New Phytologist

Review

Tansley insight

DNA methylation in *Marchantia polymorpha*

Authors for correspondence:
Mario Alberto Arteaga-Vázquez
Tel: +52 228 8421700 Ext 10880
Email: maarteaga@uv.mx

Daniel Grimanelli
Tel: +33 (0)4 6741 6376
Email: daniel.grimanelli@ird.fr

Jim Haseloff
Tel: +44 (0)1223 766 546
Email: jh295@cam.ac.uk

Received: *29 November 2018*
Accepted: *13 March 2019*

Adolfo Aguilar-Cruz[1], Daniel Grimanelli[2], Jim Haseloff[3] and Mario Alberto Arteaga-Vázquez[1]

[1]Instituto de Biotecnología y Ecología Aplicada, Universidad Veracruzana, Avenida de las Culturas Veracruzanas 101, Col. Emiliano Zapata, C.P. 91090, Xalapa, Veracruz, México; [2]Institut de Recherche pour le Développement (IRD), UMR232, Université de Montpellier, Montpellier 34394, France; [3]Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EA, UK

Contents

New Phytologist (2019) **223:** 575–581
doi: 10.1111/nph.15818

Key words: DNA methylation, epigenetic reprogramming, epigenetics, *Marchantia polymorpha*, RNA-directed DNA methylation (RdDM).

Summary

Methylation of DNA is an epigenetic mechanism for the control of gene expression. Alterations in the regulatory pathways involved in the establishment, perpetuation and removal of DNA methylation can lead to severe developmental alterations. Our understanding of the mechanistic aspects and relevance of DNA methylation comes from remarkable studies in well-established angiosperm plant models including maize and Arabidopsis. The study of plant models positioned at basal lineages opens exciting opportunities to expand our knowledge on the function and evolution of the components of DNA methylation. In this Tansley Insight, we summarize current progress in our understanding of the molecular basis and relevance of DNA methylation in the liverwort *Marchantia polymorpha*.

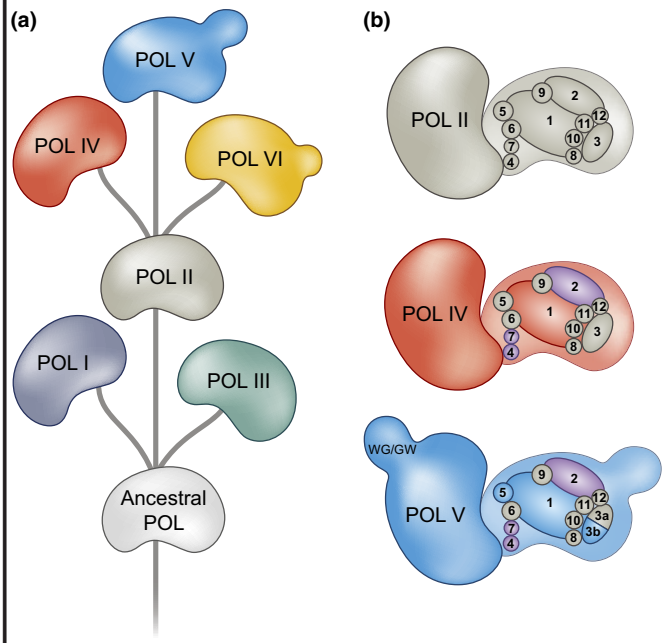
I. Introduction

DNA methylation is an important epigenetic mechanism of control of gene expression that is based on the covalent transfer of a methyl (-CH_3) group to the fifth carbon of a cytosine (C) residue on the DNA sequence to form 5-methylcytosine (5mC). Cells employ DNA methylation in association with mechanisms that regulate gene expression to silence mobile fragments of DNA known as transposable elements (TEs). TEs are highly abundant in plants and, based on their ability to reshape the host genome through mutagenesis and establishment of new states of gene expression, they constitute a major evolutionary driving force (McClintock, 1956; Lanciano & Mirouze, 2018).

II. Establishing DNA methylation through the RNA-directed DNA methylation (RdDM) pathway

In plants, DNA methylation can occur in three sequence contexts: CG, CHG and CHH (where H = A, T or C). Functional

Box 1 Plants have specialized RNA polymerases involved in DNA methylation.



Eukaryotes have three essential nuclear RNA polymerases. NRPA or POLI transcribes large ribosomal RNA, NRPB or POLII transcribes messenger RNAs (mRNAs), and NRPC or POLIII transcribes precursors of transfer RNA (tRNA), 5S ribosomal RNA and other noncoding RNA. (a) Plants contain two RNA polymerases (NRPD or POLIV and NRPE or POLV) specialized in gene silencing and control of transposable elements that evolved from POLII and function at the core of the RNA-directed DNA methylation pathway (RdDM) (see Fig. 1). (b) POLII, POLIV and POLV are composed of 12 subunits (seven of them encoded by the same genes). Subunits 6, 8, 9, 10, 11 and 12 are shared among all three polymerases (colored in gray). POLIV subunit 1 is the only POLIV-specific subunit (colored in red). POLV has two distinct subunits (1 and 5) and employs nearly equally two alternative forms of the third subunit (3a and 3b, colored in gray and blue, respectively). Both POLIV and POL V share three subunits (2, 4 and 7) (colored in purple). POLV has a distinctive carboxy terminal domain (CTD) with tryptophan-glycine/glycine-tryptophan (WG/GW), responsible for mediating interactions with members of the ARGONAUTE (AGO) family. It is important to highlight that despite the presence of shared subunits, POLIV produces 26–45 nucleotide short transcripts and POLV transcripts are long with different 5′-ends with modifications characteristic of POLII transcripts such as 7-methylguanosine caps at their 5′-end but that lack polyadenylation at their 3′-ends. Recently, NRPF or POLVI was discovered in the grass family Poacea (see main text) but its function is yet to be characterized (the subunit composition of POLVI is not depicted here).

