Arabidopsis epigenetics: when RNA meets chromatin
Anne-Valérie Gendrel and Vincent Colot

Recent work in plants and other eukaryotes has uncovered a major role for RNA interference in silent chromatin formation. The heritability of the silent state through multiple cell division cycles and, in some instances, through meiosis is assured by epigenetic marks. In plants, transposable elements and transgenes provide striking examples of the stable inheritance of repressed states, and are characterized by dense DNA methylation and heterochromatin histone modifications. Arabidopsis is a useful higher eukaryotes model with which to explore the crossroads between silent chromatin and RNA interference both during development and in the genome-wide control of repeat elements.

Address
Unité de Recherche en Génomique Végétale, INRA/CNRS/UEVE, 2 rue Gaston Crémieux, 91057 Evry Cedex, France

Corresponding author: Vincent Colot (colot@evry.inra.fr)

Introduction
Epigenetics is most commonly defined as the study of mitotically and/or meiotically heritable changes in gene expression that do not entail a change in DNA sequence [1], and these days it is usually equated with the study of chromatin inheritance [2]. Alongside DNA methylation, which is considered as a classic epigenetic mark, histone modifications and variants have become increasingly accepted as likely conveyors of epigenetic information [2]. The most recent, and versatile, partner of chromatin to have emerged across species over the past few years is RNA, which comes in many different guises, ranging from long non-coding RNAs, which are involved in X-chromosome inactivation and genomic imprinting in mammals, to short interfering RNAs (siRNAs), which originate, like miRNAs, from transcripts that bear little resemblance to the mRNAs targeted for degradation [10,11]. Moreover, several Arabidopsis miRNA genes are likely to have evolved from partial inverted duplications of target-protein-coding genes, which first produced dsRNAs that were processed into siRNAs, suggesting an
evolutionary continuum between these two types of small RNAs [12]. Finally, in addition to its function in targeting mRNAs for either degradation or translational repression, the RNAi machinery has been implicated in a third, chromatin-based silencing pathway in several organisms, including *S. pombe*, *Drosophila* and *Arabidopsis* [13]. Thus the emerging picture is one of an ancestral mechanism that has evolved through at least three functional divergences, the products of which can all be found in extant plants (Figure 1). As outlined below, the combination of *Arabidopsis* genomics with forward and reverse genetics has provided a powerful system for deciphering the links between the RNAi machinery and the modulation of chromatin at the levels of DNA methylation and histone changes in a complex organism.

**RNA-dependent DNA methylation**

The methylation of cytosine residues (C) is widespread among eukaryotes, in which it is thought to act primarily as a defense system against the transcriptional activity and mobility of repeated elements, as well as against recombination between such repeats [14]. Not all eukaryotes have DNA methylation, however, and there are differences between some organisms, such as plants and mammals, in the way in which methylation adorns their DNA. In mammals, methylated Cs are located almost exclusively within CpG dinucleotides, whereas in plants, Cs in the symmetric triplet CpNpG and in non-symmetric sequence contexts are also methylated, although less efficiently than Cs in CpGs [15]. Moreover, whereas DNA methylation in mammals affects both genes and
‘intergenic’ regions, genes are typically unmethylated in plants. This difference can be explained by the paucity of transposable elements and other repeats in plant introns compared to those of mammals, and by the fact that DNA methylation appears to be targeted specifically to repetitive elements in plants whereas it affects exons as well as repeats in mammals [16]. In turn, these differences might provide part of the explanation for the relatively benign phenotypic consequences of reduced DNA methylation observed in Arabidopsis compared to those seen in mammals [17–20].

The question of how genomes dictate the choice of sequences to be methylated has received much attention over the past 15 years, following the unexpected observation that transgenes in plants and fungi are often affected by DNA methylation and stable transcriptional silencing. At first, the discovery of two related duplication-dependent DNA methylation and silencing processes in the fungi Neurospora and Ascomolus led to a hypothesis in which DNA methylation is triggered by DNA–DNA pairing interactions between dispersed or tandem copies of the same sequence [15]. Soon afterwards, however, the observation in tobacco that a viroid (i.e. a non-encapsidated RNA pathogen) could direct the methylation of its cognate DNA sequence harbored on a transgene provided direct experimental support for an alternative, RNA-based mechanism [5]. This mechanism was subsequently shown to have an exquisite sequence specificity [21], and most of the evidence now points to RNA playing a central role in directing DNA methylation to specific regions of the genome in plants [22,23].

