Small RNAs guide histone methylation in Arabidopsis embryos

Jean-Sébastien Parent,1,2 Jonathan Cahn,1 Rowan P. Herridge,1,5 Daniel Grimanelli,3,4 and Robert A. Martienssen1

1Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA; 2Agriculture and Agri-Food Canada, Ottawa Research and Development Center, Ottawa Ontario K1A 0C6, Canada; 3Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA; 4Epigenetic Regulations and Seed Development, UMR232, Institut de Recherche pour le Développement (IRD), Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Université de Montpellier (UM), 34394 Montpellier, France

Epigenetic reprogramming occurs during gametogenesis as well as during embryogenesis to reset the genome for early development. In flowering plants, many heterochromatic marks are maintained in sperm, but asymmetric DNA methylation is mostly lost. Asymmetric DNA methylation is dependent on small RNA but the re-establishment of silencing in embryo is not well understood. Here we demonstrate that small RNAs direct the histone H3 lysine 9 dimethylation during Arabidopsis thaliana embryonic development, together with asymmetric DNA methylation. This de novo silencing mechanism depends on the catalytic domain of SUVH9, a Su(Var)3-9 homolog thought to be catalytically inactive.

Supplemental material is available for this article.

Received August 19, 2020; revised version accepted April 16, 2021.

Constitutive heterochromatin in many eukaryotes is characterized by the presence of methyl groups on the cytosine bases of DNA [mC] and by di- or trimethylation of lysine 9 of histone H3 [H3K9me2/3] (Du et al. 2015). Maintenance of these marks through cell divisions is essential to keep transposable elements [TEs] silenced and ensure genome stability [Allshire and Madhani 2018]. In plants, DNA methylation is found in three functionally distinct sequence contexts, namely, symmetric CG, CHG, and asymmetric CHH [where H is either A, C, or T], with the non-CG contexts tightly associated with H3K9me2 in heterochromatin [Stroud et al. 2014]. In Arabidopsis thaliana, this association relies on the activity of three H3K9 methyltransferases [KMTs], SU(VAR)3-9 HOMOLOG4 (SUVH4), SUVH5, and SUVH6 that bind methylated cytosines in different contexts through their SRA domain [Du et al. 2015]. The methylated histones are then recognized by the DNA methyltransferases [DMTs] CHROMOMETHYLASE2 (CMT2) and CMT3 that perpetuate CHH and CHG methylation, respectively [Stroud et al. 2014]. The interdependence between the two marks is thought to be responsible for the propagation of H3K9me2 through cell generations.

During the flowering plant male germline development, CG and CHG methylation are retained but CHH methylation is lost from the paternal genome, first during meiosis and then through the microspore and mature sperm cells [Calarco et al. 2012; Ibara et al. 2012; Ingouff et al. 2017; Walker et al. 2017]. Following fertilization, this methylation must be restored de novo and has been proposed to be guided by small RNA [sRNA] molecules [Calarco et al. 2012; Ibara et al. 2012]. In Arabidopsis, these sRNA molecules are 24 nt long and produced by RNA POLYMERASE IV [POLIV] [Matzke et al. 2015]. They are loaded onto a specific clade of ARGONAUTE proteins [mainly AGO4, AGO6, and AGO9], which can then detect homology in noncoding transcripts produced by RNA POLYMERASE V [POLV]. POLV is recruited to methylated DNA by SUVH2 and SUVH9, the latter having specific affinity for CHH methylation bestowed by its SRA domain [Johnson et al. 2014; Liu et al. 2014; Ingouff et al. 2017]. The recognition of POLV transcripts by AGO proteins ultimately recruits DOMAIN REARRANGED METHYLTRANSFERASE1 [DRM1] and DRM2 proteins that methylate cytosines in all three sequence contexts, a process known as RNA-directed DNA methylation [RdDM] [Matzke et al. 2015].