information residing in the sequence motif (H) adjacent to the C can influence both the frequency of methylation and the recruitment of specific methyltransferases acting upon the trinucleotide (Law & Jacobsen, 2010; Gouil & Baulcombe, 2016). The RNA-directed DNA methylation (RdDM) pathway employs specialized plant-specific RNA polymerases to establish DNA methylation: NRPD or POLIV, NRPE or POLV and the recently described grass- (Poacea) specific NRPF or POLVI (Box 1) (Haag & Pikaard, 2011; Zhou & Law, 2015; Trujillo *et al.*, 2018). RNA POLYMERASE IV (POLIV) transcribes short (26–45 nucleotides) single-stranded RNA (ssRNA) from the template/target loci. The ssRNA then serves as a substrate for RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) that physically interacts with POLIV and immediately converts it into double-stranded RNA (dsRNA). The dsRNA is subsequently processed into 24-nucleotide small interfering RNAs (siRNAs) by the RNase DICER-LIKE 3 (DCL3). The resulting pool of 24-nucleotide siRNAs is methylated at their 3′-end by HUA ENHANCER 1 (HEN1) and recruited by ARGONAUTE 4 (AGO4) (Zilberman *et al.*, 2003) and other members including AGO6 and AGO9 (Havecker *et al.*, 2010) and function as guides to direct an AGO4–siRNA complex to transcripts generated by POLV, another plant-specific RNA polymerase also specialized in transcriptional regulation (Box 1), through an RNA–RNA pairing interaction between the siRNA present in the AGO4–siRNA complex with the chromatin bound POLV-dependent transcripts and a protein–protein interaction between AGO4 and the C-terminal domain (CTD) of POLV (reviewed by Matzke & Mosher, 2014). Recruitment of the DNA methyltransferase DOMAINS REARRANGED 2 (DRM2) to the template/target DNA and subsequent *de novo* methylation of DNA in all sequence contexts (Cao & Jacobsen, 2002) occur through RNA-DIRECTED DNA METHYLATION 1 (RDM1) that can bind methylated ssDNA and links AGO4 with DRM2 (Gao *et al.*, 2010) (Fig. 1). The molecular basis of the mechanisms positioning POLIV and POLV at their target loci are not yet fully understood as no consensus binding motifs have been identified by chromatin occupancy mapping. However, it has been shown that POLIV can be recruited to loci harboring chromatin with methylated residues in the lysine 9 of histone 3 (H3K9me) through its interaction with SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1) (Law *et al.*, 2013), which can bind H3K9me. Recent evidence from Arabidopsis indicates that CLASSY1 (CLSY1) and CLSY2 are required for the association between SHH1 and POLIV and, in fact, all members of the CLSY family are absolutely required for the locus-specific positioning of POLIV through chromatin features (including CG and H3K9 methylation) and are also required for the generation of practically all 24-nucleotide siRNAs (Zhou *et al.*, 2018). By contrast, POLV interacts with DEFECTIVE IN RNA DIRECTED DNA METHYLATION 1 (DRD1) (Law *et al.*, 2010), which also interacts with two histone methyltransferases belonging to the SUPPRESSOR OF VARIEGATION [SU(VAR)] HOMOLOG (SUVH) family, namely SUVH2 and SUVH9, that bind methylated DNA (but seem to have lost their methyltransferase activity) (Johnson *et al.*, 2014) (Fig. 1). Thus, while details of the molecular mechanisms targeting *de novo* DNA methylation at loci with pre-existing patterns of H3K9 and/or DNA methylation are emerging, the establishment of such pre-existing patterns, and ultimately the establishment of DNA methylation at naïve loci without pre-existing silencing-associated epigenetic marks is yet to be described.