Various RNA sources, such as viroids, cytoplasmic viruses, transgenes or particular endogenous loci, can trigger RNA-dependent DNA methylation (RdDM) experimentally [22]. For instance, the introduction of a construct that enables the transcription of an inverted repeat containing the nopaline synthase promoter (NOS-pro) led to the methylation and transcriptional silencing in trans of an identical promoter sequence carried by a second transgene [6]. In this case, the production of dsRNA was associated with the appearance of siRNAs of 23–25 nt [6], siRNAs have not, however, been detected yet in the case of the RNA-mediated DNA methylation of the Arabidopsis paralogous PAI2 and PAI3 transproobsynthetic genes by the genetically unlinked PAI1–PAI4 inverted repeat that is present in some ecotypes, although intramolecular dsRNAs are clearly produced from the inversion [24*] and appear to be necessary for PAI2 and PAI3 methylation [25]. Whether the signal for directing DNA methylation is siRNAs or precursor dsRNAs cannot therefore be deduced from these observations. However, several other lines of evidence strongly favor the former possibility. First, re-methylation of the naturally methylated Arabidopsis gene FWA following its re-introduction into plants by Agrobacterium-mediated transformation was shown to depend on several RNAi genes, including the Dicer-like gene DCL3, the RNA-dependent RNA polymerase2 (RDR2) gene and ARGO-NAUTE (AGO4) [26*], and siRNAs have been detected that correspond to FWA promoter sequences in wildtype plants [27**]. Second, ago4, rdr2 and dcl3 mutant plants are also defective in siRNA accumulation and DNA methylation (as well as in histone H3 lysine 9 di-methylation [H3K9me2]; see next section) at several endogenous and transgenic loci [28–30*]. Third, two complexes have been biochemically characterized in the fission yeast S. pombe that contain Aргонаута and RdRP, respectively, and that physically associate with each other and with pericentromeric heterochromatin in a Dicer- and siRNAs-dependent manner [31**,32**]. Although S. pombe does not methylate its DNA, the RNAi-mediated heterochromatin assembly pathway that was first identified in this yeast is likely to be conserved to a great extent in plants [13,33].

A hallmark of RdDM is the methylation of cytosines at both symmetric and non-symmetric sites [34], which begs the question of how these complex methylation patterns are established. The Arabidopsis genome sequence has revealed the existence of at least 10 genes that encode DNA methyltransferases (MTases), and genetic analysis has, to date, implicated four MTases in RdDM. DRM1 and DRM2, which are related to the Dmnt3 group of mammalian de novo MTases, are essential for the establishment of RdDM and effect some degree of methylation in all sequence contexts. The chromodomain-containing MTase CMT3, which belongs to a family that is unique to plants, appears to be responsible for additional levels of CNG and asymmetric methylation [35]. As for MET1, which is related to the maintenance MTase Dmnt1 of mammals, this MTase has a dual role, leading to higher levels of CG methylation and being responsible for the maintenance of methylation at CG sites that is observed once the RNA trigger has been removed [24*,35,36].

Forward-genetic screens have also uncovered a putative histone deacetylase and a putative chromatin protein as components of the RdDM machinery [37,38]. This is consistent with the notion that RdDM is part of a chromatin modification system that extends beyond DNA methylation ([13,22,23,39,40]; see next section).