However, the current model for RdDM implies pre-existing DNA methylation and so cannot account for purely de novo silencing. One possible explanation is that sRNAs targeting TEs instead guide KMTs as they do in fission yeast, Caenorhabditis elegans, and Drosophila [Castel and Martienssen 2013], all of which lack these DMTs. H3K9me2 could then recruit POLIV via the SAWADEE HOMEODOMAIN HOMOLOG1 [SHH1] [Law et al. 2011] and maintain silencing through RdDM. Here we investigated the genetic dependencies of methylated DNA and histone H3K9 in mature embryos of Arabidopsis thaliana. We observe that H3K9 methylation at small RNA targets is independent of the maintenance KMTs SUHVH4/5/6 and does not correlate perfectly with non-CG methylation at the same loci. We propose a role for a new family of KMT that is directly guided by small RNA molecules for de novo silencing.

Results and Discussion

sRNA molecules guide CHH and CHG DNA methylation to specific targets in embryos

It was recently shown that RdDM activity peaks during embryogenesis in Arabidopsis as well as in soybeans and...
chickpeas (An et al. 2017; Bouyer et al. 2017; Kawakatsu et al. 2017; Rajkumar et al. 2020), which is otherwise consistent with a role in epigenetic reprogramming. We therefore examined both the repressive mC and H3K9me2 modifications in Arabidopsis embryos. First, whole-genome bisulfite sequencing (WGBS) was performed on dissected mature embryos. We tested an ago4 ago6 ago9 triple mutant, given that the three proteins are present in the cell nucleus of different embryonic tissues throughout development (Supplemental Fig. S1) and all are required for 24-nt sRNA activity (Liu et al. 2018). We also tested a polIV mutant devoid of 24-nt sRNAs, as well as drm1 drm2, and suvh2 suvh9 double mutants to compare them with the suvh4 suvh5 suvh6 triple mutant with impaired H3K9me2 maintenance (Stroud et al. 2014). Methylation ratios were separated by context, and CHH was used to call differentially methylated regions (DMRs).

We observed that sRNA-dependent DMRs in embryos (defined as hypomethylated in ago4 ago6 ago9) mapped mostly to DNA transposons (Fig. 1A) and were largely excluded from sRNA-independent DMRs (defined as hypomethylated in suvh4 suvh5 suvh6), comprising mostly Copia and Gypsy elements (retrotransposons). As in vegetative tissues (Stroud et al. 2013), CHG methylation at sRNA targets is reduced in mutant embryos but not entirely lost (Fig. 1B). CHH methylation is also exclusive to each target group with suvh4 suvh5 suvh6 mutants being unaffected at sRNA targets while losing all CHH methylation at sRNA-independent targets. Finally, the two pathways are able to repress their respective targets without influencing one another, as exemplified by neighboring TEs at many different locations in the genome (Fig. 1C).

sRNA molecules target H3K9me2 independently of SUVH4, SUVH5, and SUVH6 in embryos

Next, we performed chromatin immunoprecipitation followed by sequencing (ChIP-seq) on dissected mature embryos. The genotypes used in WGBS experiments were again tested in duplicates to ascertain reproducibility (Supplemental Fig. S2). We used previously identified DMRs to map H3K9me2 levels and showed that RdDM mutants have lower H3K9me2 levels at sRNA-dependent regions but wild-type levels at sRNA-independent regions (Fig. 2A). We noted that the drm1 drm2 mutant appeared less affected, suggesting sRNA-dependent H3K9me2 did not fully depend on mC in embryos (Fig. 2A, top panel, Supplemental Fig. S3). In contrast, suvh4 suvh5 suvh6 H3K9me2 levels were not decreased compared with wild type at sRNA-dependent regions (Fig. 2A, top panel). This observation argues that neither SUVH4, SUVH5, nor SUVH6 is responsible for depositing H3K9me2 at these loci.

