic expression states. Together with reference 17, this paper provides new mechanistic insight into imprinting and a putative demethylation pathway.
115. Grossniklaus, U., Vielle-Calzada, J. P., Hoeppner, M. A. & Gagliano, W. B. Maternal control of embryogenesis by MEDEA, a polycomb group gene in *Arabidopsis*. *Science* **280**, 446–450 (1998).
116. Xiao, W. *et al.* Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Dev. Cell* **5**, 891–901 (2003).
117. Tompa, R. *et al.* Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE 3. *Curr. Biol.* **12**, 65–68 (2002).
118. Mockler, T. C. *et al.* Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**, 1–15 (2005).
119. Vongs, A., Kakutani, T., Martienssen, R. A. & Richards, E. J. *Arabidopsis thaliana* DNA methylation mutants. *Science* **260**, 1926–1928 (1993).
Groundbreaking forward genetic screen for DNA methylation mutants in *A. thaliana*, which isolated key factors controlling CG DNA methylation.
120. Bao, N., Lye, K. W. & Barton, M. K. MicroRNA binding sites in *Arabidopsis* class III HD-ZIP mRNAs are required for methylation of the template chromosome. *Dev. Cell* **7**, 653–662 (2004).
121. Ashapkin, V. V., Kutueva, L. I. & Vanyushin, B. F. The gene for domains rearranged methyltransferase (DRM2) in *Arabidopsis thaliana* plants is methylated at both cytosine and adenine residues. *FEBS Lett.* **532**, 367–372 (2002).

Acknowledgements

We apologize to the many authors whose work was not cited because of space limitations. DNA-methylation studies in our laboratory have been supported by Searle Scholar, Beckman Young Investigator and National Institutes of Health grants to S.E.J. S.W.-L.C. is a DoE Energy Biosciences fellow of the Life Sciences Research Foundation. I.R.H. is supported by an EMBO postdoctoral fellowship.

Competing interests statement

The authors declare no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:
The Arabidopsis Information Resource:
http://www.arabidopsis.org/servlets/Search?action=new_search&type=gene
AG | *AGO4* | *BAL* | CMT3 | *DCL3* | *DDM1* | *DRM1* | *DRM2* | *FWA* | *HDA6* | HDT1 | *KYP* | LHP1 | *MEDEA* | *MET1* | *PAI1* | *PAI2* | *PAI4* | *RDR2* | *ROS1* | RPD1 | *SUP*

FURTHER INFORMATION

Jacobsen laboratory home page:
http://www.mcdb.ucla.edu/Research/Jacobsen/index.html
Access to this interactive links box is free online.