Expansion and Contraction of Small RNA and Methylation Machinery Throughout Plant Evolution

Tania Chakraborty[a], Hayden Payne[a], and Rebecca A. Mosher[a,b]

[a] School of Plant Sciences, University of Arizona, Tucson, AZ 85721-0036, USA.

[b] Corresponding author: rmosher@email.arizona.edu

Abstract

The revolution in sequencing has created a wealth of plant genomes that can be mined to understand the evolution of biological complexity. Complexity is often driven by gene duplication, which allows paralogs to specialize in an activity of the ancestral gene or acquire novel functions. Angiosperms encode a variety of gene silencing pathways that share related machinery for small RNA biosynthesis and function. Recent phylogenetic analysis of these gene families plots the expansion, specialization, and occasional contraction of this core machinery. This analysis reveals the ancient origin of RNA-directed DNA Methylation in early land plants, or possibly their algal ancestors, as well as ongoing duplications that evolve novel small RNA pathways.

Introduction

In plants, small RNAs silence genes through various mechanisms, but each type of small RNA uses similar core machinery for biosynthesis and function (**Figure 1**). Small RNA production starts with an RNA transcript produced by Pol II or a specialized RNA polymerase [1]. This transcript becomes double-stranded through folding or by the action of an RNA-dependent RNA polymerase (RDR), and double-stranded RNA is processed into small RNAs by Dicer-Like (DCL) endonucleases [2–8]. Small RNAs are loaded into Argonaute (AGO) proteins and direct these effectors to complementary transcripts, resulting in various forms of gene silencing [9–15]. The fission yeast *Schizosaccharomyces pombe* contains single copies of *RDR*, *Dicer*, and *AGO*, while *Arabidopsis thaliana* contains several paralogs of each gene, along with two specialized RNA polymerases [1,16]. Recent years have generated an abundance of plant genome sequences, allowing us to trace duplications of small RNA core machinery across the evolution of land plants and into their closest algal ancestors.

Plant-specific RNA polymerases

In angiosperms, most small RNAs are 24-nt siRNAs from the RNA-directed DNA Methylation (RdDM) pathway, which is distinct from other small RNA pathways in its requirement for two plant-specific DNA-dependent RNA polymerases, Pol IV and Pol V [1,16]. The specialized activities of Pol IV and V arise through the duplication and specialization of at least five of the twelve subunits that make up these complexes. In all cases, specialized Pol IV and V subunits evolved through duplication and divergence of Pol II subunits [17,18], although these duplications occurred at different times across plant evolution (**Figure 2**). All land plant lineages contain at least two additional copies of the largest subunit [19–22], although You and colleagues argue that true differentiation of Pol IV and V only occurred following the divergence between ferns and seed plants [22]. However, other research groups identified distinct Pol IV and Pol V largest subunits in bryophytes, suggesting that the differentiation of Pol II, Pol IV, and Pol V occurred prior to plant terrestrialization [19–21]. Pol IV and V share second and seventh subunits that are paralogous to Pol II subunits, and these are also found in all extant land plant lineages [18–20,22]. Duplication and specialization of the Pol V-specific fifth subunit occurred following the divergence of ferns from seed plants, while the Pol IV/V-specific fourth subunit is only observed in angiosperms [19,20,22] (**Figure 2**).

While there is no evidence for specialized polymerase subunits in green algae (Chloroplastida), a partial sequence of an NRPD1-like gene was identified from multiple genera of Charophycean green algae (CGA) [17]. CGAs comprise a series of lineages, from the Mesostigmatophyceae, which are the most diverged from land plants (Embryophyta), to the Zygnemophyceae, which are sister to land plants [23–25]. As described below, CGAs might encode other components related to RdDM, raising the possibility that this pathway evolved prior to terrestrialization. However, a recent study of CGA genomes and transcriptomes found no evidence for Pol IV or Pol V subunits [26], leaving ambiguity about when these characteristic polymerases, and the RdDM pathway, first arose.

Core small RNA machinery

Of the six *RDR* genes in Arabidopsis, three are associated with small RNA production: *RDR1* is associated with resistance to viral infection, *RDR2* functions in RdDM, and *RDR6* is primarily associated with the production of phased siRNAs following microRNA cleavage of mature transcripts [27,28]. A distinct *RDR6* is ubiquitous in extant land plant genomes and most CGA lineages, suggesting that specialization among RDR proteins began in the algal ancestors of land plants [19,22,26,29]. CGAs, bryophytes, and other non-seed plants also encode a single-copy *RDR1*/*RDR2* ancestor, which resolved into distinct *RDR1* and *RDR2* genes in the

common ancestor of seed plants [19,26,29]. Mutation of the *RDR1*/*RDR2* gene in the moss *Physcomitrium patens* (a bryophyte) results in loss of 23-24-nt siRNAs, indicated that the pre-duplication gene likely functioned with Pol IV to produce siRNAs [21].

Arabidopsis encodes four DCL proteins, all of which are involved in small RNA production. DCL1 homologs, which cleave structured single-stranded RNA to produce microRNAs, exist in all major land plant lineages and in some CGA groups [19,26], while Chlorophytic algae encode algae-specific DCLs that likely provide the same function [26,30]. Zygnemophyceae, the group of CGAs most closely related to bryophytes and other land plants [24,25,31], encode an ancestral DCL2/DCL3/DCL4 sequence, which is duplicated to form DCL3 and DCL2/4 clades in bryophytes [19,26]. In the bryophyte *Physcomitrium patens*, loss of DCL3 causes reduced 24-nt siRNA accumulation and developmental phenotypes similar to Pol IV and RDR2 mutants, supporting its role in RdDM [21]. Whether the DCL2/DCL3/DCL4 ancestral CGA gene also functions in RdDM is unknown.