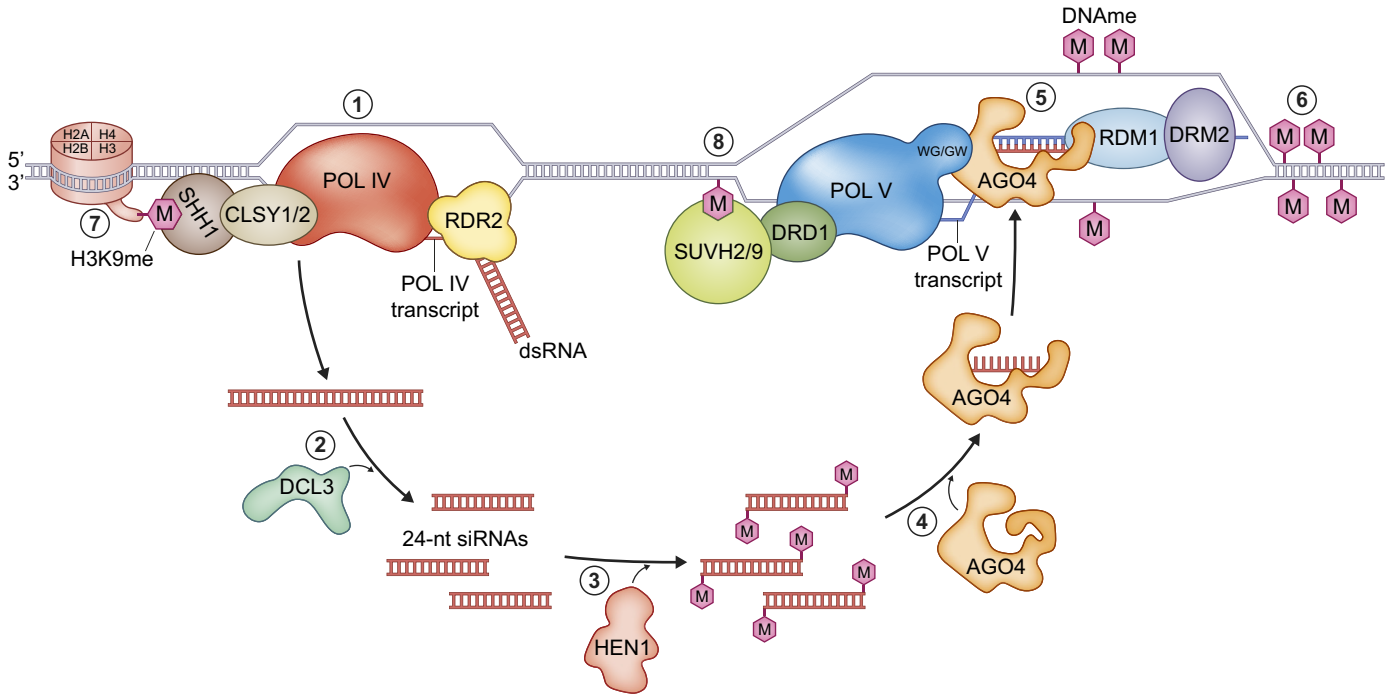


Fig. 1 The RNA-directed DNA methylation pathway (RdDM). (1) RNA polymerase IV (POLIV) generates short single-stranded RNA (ssRNA) that is immediately converted into double-stranded RNA (dsRNA) by the physically coupled activity of RNA DEPENDENT RNA POLYMERASE 2 (RDR2). (2) dsRNA is processed into 24-nucleotide small interfering RNAs (siRNAs) by DICER-LIKE 3 (DCL3). (3) The short duplexes are subsequently methylated at their 3′-ends by HUA ENHANCER 1 (HEN1). (4) A single strand of the 24-nucleotide siRNA duplex is integrated into ARGONAUTE 4 (AGO4) (and into other members of the family including AGO6 and AGO9). (5) RNA POLYMERASE V (POLV) itself helps recruit the activity of the *de novo* DNA methyltransferase DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) by generating long ssRNA that through Watson–Crick base pairing interacts with the 24-nucleotide siRNA incorporated in AGO4 and by physically interacting with AGO4 through the tryptophan-glycine/glycine-tryptophan (WG/GW) platform (or AGO-hook domain). The plant-specific RNA-DIRECTED DNA METHYLATION 1 (RDM1) protein that is able to bind methylated ssDNA helps establish the interaction between AGO4 and DRM2. (6) DRM2 methylates DNA in three sequence contexts. (7) Recent evidence indicates that CLASSY (CLSY) proteins are required for the generation of the vast majority of 24-nucleotide siRNAs. CLSY3 and CLSY4 (not depicted here) rely on CG methylation for the generation of a subset of 24-nucleotide siRNAs. CLSY1 and CLSY2 are required for the physical association between SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1) (which binds to methylated lysine 9 in histone 3 (H3K9me)) and POLIV, for the generation of a large fraction of 24-nucleotide siRNAs. (8) POLV can be recruited to regions with pre-existing DNA methylation through indirect interaction with the histone methyltransferases SU(VAR) HOMOLOG 2 (SUVH2) and SUVH9 (able to bind methylated DNA but that do not show methyltransferase activity) through the interaction with DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1) and this interaction is thought to reinforce DNA methylation.

III. Maintenance and removal of DNA methylation

The newly established patterns of DNA methylation are maintained and transmitted across mitotic divisions by different mechanisms. Maintenance of CG sites occurs by the activity of members of the family VARIANT IN METHYLATION 1-3 (VIM1-3) that bind hemimethylated CG sites on the parental strand and recruit METHYLTRANSFERASE 1 (MET1) to methylate the complementary CG sites on the daughter strand (Vongs *et al.*, 1993; Woo *et al.*, 2007, 2008). Evidence from *in vitro* experiments has shown that VIM1 can bind both CG and CHG sites, suggesting that MET1 might be involved in maintaining methylation at the CHG context *in planta* either at a particular set of loci or in a time/cell-specific fashion. By contrast, maintenance of CHG is enforced by the activity of CHROMOMETHYLASE 3 (CMT3) and CMT2 through a feedback loop acting upon the crosstalk between both DNA methyltransferases and the activity of members of the SUVH family responsible for the deposition of H3K9me2 (reviewed by Stroud *et al.*, 2014). While SUVH4 (also known as KRYPTONITE) (Jackson *et al.*, 2002) is responsible for the majority of H3K9me2, other members including SUVH5 and SUVH6 also contribute (reviewed by Li *et al.*, 2018). This results in the methylation of DNA in CWA and CWG contexts (W = A or T) by CMT2 and CMT3, respectively.