This prompted us to use immunofluorescence to test whether the global levels of H3K9me2 were perceptibly changed in these mutants during embryo development, something not previously reported. We identified a developmental window between the zygote and the 64-cell stage where H3K9me2 immunofluorescence at heterochromatic foci is sharply reduced in ago4 ago6 ago9 and suvh2 suvh9 mutants (Fig. 2B; Supplemental Fig. S4), as it is in suvh4 suvh5 suvh6 mutants at these same embryonic stages. This was unexpected because, as small RNA-dependent regions are scattered throughout the genome, loss of RdDM does not lead to loss of immunofluorescent signal in somatic cells [Johnson et al. 2008; Jing et al. 2016]. However, ago4 ago6 ago9 and suvh2 suvh9 mutants recover H3K9me2 signal at later stages unlike suvh4 suvh5 suvh6 (Fig. 2C; Supplemental Table S1), in agreement with our ChIP-seq results that show massive loss of H3K9me2 signal in the pericentromeric regions only in the latter.

**SUUVH9 catalytic site is required for epigenetic reprogramming**

Our results therefore pointed to the existence of one or more KMTs acting in concert with sRNA molecules in embryos to impose a silencing mark. We rationalized that SUVH2 and SUUVH9 were the most likely candidates and, like G9a in human, might
both recruit DMTs and deposit methyl groups on lysine 9 (Dong et al. 2008). However, these two proteins are not thought to be active KMTs due mainly to their lack of a post-SET domain (Johnson et al. 2008, 2014). Despite this short truncation, they do have a SET domain with the catalytic tyrosine residue conserved, and the suvh2 suvh9 mutant has sharply reduced H3K9me2 foci in early embryos with fluorescent signal in the case of SUVH2, but only the WT form was able to restore non-CG methylation to the SDC locus (Supplemental Table S2), further demonstrating the importance of SUVH9 catalytic activity for epigenetic reprogramming.

The SUVH9 family appeared early in the evolution of flowering plants, as suggested by the presence of a single gene in the basal angiosperm Amborella trichopoda (Li et al. 2015). This copy already displays the truncation of the C-terminal post-SET domain that is essential for completing the catalytic site of other KMTs (Johnson et al. 2014). This supported the notion that SUVH9 lost catalytic function early in its evolution and served only as an adaptor between DNA methylation and POLV (Johnson et al. 2008; Liu et al. 2014). However, all the SUVH9 orthologs we have examined possess the catalytic tyrosine as well as many other structurally important residues of the SET domain. Such a level of sequence conservation in an inactive SET domain seems unlikely, and indeed, mutating the catalytic residue abolished SUVH9 developmental phenotypes (Johnson et al. 2008). Since the activated SDC allele requires functional de novo silencing activity to restore silencing (Henderson and Jacobsen 2008), crossing suvh2 suvh4 suvh9 to suvh2 suvh9 results in 100% F1 plants with the SDC phenotype, while crossing suvh2 suvh4 suvh9 to WT yields normal progeny (Fig. 3B). When the triple mutant was crossed with two different lines of suvh2 suvh9 complemented with the wild-type transgene, the offspring showed a wild-type phenotype, confirming the rescue of SUVH9 function. However, when the catalytic tyrosine Y636 is mutated, the rescue is abolished in two independent lines (Fig. 3B; Supplemental Table S2). This suggests that catalytic activity is essential for the reprogramming of the SDC allele by the RdDM pathway.