**DNA methylation and histone modifications**

As in mammals and the fungus Neurospora, DNA methylation in plants is closely associated with specific histone modifications [22,39,40]. The extent to which this association holds true has been assessed using a genomic DNA tiling microarray that covers 0.5 Mb of heterochromatin and 1 Mb of neighboring euchromatin on Arabidopsis chromosome 4. A near-perfect match was found between DNA methylation and H3K9me2, a typical
heterochromatin mark. Moreover, these two marks map almost exclusively to transposable elements and related repeats, both within and outside cytologically visible heterochromatin. They also adorn most of the sequences within the 1.5 Mb region that match cloned siRNAs, consistent with a role for these siRNAs in guiding silent chromatin formation [27**]. In another study, the chromodomain of the DNA methyltransferase CMT3 was shown to interact in vitro with histone H3 trimethylated at lysine 9 (H3K9me3), but only when this histone was also trimethylated at lysine 27 (H3K27me3) [41]. The significance of this observation is unclear, however, given that H3K9me3 and H3K27me3 have not been reliably detected in plants, though they have in mammals [41,42]. Furthermore, the H3K9me2 and H3K27me1 (or 2) co-localization seen in vivo is not always associated with DNA methylation, ruling out a simple relationship of causality [41,43,44*]. How H3K9me2 controls CMT3-mediated CNG methylation therefore remains an open question [22,41]. In addition, a null allele of the MET1 gene has been shown to lead to a loss of H3K9me2 methylation in heterochromatin [45], which would suggest opposite causal relationships between H3K9me2 and CG and CNG methylation, and some form of mutual reinforcement between DNA methylation and histone modifications [39,40,46].

**Epigenetic control, plant development and natural variation**

Few genes have convincingly been shown to be epigenetically regulated during plant development, a situation that probably reflects the different life strategies of plants and animals [47]. Among the best-known cases of epigenetically regulated plant genes are the three Arabidopsis genes MEDEA (MEA), FWA and FLOWERING LOCUS C (FLC). The floral repressor gene FLC becomes stably repressed during vegetative growth following exposure to a prolonged period of cold in vernalization-responsive genetic backgrounds [48,49]. MEA and FWA are subject to genomic imprinting in the endosperm, where they are expressed from the maternal chromosome [50,51**]. Imprinting of MEA and FWA involves the specific reactivation of the maternal allele in the female gamete before fertilization, and is mediated by the antagonistic actions of the DNA methyltransferase MET1 and the DNA glycosylase DEMETER (DME), which may serve as a DNA demethylase [51**,52]. In support of this latter role, reactivation of maternal FWA was found to be associated with hypomethylation of its promoter sequence specifically in the endosperm, and it has been proposed that DNA-methylation-associated gene silencing could serve as a ‘one-way’ control system during plant development [51**]. Indeed, because DNA hypomethylation is usually transmitted stably through meiosis in Arabidopsis [20,53,54], gene reactivation or gene demethylation in a developmental context would only occur in tissues such as the endosperm that do not contribute to the next generation. Although this proposition requires further studies, it is worth noting that vernalization-induced inactivation of the FLC gene, which needs to be erased at each generation, relies on a mechanism that involves H3K9me2 and H3K27me2, as well as Polycomb group and other chromatin proteins, but (apparently) not DNA methylation [43*,44*,48,49].

In contrast to its limited role during development, DNA methylation is associated with the stable silencing of transposable elements (TEs) in Arabidopsis [27**,55*], and significantly, most Arabidopsis siRNAs that have been cloned and sequenced correspond to TEs and related repeats [13,29*]. In addition, the appearance of ‘epimutant’ forms of genes is often associated with hyper- or hypomethylation of the affected loci [15,56]. These ‘epialleles’, despite being metastable by nature, can be transmitted through many generations in some instances, and could represent a significant fraction of the natural genetic variation seen in plants [15,56]. As predicted by Barbara McClintock more than 50 years ago on the basis of observations made with maize [57], the insertion of TEs is often involved in the formation of gene structures that have the capacity to ‘epimutate’ [18]. Significantly, the Arabidopsis genome is not exempt from the formation of such gene structures, despite the relatively small number of transposable elements it contains [27**,58,59].