The particular pattern of DNA methylation in the CG context located between the transcription start site (TSS) and transcription termination site (TTS) of genes present in angiosperms is referred to as gene body methylation (gbM) (Bewick & Schmitz, 2017). In Arabidopsis, gbM is linked to a relative enrichment in linker histone H1 and histone variant H3.3 (Wollmann *et al.*, 2017) and most of the genes affected in H3.3 knockdown alleles correspond to genes involved in either environmental response or development. Interestingly, there is also evidence that loss of function in the *increase in bonsai methylation 1* (*IBM1*) gene affects gbM negatively, replacing CG methylation with CHG methylation and H3K9me2 (Saze *et al.*, 2008). In Arabidopsis, rice and maize, around *c.* 30% of their genes show gbM and even though gbM has been reported as a feature present in orthologous genes across different species and it has also been associated with constitutive levels of expression (reviewed by Bewick & Schmitz, 2017), its functional role is

currently unknown. gbM is dependent on the function of MET1 (reviewed by Bewick & Schmitz, 2017) but recent evidence suggests that CMT3 might be responsible for long-term maintenance of gene body methylation because the absence of CMT3 across a number of angiosperm phyla correlates with concomitant losses of gbM (Bewick *et al.*, 2017).

Removal of DNA methylation (demethylation) can occur either through the enzymatic function of DNA glycosylases that replace 5mC with cytosine or passively in the absence of DNA methyltransferase activity or due to shortage of methyl group supplies (Zhang & Zhu, 2012). In Arabidopsis, there are four DNA glycosylases able to recognize and remove 5mC from all sequence contexts, namely REPRESSOR OF SILENCING 1 (ROS1), DEMETER (DME), DEMETER-LIKE 2 (DML2) and DML3 (Ortega-Galisteo *et al.*, 2008). While ROS1, DML2 and DML3 are expressed in nearly all vegetative tissues, the expression of DME is preferentially expressed in the gametes where it helps establish differential patterns of DNA methylation at some imprinted loci (i.e. loci expressed in a contrasting manner depending on whether they are inherited from the mother or the father). ROS1 is the founding member of the family and in addition to its regulatory roles in diverse developmental programs it is also involved in responses to biotic and abiotic stress (reviewed by Zhang *et al.*, 2018). ROS1 is involved in establishing boundaries between TEs and genes at thousands of loci by preventing the spread of DNA methylation and DNA hypermethylation as it counteracts RdDM through a postulated mechanism based on the ability of ROS1 to demethylate its own promoter from RdDM hypermethylation (Zhang *et al.*, 2018). Recently, in addition to its role in the generation of 24-nucleotide siRNAs (Zhou *et al.*, 2018), CLSY4 was found to function in DNA demethylation in the same genetic pathway of the ROS1 family (Yang *et al.*, 2018).

IV. Noncanonical RdDM

RdDM can also be achieved through alternative routes feeding into the RdDM pathway that involve the activity of diverse populations of small RNAs with different modes of biogenesis but that ultimately rely on the function of DRM2 and POLV and their associated noncoding transcripts (reviewed by Cuerda-Gil & Slotkin, 2016). These routes include the RDR6-RdDM pathway that targets transcriptionally active TEs and depends on POLII-generated noncoding transcripts that are cleaved by the activity of microRNAs (miRNAs) and used as substrate by RNA-DEPENDENT RNA POLYMERASE 6 (RDR6) to generate dsRNA. This dsRNA is processed by DCL1, DCL2 or DCL4 into 21–22 nucleotide siRNAs that are recruited by AGO6. The AGO6–21 nucleotide siRNA complex subsequently feeds into RdDM by the interaction with POLV and DRM2 (Nuthikattu *et al.*, 2013). The second route is the RDR6-DCL3-RdDM pathway that targets high copy TEs and transgenes and basically differs from the RDR6-RdDM pathway in the use of DCL3 for the generation of 24-nucleotide siRNAs from RDR6-dependent dsRNA. Third, the POLIV-NERD pathway targets a subset of intergenic loci by the NEEDED FOR RDR2 INDEPENDENT METHYLATION (NERD)-mediated interaction of AGO2 with POLIV and POLV, and employs siRNAs whose biogenesis depends on canonical and noncanonical components including RDR1, RDR6, DCL2, DCL3 and DCL4 (Pontier *et al.*, 2012). Finally, the recently described DICER-independent RdDM pathway is postulated to target TEs and relies on the generation of 21 and 24 nucleotide siRNAs from dsRNA bound to AGO4 and trimmed to their final size by the activity of exosome–core complex exonucleases (Yang *et al.*, 2016).