We then performed ChIP-seq and WGBS on developing seeds at the mature embryo stage from suvh2 suvh9 mutants complemented by mutated and wild-type SUVH9. Using MACS2 (Zhang et al. 2008), we identified 203 regions with H3K9me2 peaks in seeds complemented with SUVH9 (WT) that did not show enrichment in the control (Fig. 3C). The majority of regions identified overlapped with DNA transposons of the MuDR (44%) and Helitron (19%) families, in line with the target preference of this silencing pathway. H3K9me2 levels at these loci were also increased in plants complemented with the Y636F version but to a significantly lower extent (Fig. 3C). Similarly, DNA methylation in the CHH context at the same loci is significantly increased in both complemented lines but to a lower extent with the mutant version (Fig. 3D). This suggests that the catalytic residue also has a direct or indirect influence on the ability to recruit DMTs. Importantly both WT and catalytic mutants were also able to restore non-CG methylation to the SDC locus (Supplemental Fig. S6), but only the WT form was able to restore silencing (Fig. 3B; Supplemental Table S2), further demonstrating the importance of SUVH9 catalytic activity for epigenetic reprogramming.

The SDC gene that is activated by complete loss of epigenetic silencing in the suvh2 suvh4 suvh9 mutant, causing
H3K9me2 ratios \[\log_2(\text{H3K9me2}/\text{H3})\] were calculated over the genomic regions identified by – es between distributions were calculated using a Kolmogorov were calculated over the genomic regions identified by the MACS2 broad peak analysis. Distanc- ent contexts are identified with a need for epigenetic reprogram- ming in early embryogenesis where asymmetric DNA methylation levels are low. Pericentromeric regions in suvh2 suvh9 mutants are nonetheless able to recover normal H3K9me2 levels later in development (Fig. 2C, bottom panel), presumably through the persistence of symmetric DNA methylation and the activity of SUVH4/5/6 (To et al. 2020).

Flowering plants have therefore evolved a specialized clade of SuVar3-9 homologs, which we show are associated with small RNA-dependent transcriptional silencing in the develop- ing seed. Interestingly, this pathway appeared in angiosperms at the same time as the appearance of dosage im- balance and imprinting in the endo- sperm, due to double fertilization [Wang and Köhler 2017]. Remarkably, both SUVH9 and H3K9me2 are important for imprinting and chromosome dosage response in the endosperm [Ji- ang et al. 2017]. Indeed, the sRNA-de- pendent DMRs we have identified in the embryo also gain H3K9me2 in the unbalanced 4n endosperm [Supple- mental Fig. S7], implicating sRNA-de- pendent H3K9me2 in triploid seed abortion. Our study therefore links the emergence of SUVH9 homologs to the appearance of interploidy hy- bridization barriers in endosperm development, an impor- tant driver of plant speciation. It has previously been speculated that the small RNA silencing pathway could be a facilitator of genome duplications that are frequent in the history of flowering plants [Matzke et al. 2015]. A mechanism that acts to control both TE expansion during genome duplication and the establishment of interploidy barriers would undoubtedly have played a crucial role in Darwin’s “abominable” diversification of angiosperms (Cibrian-Jaramillo and Martienssen 2009).

function (Fig. 3B,C). The crystal structure of SUVH9 has revealed that the substrate binding pocket is too widely opened and incomplete, explaining the lack of in vitro ac- tivity in recombinant protein [Johnson et al. 2008, 2014]. This is reminiscent of other well-known KMTs, namely, CURLY LEAF (CLF) and MEDEA (MEA), two E[z] homologs in plants. E[z] homologs do not possess a cysteine- rich post-SET domain and yet are known to catalyze the/CURLY LEAF (CLF) and MEDEA (MEA), two E(z) homo- logs in plants. E(z) homologs do not possess a cysteine-rich post-SET domain and yet are known to catalyze the activity in recombinant protein (Johnson et al. 2008, 2014). Like their animal counterparts, these proteins need additional subunits to be active and are therefore pu- rified as complexes for in vitro assays (Schmitges et al. 2014). Indeed, the sRNA-de- pendent H3K9me2 in endosperm of early seed is not dependent on SUVH2/9 but only SUVH4/5/6 [Fig. 2B,C; Supple- mental Fig. S4], similar to what is observed in leaves. This is consistent with a need for epigenetic reprogram- ming in early embryogenesis where asymmetric DNA methylation levels are low. Pericentromeric regions in suvh2 suvh9 mutants are nonetheless able to recover normal H3K9me2 levels later in development [Fig. 2C, bottom panel], presumably through the persistence of symmetric DNA methylation and the activity of SUVH4/5/6 (To et al. 2020).