**Conclusions**

Although a role for RNAi in silent chromatin formation is now firmly established in Arabidopsis, the molecular details remain to be elucidated for the most part. For instance, the accumulation of siRNAs that correspond to TEs and other repeats is recurrently found not to be uniformly affected in any given RNAi or chromatin mutant, with some siRNAs disappearing whereas others remain unchanged or overaccumulate [27**,28*–30*,37,38,55*]. This might indicate that different complexes or pathways are involved in RdDM and transcriptional silencing, depending on the nature of the repeats [30*,55*]. Another intriguing observation is that although the inactivation of the PAI2 and PAI3 genes, or of various transgenes, is readily achieved by the introduction of unlinked inverted repeats through crossing [24*,30*], similar inactivation rarely occurs when mutant plants with reactivated endogenous TEs are crossed to wildtype plants [27**,55*]. This could be due to a strong cis-preference in chromatin targeting by siRNAs, which would only be overcome in instances in which siRNAs accumulate to high levels. This explanation can only be partial, however, as trans-silencing of the PAI2 and PAI3 genes occurs in the absence of detectable siRNAs [24*]. More generally, we know little about the way siRNAs might interact with chromatin to direct DNA and histone modifications. For instance, does recognition occur through direct RNA–DNA pairing, or through pairing with nascent transcripts produced from the target?
Finally, we also know little about the nature and timing of the transcriptional activity that is (paradoxically) required across silent repeat elements to feed the RNAi machinery that produces the corresponding endogenous siRNAs. Thanks to the ever-increasing power of Arabidopsis genetics and genomics, we can expect answers to such questions over the next few years.

Acknowledgements
We apologize to our colleagues whose work we did not cite because of space limitations. We thank our collaborators R. Martienssen and Z. Lippman for discussions and E. Heard for comments on the manuscript. A-VG is the recipient of a Graduate Studentship from the French Ministry of Education and Research. VC is supported by grants from Genopole, the Centre National de la Recherche Scientifique (CNRS), and the European Union (NoE “The Epigenome”).

References and recommended reading
Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest


8. Using a wheat-germ extract, the authors provide the demonstration of RNAi in plants and of the existence of at least two siRNAs, but not siRNAs, corresponding to the inverted repeat.


14. The authors of this paper demonstrate that transcription of the PAI1-PAI4 inverted repeat is necessary for DNA methylation and silencing of the genetically unlinked PAI2 and PAI3 genes. Moreover, they detected long dsRNAs, but not siRNAs, corresponding to the inverted repeat.


17. The work described in this short paper demonstrates the role of AGO4, DCL3, RDR2 and SDE4 (an as yet uncharacterized RNAi gene) in initiating DNA methylation of an FWA transgene upon transformation, and in perpetuating its non-CG DNA methylation.


The authors describe the first extensive analysis of transcription and of DNA and histone modification across a large chromosomal region in Arabidopsis. Transposable elements and related repeats were found to define heterochromatin, and to bring genes under its epigenetic control when inserted within them, as in the case of the FWA gene. In addition, a mutant allele of the DDM1 gene, which encodes a chromatin remodeling factor, was found to lead to widespread loss of heterochromatin marks over repeated elements and to a heritable accumulation of transcripts corresponding to these elements.


20. The authors report the first identification of an RNAi gene that is involved in transcriptional silencing and chromatin modification in Arabidopsis.


22. Using a genetic approach, the authors provide evidence of the existence of at least three RNAi pathways in plants that are involved in the formation of functionally distinct small RNAs. In particular, they show that DCL3 is involved in the production of siRNAs that correspond to transposable elements and other repeats. Loss of these siRNAs in dcl3 mutants was
associated with transcript accumulation and loss of heterochromatin marks at some loci.


Following the identification of AGO4 (see [28]), the authors address the question of its involvement in diverse aspects of transcriptional or post-transcriptional transgene silencing triggered by inverted repeats. The authors conclude that different RNAi pathways might control DNA methylation at direct repeats and inverted repeats.


This and the following landmark paper [32] describe the biochemical characterization of two RNAi complexes that are required for heterochromatin assembly in the fission yeast S. pombe.


See annotation for [31]**.


This paper describes the transcriptional and chromatin modification status of five Arabidopsis transposable elements in wildtype and several mutant backgrounds. The presence of siRNAs was also investigated in some cases. The results indicate the existence of distinct, but interacting, complexes or pathways that are involved in transposon silencing. Furthermore, transposable elements were found to differ in their susceptibility to different forms of epigenetic regulation.


See annotation for [43].


The authors show that imprinting of FWA in the endosperm is not established by allele-specific de novo DNA methylation, unlike that of imprinted genes in mammals. Instead, it occurs through maternal gametophyte-specific demethylation and gene activation. The authors propose that, given the lack of widespread de novo methylation activity during plant development, DNA methylation is used as a ‘one way’ control system of gene expression.