V. Mechanisms of DNA methylation in *Marchantia*

Bryophytes (liverworts, mosses and hornworts) are among the plants that first colonized the land (Bowman *et al.*, 2017; Morris *et al.*, 2018). *Marchantia polymorpha* (hereafter Marchantia) is a liverwort that re-emerged as a very powerful model to study the function and evolution of fundamental mechanisms common to land plants including those underlying morphological adaptations to a terrestrial life (Bowman *et al.*, 2017). The dominant phase of Marchantia's life cycle is the gametophytic (haploid) one and it is characterized by a dorsiventral thallus that can produce asexual propagules (gemmae) inside specialized gemma cups. The haploid gametophyte then generates male (antheridiophore) or female (archegoniophore) sexual reproductive structures. Motile sperm, formed in the antheridia, fertilize eggs contained in the archegonia resulting in the formation of haploid spores, and as sporelings develop into a thallus, the cycle is completed (Fig. 2).

While the core components of the small RNA regulatory pathways characterized in vascular plants are present in the Marchantia genome (Huang *et al.*, 2015; Bowman *et al.*, 2017) and across the green plant clade (Table 1) (Ma *et al.*, 2017), there are interesting peculiarities worth highlighting. In addition to the largest subunits of POLIV (MpNRPD1) and POLV (MpNRPE1a) (Box 1), Marchantia has an additional copy (MpNRPE1b) of the large subunit of POLV that in contrast to MpNRPE1a it contains a reduced version of the so-called WG/GW platform or AGO-hook domain, a region formed by repetitions of the amino acids tryptophan (W) and glycine (G) located at the CTD of POLV that is responsible for mediating the physical interaction with AGO proteins. This composition seems to be a feature shared by bryophytes as the moss *Physcomitrella patens* also contains two copies of POLV; however, their function is not clear, as mutations in *PpPOLV* that harbors an AGO-hook domain only affect the expression of a handful of loci and it has not been possible to recover mutants in the second copy (Coruh *et al.*, 2015). The DICER family in Marchantia consists of at least four members grouped into Mp*DCL1a*, Mp*DCL1b*, Mp*DCL3* and Mp*DCL4* (Huang *et al.*, 2015; Bowman *et al.*, 2017). There are no members of the DCL2 family. There is a single member of the Mp*HEN1* family in Marchantia and its expression is highly enriched in antherozoids or sperm (Fig. 2(9)). By contrast to observations in Arabidopsis and most land plants, the AGO family in Marchantia contains representatives of only two classical groups with a single member in the AGO1 clade and four members in the AGO4/6/9 clade (Bowman *et al.*, 2017). Additionally, there are two AGO-LIKE members named MpPIWIa and MpPIWIb that are missing the PAZ domain but with a distinctive PIWI domain and show an increased expression