Materials and methods

**Plant material and growth conditions**

Plants were grown under long-day conditions at 22°C. Seeds were sown directly on soil after stratification in distilled water for 3 d at 4°C. For de- veloping seed isolation, the mature fruits were collected from the primary inflorescence after 7–8 wk of growth and cut open, and fresh seeds were harvested and frozen. For embryo isolation, siliques were cut open with in- sulin syringe needles; whole seeds were transferred on microscope slide in a drop of PBS; and embryos were squeezed out, applying pressure with a microslide. Embryos were then manually separated from the other tissues and washed three times in PBS before freezing in liquid nitrogen. We com- bined the alleles ago4-5, ago6-2, and ago9-2 (used in Stroud et al. 2013) to generate the triple ago mutant. trypd1-a3 was described before [Herr et al. 2005]. drm1 drm2 (Stroud et al. 2013), suvh2 suvh9 [Johnson et al. 2008],
**Bisulfite sequencing and DNA methylation analysis**

DNA was isolated from 2000–3000 mature embryos using the Quick-DNA microprep Plus kit (Zymo Research), according to the manufacturer’s instructions. DNA (150 ng) was sheared using a Bioruptor to sizes ranging from 200–600 bp. The fragments were ligated with NEBNext methylated adaptors [New England Biolabs] and AMPure bead-purified (Beckman Coulter) before being treated with the EZ DNA Methylatation-Lightning kit [Zymo Research], according to the manufacturer’s instructions. Converted DNA was eluted with 20 μL of buffer, and 10 μL was used for amplification with uracil-tolerant polymerase [KAPA Biosystems] with NEBNext index containing primers.

For developing seeds, DNA was isolated from 250–500 young seeds at the mature embryo stage using the plant/seed DNA minipreparation kit [Zymo Research]. Five nanograms of DNA was converted and processed into sequencing libraries with the Zymo Pico Methyl-seq library preparation kit, according to the manufacturer’s instructions. A single sample was made and analyzed for each genotype.

Libraries from each genotype were pooled and sequenced on a NextSeq 500 platform [Illumina] with single-end 75 nt reads at high output. Adaptor sequences were removed using Trimomatic [Bolger et al. 2014], trimmed reads were then aligned on the Arabidopsis TAIR10 genome using Bismark [Krueger and Andrews 2011]. Alignment files were then analyzed with the Methpipe programs [Song et al. 2013] for quality assessment, and biological replicates (embryos) were compared to assess reproducibility. The replicates were then merged to give more statistical power to the DMR analysis. In short, only cytosines covered more than five times were considered; the cytosines were separated in three contexts (CG, CHG, and CHH), and the CHH methylation was used to define hypomethylated DMRs in mutant backgrounds. The algorithm [Song et al. 2013] used a hidden Markov model to identify methylated regions in each data set; these regions were then compared between data sets [mutants compared with WT] to identify regions with a minimum of 10 methylated cytosines, at least five of which were hypomethylated in the mutant. DMRs hypomethylated in agd6 agd8 agd9 compared with wild type were defined as small RNA dependent, and DMRs hypomethylated in svhd svh5 svh6 compared with wild type were defined as small RNA independent. Regions appearing in both sets were discarded; all distributions were compared with the WT distribution using the Kolmogorov-Smirnov test with R.