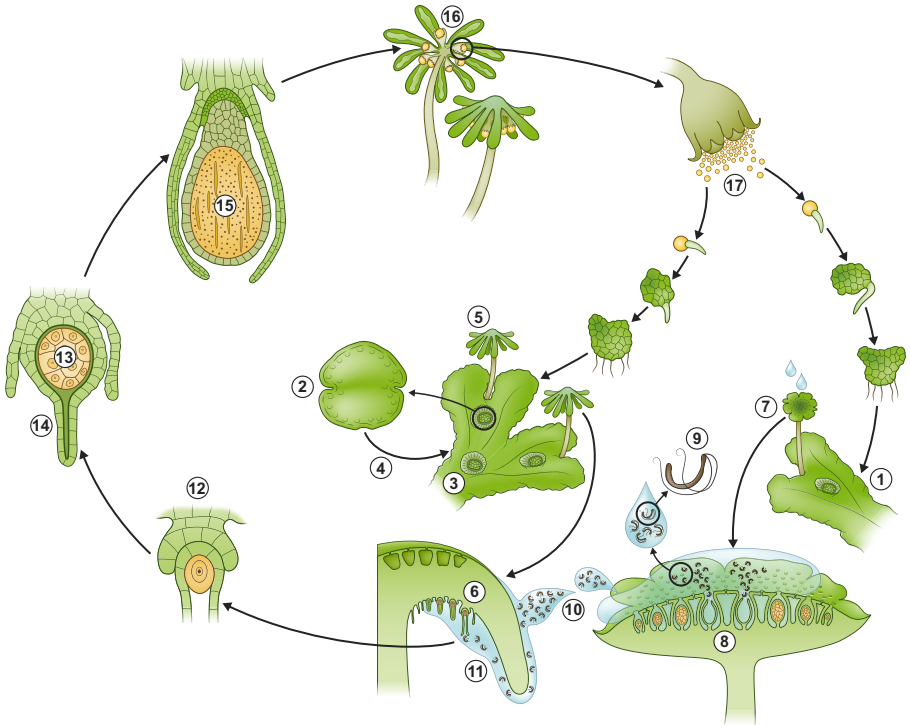


Fig. 2 Life cycle of *Marchantia polymorpha* (Marchantia). Marchantia is a thalloid liverwort that can reproduce asexually and sexually. Marchantia is a dioecious plant (male and female reproductive organs develop in distinct thalli). Sex is determined by the presence of a sexual chromosome Y in the case of males and X in the case of females. The haploid dominant phase is mainly constituted by a sheet-like structure named thallus (1). Asexual propagules known as gemmae (singular: gemma) (2) are produced inside a specialized structure called the gemma cup (3) that develops on the dorsal side of the thallus. Multicellular gemmae develop into a thallus, in this way completing the asexual cycle (4). Gemmae contain meristematic apical stem cells (also known as apical notches) located in the notches at the extremities of the thallus that direct growth of the thallus in a bifurcated pattern characteristic of this species. When mature thalli are exposed to a particular combination of far-red and white light, archegoniophores (5) harboring archegonia (6) (singular: archegonium) and antheridiophores (7) harboring antheridia (8) (singular: antheridium) are formed directly from the apical notches of the thallus. Motile sperm (9) produced in the antheridia take advantage of drops of water (10) (rain being the most common water source) to swim towards the archegonia (11) where they fertilize the egg cells (12) within. The embryo (13) develops inside a calyptra (14) that protects the growing sporophyte and after meiosis will give rise to spores (15). Thousands of spores contained in the sporangia (16) are released (17) and develop into young sporelings that will grow into a thallus, thereby completing the sexual cycle.

in the antheridiophores (Fig. 2(7)) (Schmid *et al.*, 2018). Marchantia contains two *de novo* DNA methyltransferases, two CMT-like methyltransferases (that do not reside in a clade with CMT2 or CMT3 clades) and two members of the ROS1 DNA glycosylase family. Interestingly, the expression pattern of several genes involved in DNA methylation is highly enriched in reproductive tissue (Fig. 2(5), (7)) and sporophytes (Fig. 2(14), (15)). For example, expression of Mp*DRM1b* is increased in archegonia (Fig. 2(6)) and sperm (Fig. 2(9)). Mp*MET1*, Mp*CMTa*, Mp*PIWIa* and Mp*PIWIb* are highly expressed in antheridiophores (Fig. 2(5)). The expression of Mp*DRMa* and Mp*CMTa* is increased in antheridiophores and sperm (Fig. 2(7), (9)). Mp*CMTb* is highly enriched in young sporophytes (Fig. 2(13)) (Bowman *et al.*, 2017; Schmid *et al.*, 2018). This opens the possibility of diversified RdDM modules acting with temporal or tissue specificity. Marchantia is mostly deprived of *sensu stricto* gbM (as defined above) during the gametophytic phase of its life cycle (Bewick *et al.*, 2017).

Recent genome-wide analyses of DNA methylation patterns in tissues ranging from thalli (Fig. 2(1)) to reproductive tissues including sperm cells, archegonia, vegetative meristems and different stages of sporophyte development have shown that extensive reprogramming occurs during the life cycle of Marchantia (Fig. 2); there is a marked gain of DNA methylation in archegonia, sperm cells and during sporophyte development, with the lowest levels of DNA methylation observed in vegetative gametophytic tissues and the highest levels present in sperm in all sequence contexts including not only gene bodies but also gene flanking regions (Schmid *et al.*, 2018). Similar to observations in Arabidopsis, mutations in Mp*MET1* result in a marked decrease of DNA methylation in the CG context in TEs although in Marchantia it is accompanied by a significant increase of non-CG methylation, suggesting that the increase in non-CG methylation serves to alleviate some of the CG loss (Ikeda *et al.*, 2018). The Mp*met1* mutant exhibits dramatic alterations during gametophyte development that include the formation of undifferentiated tissue that is not able to form reproductive structures (Ikeda *et al.*, 2018). The characterization of mutants affected in the core components of RdDM and other small RNA-based regulatory pathways will shed light on the evolution and the common and distinctive functions of these ancestral regulatory modules relative to other land plants. In addition, the availability of mutants with altered gene silencing mechanisms will be a great asset for ongoing efforts aiming to

Table 1 Conservation and diversification of DNA methylation pathways in selected species of the green plant clade.

		Core components of DNA methylation					
Taxa	Species	RDR	DCL	AGO/PIWI	DRM	CMT	MET
Chlorophyta	*Chlamydomonas reinhardtii*	0	3	3	3	0	3
	Volvox carteri	0	1	2	0	0	1
	Cholerella sorokiniana	0	1	1	0	1	1
	Chlorella variabilis	0	1	1	0	1	1
	Coccomyxa subellipsoidea	1	1	2	0	0	0
	Batrhycoccus prasinos	0	0	0	0	0	0
	Ostreococcus lucimarinus	0	0	0	0	0	0
Streptophyta	*Chara braunii*	6	2	5	2	1	0
	Klebsormidium flaccidum	4	2	6	2	1	1
	Marchantia polymorpha	4	4	6	2	2	1
	Physcomitrella patens	2	4	7	2	1	1
	Arabidopsis thaliana	6	4	10	2	3	4

The presence of genes involved in DNA methylation across representative species with sequenced genomes from Chlorophyta and Streptohpyta (adpated from Ma *et al.*, 2017). RNA DEPENDENT RNA POLYMERASE (RDR), DICER-LIKE (DCL). ARGONAUTE/P-element Induced Wimpy testis (AGO/PIWI), DOMAINS REARRANGED METHYLTRANSFERASE (DRM), CHROMOMETHYLASE (CMT), METHYLTRANSFERASE (MET). Chlorophytes exhibit the greatest variation, ranging from representatives lacking clear homologs for any of the gene families studied to species with a nearly a full set (i.e. *Chlamydomonas reinhardtii*). Species in the streptophyta clade show a more consistent distribution with just a single species not containing at least one representative of the gene families analyzed. The *Physcomitrella patens* genome was the first bryophyte genome to be sequenced more than a decade ago and it has been intensively studied for its nonconventional (relative to angiosperms) DNA methylation toolkit that includes four *DCL* genes: *PpDCL1a* is a member of the *DCL1* clade originally proposed to contain two members with *PpDCL1b* having a function in DNA methylation although recent evidence suggests strongly that *PpDCL1b* might be a spliced pseudogene; and single members of *DCL3* and *DCL4* clades and a *MINIMAL DICER-LIKE* (*mDCL*) involved in the production of 23 nucleotide small interfering RNAs (siRNAs) (reviewed by Coruh *et al.*, 2015).

use Marchantia as a natural chassis for bioengineering and synthetic biology.

VI. Conclusions and future perspectives

DNA methylation is essential for plant survival and it plays a fundamental role during the formation of seeds. In recent decades, we have gained knowledge on the function and the molecular mechanisms controlling the establishment, perpetuation and removal of DNA methylation in plants mainly from vascular model systems. *Marchantia polymorpha* is a powerful model with idiosyncratic features, including sexual reproduction, asexual natural clonal propagation and a full set of molecular tools for functional genomics that include genetic transformation, homologous recombination and CRISPR-Cas9-mediated gene editing. This will greatly enhance the repertoire of model systems positioned at basal evolutionary branches. Most interestingly, given the well-defined events of epigenetic reprogramming occurring during its life cycle, Marchantia appears to be an excellent model to study epigenetic regulation of gene expression and the molecular basis of transgenerational epigenetic inheritance.

Acknowledgements

We thank Luis Alfredo Cruz-Ramírez and Mario Enrique Zurita Ortega for their critical reading of the manuscript. We are grateful to Mario A. Arteaga-Sánchez for his help in the generation of illustrations. This work was funded by The Royal Society through a Newton Fund Advanced Fellowship (NA150181), the Mexican National Council on Science and Technology (CONACYT) (CB-158550) and the Agropolis Fondation (1502-307). AAC is the recipient of a scholarship (629316) from CONACYT.

ORCID

Mario Alberto Arteaga-Vázquez https://orcid.org/0000-0002-9435-0049
Daniel Grimanelli https://orcid.org/0000-0001-5424-114X
Jim Haseloff https://orcid.org/0000-0003-4793-8058

References

Bewick AJ, Niederhuth CE, Ji L, Rohr NA, Griffin PT, Leebens-Mack J, Schmitz RJ. 2017. The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. *Genome Biology* **18**: 65.

Bewick AJ, Schmitz RJ. 2017. Gene body DNA methylation in plants. *Current Opinion in Plant Biology* **36**: 103–110.

Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K, Yamaoka S, Nishihama R, Nakamura Y, Berger F *et al.* 2017. Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* **171**: 287–304.e15.

Cao X, Jacobsen SE. 2002. Locus-specific control of asymmetric and CpNpG methylation by the *DRM* and *CMT3* methyltransferase genes. *Proceedings of the National Academy of Sciences, USA* **99**: 16491–16498.

Coruh C, Cho SH, Shahid S, Liu Q, Wierzbicki A, Axtell MJ. 2015. Comprehensive annotation of *Physcomitrella patens* small RNA loci reveals that the heterochromatic short interfering RNA pathway is largely conserved in land plants. *Plant Cell* **27**: 2148–2162.

Cuerda-Gil D, Slotkin RK. 2016. Non-canonical RNA-directed DNA methylation. *Nature Plants* **2**: 16163.

Gao Z, Liu H-L, Daxinger L, Pontes O, He X, Qian W, Lin H, Xie M, Lorkovic ZJ, Zhang S *et al.* 2010. An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature* **465**: 106–109.

Gouil Q, Baulcombe DC. 2016. DNA methylation signatures of the plant chromomethyltransferases. *PLoS Genetics* **12**: e1006526.

Haag JR, Pikaard CS. 2011. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nature Reviews Molecular Cell Biology* **12**: 483–492.

Havecker ER, Wallbridge LM, Hardcastle TJ, Bush MS, Kelly KA, Dunn RM, Schwach F, Doonan JH, Baulcombe DC. 2010. The *Arabidopsis* RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell* **22**: 321–334.

Huang Y, Kendall T, Forsythe ES, Dorantes-Acosta A, Li S, Caballero-Pérez J, Chen X, Arteaga-Vázquez M, Beilstein MA, Mosher RA. 2015. Ancient origin and recent innovations of RNA polymerase IV and V. *Molecular Biology and Evolution* **32**: 1788–1799.

Ikeda Y, Nishihama R, Yamaoka S, Arteaga-Vazquez MA, Aguilar-Cruz A, Grimanelli D, Pogorelcnik R, Martienssen RA, Yamato KT, Kohchi T *et al.* 2018. Loss of CG methylation in *Marchantia polymorpha* causes disorganization of cell division and reveals unique DNA methylation regulatory mechanisms of non-CG methylation. *Plant and Cell Physiology* **59**: 2421–2431.