**ChIP-seq and analysis**

The detailed ChIP protocol is in the Supplemental Material. For sequencing, DNA fragments were made into libraries using the NEBNext Ultra II FS kit [New England Biolabs]. Libraries were pooled and sequenced on a NextSeq 500 platform [Illumina] with paired-end 150-nt reads at high output [developing seeds] or single-end 75-nt reads at high output [mature embryos]. Adaptor sequences were removed using Trimomatic [Bolger et al. 2014], trimmed reads were then aligned on the Arabidopsis TAIR10 genome using Bowtie2 [Langmead and Salzberg 2012]. Duplicated reads are removed using the Picard Tool suite, and coverage was calculated and mapped to DMRs using Bedtools. All distributions were compared with the first WT sample distribution using the Kolmogorov-Smirnov test with R.

**Immunofluorescence**

Cytological technique was essentially described elsewhere [Ingouff et al. 2017]. Young siliques of different stages were harvested and incubated in fixating solution for 5 min under vacuum and then for 2 h with rotation. Developing seeds were then dissected, embedded in acrylamide, digested, and incubated with antibodies anti-H3K9me2 [ab12230] and anti-H3K4me1 [Abcam ab8895] and secondary antibodies coupled with Alexa Fluor 488 and 564 [Thermo Fisher Scientific] together with DAPI and ProLong Gold antifade mountant [Thermo Fisher Scientific]. Imaging was performed using either a Zeiss LSM780 or LSM880 with Axyscan detector.

**Cloning**

SUVH2 and SUVH9 genomic sequences were amplified and then cloned into pDONR-221 and pDONR-207 [Thermo Fisher Scientific] using the BP enzyme mix according to the manufacturer’s instructions. The mCtrynite fragment was then added before the start codon using the HiFi DNA assembly master mix [New England Biolabs]. PCR mutagenesis was achieved, amplifying the entry clone with mutation-containing primers and the KOD Xtrae hot start DNA polymerase [MilliporeSigma]. Entry clones were transferred to the binary vector PMDC99 [Curtsis and Grossniklaus 2003]. Binary vectors were transformed into Agrobacterium tumefaciens strain GV3101 and used to transform svhd svh9 plants by floral dip (Clough and Bent 1998). All primers used are in Supplemental Table S3.

**Accession numbers**

The data sets generated in this study are available at Gene Expression Omnibus (GSE147245).

**Competing interest statement**

The authors declare no competing interests.

**Acknowledgments**

We thank E. Ernst and B. Berube for help with computational analysis, M. Ingouff for help with microscopy, S. Bhattacharjee, U. Ramu, and M. Callahan for technical help, and T. Mulligan for plant care. We thank F. Borges, B. Roche, and J.-F. Couture for fruitful discussions, as well as I. R. Henderson for critical reading of this manuscript. This work was supported by the Howard Hughes Medical Institute and grants from the National Institutes of Health [ROI GM070714] and the National Science Foundation Plant Genome Research Program [to R.A.M.]. D.G. was supported by grants from the Agence Nationale de la Recherche [REMETH; ANR-15-CE12-0012], Chromobore [ANR-18-CE92-0014], and Marie Sklodowska-Curie Actions [REP-658900-2]. We acknowledge assistance from the Cold Spring Harbor Laboratory Shared Resources, which are funded in part by the Cancer Center [Support Grant SP0030CA045508].

**Author contributions**


**References**


Parent et al.


Law JA, Vashishth AA, Wohlschlegel JA, Jacobsen SE. 2011. SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. *PLoS Genet* 7: e1002195. doi:10.1371/journal.pgen.1002195


Small RNAs guide histone methylation in *Arabidopsis* embryos

Jean-Sébastien Parent, Jonathan Cahn, Rowan P. Herridge, et al.

*Genes Dev.* published online May 20, 2021
Access the most recent version at doi:10.1101/gad.343871.120

**Supplemental Material**

http://genesdev.cshlp.org/content/suppl/2021/05/17/gad.343871.120.DC1

Published online May 20, 2021 in advance of the full issue.

**Creative Commons License**

This article, published in *Genes & Development*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

**Email Alerting Service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).