Jackson JP, Lindroth AM, Cao X, Jacobsen SE. 2002. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**: 556–560.

Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, Zhong X, Marson G, Pellegrini M, Segal DJ *et al.* 2014. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* **507**: 124–128.

Lanciano S, Mirouze M. 2018. Transposable elements: all mobile, all different, some stress responsive, some adaptive? *Current Opinion in Genetics & Development* **49**: 106–114.

Law JA, Ausin I, Johnson LM, Vashisht AA, Zhu J-K, Wohlschlegel JA, Jacobsen SE. 2010. A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in arabidopsis. *Current Biology* **20**: 951–956.

Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, Strahl BD, Patel DJ, Jacobsen SE. 2013. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498**: 385–389.

Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **11**: 204–220.

Li X, Harris CJ, Zhong Z, Chen W, Liu R, Jia B, Wang Z, Li S, Jacobsen SE, Du J. 2018. Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation. *Proceedings of the National Academy of Sciences, USA* **115**: E8793–E8802.

Ma X, Kim E, Cerutti H. 2017. Gene silencing in archaeplastida algae. In: Dalmay T, ed. *Plant gene silencing: mechanisms and applications*. Wallingford, UK: CABI, 75–93.

Matzke MA, Mosher RA. 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics* **15**: 394–408.

McClintock B. 1956. Controlling elements and the gene. *Cold Spring Harbor Symposia on Quantitative Biology* **21**: 197–216.

Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. 2018. The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences, USA* **115**: E2274–E2283.

Nuthikattu S, McCue AD, Panda K, Fultz D, DeFraia C, Thomas EN, Slotkin RK. 2013. The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiology* **162**: 116–131.

Ortega-Galisteo AP, Morales-Ruiz T, Ariza RR, Roldán-Arjona T. 2008. Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. *Plant Molecular Biology* **67**: 671–681.

Pontier D, Picart C, Roudier F, Garcia D, Lahmy S, Azevedo J, Alart E, Laudié M, Karlowski WM, Cooke R *et al.* 2012. NERD, a plant-specific GW protein, defines an additional RNAi-dependent chromatin-based pathway in *Arabidopsis*. *Molecular Cell* **48**: 121–132.

Saze H, Shiraishi A, Miura A, Kakutani T. 2008. Control of genic DNA methylation by a jmjC domain-containing protein in *Arabidopsis thaliana*. *Science* **319**: 462–465.

Schmid MW, Giraldo-Fonseca A, Rövekamp M, Smetanin D, Bowman JL, Grossniklaus U. 2018. Extensive epigenetic reprogramming during the life cycle of *Marchantia polymorpha*. *Genome Biology* **19**: 9.

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE. 2014. Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nature Structural & Molecular Biology* **21**: 64–72.

Trujillo JT, Seetharam AS, Hufford MB, Beilstein MA, Mosher RA. 2018. Evidence for a unique DNA-dependent RNA polymerase in cereal crops. *Molecular Biology and Evolution* **35**: 2454–2462.

Vongs A, Kakutani T, Martienssen RA, Richards EJ. 1993. *Arabidopsis thaliana* DNA methylation mutants. *Science* **260**: 1926–1928.

Wollmann H, Stroud H, Yelagandula R, Tarutani Y, Jiang D, Jing L, Jamge B, Takeuchi H, Holec S, Nie X *et al.* 2017. The histone H3 variant H3.3 regulates gene body DNA methylation in *Arabidopsis thaliana*. *Genome Biology* **18**: 94.

Woo HR, Dittmer TA, Richards EJ. 2008. Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in *Arabidopsis*. *PLoS Genetics* **4**: e1000156.

Woo HR, Pontes O, Pikaard CS, Richards EJ. 2007. VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. *Genes & Development* **21**: 267–277.

Yang D-L, Zhang G, Tang K, Li J, Yang L, Huang H, Zhang H, Zhu J-K. 2016. Dicer-independent RNA-directed DNA methylation in *Arabidopsis*. *Cell Research* **26**: 66–82.

Yang D-L, Zhang G, Wang L, Li J, Xu D, Di C, Tang K, Yang L, Zeng L, Miki D *et al.* 2018. Four putative SWI2/SNF2 chromatin remodelers have dual roles in regulating DNA methylation in *Arabidopsis*. *Cell Discovery* **4**: 55.

Zhang H, Lang Z, Zhu J-K. 2018. Dynamics and function of DNA methylation in plants. *Nature Reviews Molecular Cell Biology* **19**: 489.

Zhang H, Zhu J-K. 2012. Active DNA demethylation in plants and animals. *Cold Spring Harbor Symposia on Quantitative Biology* **77**: 161–173.

Zhou M, Law JA. 2015. RNA Pol IV and V in gene silencing: rebel polymerases evolving away from Pol II's rules. *Current Opinion in Plant Biology* **27**: 154–164.

Zhou M, Palanca AMS, Law JA. 2018. Locus-specific control of the *de novo* DNA methylation pathway in *Arabidopsis* by the CLASSY family. *Nature Genetics* **50**: 865.

Zilberman D, Cao X, Jacobsen SE. 2003. *ARGONAUTE4* control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**: 716–719.