AGOs are a large gene family, with ten homologs in Arabidopsis that form three major clades: *AGO1/5/10*, *AGO2/3/7*, and *AGO4/6/8/9* [32]. Although the number of genes in each clade varies among different species, these three clades are ancient, evolving from an algae-specific AGO in the green algae [26,33]. Since *AGO4*, *AGO6*, *AGO8*, and *AGO9* are associated with RdDM [9,11,12,15,34–37], the presence of this clade in CGA species is further evidence that RdDM might have functioned in the aquatic ancestor of land plants.

Methyltransferases

Just as duplications of core small RNA machinery have led to specialized machinery for RdDM, plants encode multiple DNA methyltransferases with divergent functions. MET1, a DNMT1 homolog, maintains methylation in the CG context, while the chromomethyltransferases CMT2 and CMT3 bind to Histone 3 Lysine 9 dimethylation and catalyze CHH and CHG methylation, respectively [38,39]. Plants also encode Domains Rearranged Methyltransferase (DRM), a DNMT3 homolog that establishes DNA methylation in all three cytosine contexts in a 24-nt siRNA dependent manner [39,40] (**Figure 3**). *DRM* homologs have been identified across land plants and in at least one CGA lineage [41]. Eliminating *DRM* function in *Physcomitrium patens* has little impact on global CHH methylation levels, but causes hypomethylation of transcriptionally-active, euchromatic transposons, suggesting that *DRM* is responsible for RdDM even in the earliest land plants [41,42].

Ongoing duplications

The step-wise duplication of polymerase subunits and core small RNA machinery continues in specific angiosperm families. Genomes of Poaceae members (monocots) contain duplicates of the largest two subunits of Pol V, which have been conserved for over 50 million years [43]. There are many indications that these paralogs have distinct functions: a signature of selection along the interacting surface between the first and second subunits and differential protein-protein interactions suggest assembly into different holoenzyme complexes, while shortening of the C-terminal domain and different mutant phenotypes indicate different molecular and biological functions [43–45]. As more genomes are sequenced, we might discover additional elaboration of polymerase complexes.

DCL3 duplicated in the ancestor of monocots, generating DCL3, a paralog that is associated with canonical RdDM and silencing of transposons, and DCL5, a paralog that produces phased 24-nt siRNAs from long non-coding transcripts [46–49]. However, since both classes of 24-nt small RNAs are produced in eudicot species [50], DCL5 likely arose through sub-functionalization in monocots. Phased 24-nt siRNAs induce high CHH methylation in cis during anther development in maize (a monocot), indicating that they are part of a non-canonical RdDM pathway [51].

One of the most labile gene families associated with small RNAs is the Argonaute family. Across the three main clades of AGOs, angiosperm genomes encode an average of 13 family members [33]. What drives the expansion of AGOs is unknown, however, one reason might be that additional AGOs allow unique functions or tissue-specificity [35,52,53]. In addition to expansion of these three ancestral groups, grasses contain an additional subclade within the AGO1/5/10 group, AGO18, that might be associated with phased siRNAs [32,33,54,55]. Additional phylogenetic and functional analysis is necessary to determine whether the many Argonaute paralogs represent recent duplications or deeply-conserved, and potentially neofunctionalized, genes.

Tuning methylation through allelic diversity

In addition to duplications that elaborate and expand small RNA machinery, sometimes components of RdDM are modified, or even eliminated, in individual plant species, which can create variation in the developmental pattern or genomic context of methylation across land plants [42,56–60]. Since RdDM suppresses transposon mobility, changes in RdDM components may lead to transposon proliferation and genome expansion, potentially accounting for the gigantic genomes in lineages such as lycophytes and gymnosperms [22,61,62]. Consistent with this idea, 24-nt siRNAs, a key indicator of RdDM, constitute only a small proportion of sRNA

libraries outside of angiosperms [63], suggesting that RdDM is less active in these species. However, this lack of representation might be due to analysis of relatively few tissues, since 24-nt siRNAs are easily detected in gymnosperm reproductive tissues [64–68]. The presence of 24-nt siRNAs in gymnosperms indicates that genomes can grow very large despite active RdDM. Concordantly, the *Spirodela* (a monocot) genome remains small despite evidence that RdDM has been lost [69,70]. Together, these observations indicate that RdDM is not a primary factor in genome size variation.

Although changes in RdDM are unlikely to be responsible for genome expansion, allelic variation at Pol V and *AGO9* is associated with the level of CHH methylation at euchromatic transposons throughout the genome [71,72]. Variation at other small RNA pathway components might also influence global methylation levels [71]. These subtle changes in CHH methylation suggest that methylation pathways are not simply on or off, but instead are tunable to ecological conditions. Indeed, Arabidopsis grown at a lower temperature displays decreased CHH methylation [73], possibly due to reduced siRNA production at low temperature [74]. Selection might therefore tune the efficiency of methylation machinery to match an ecological niche [75].

Loss of methylation pathways

In some species, DNA methylation is reduced to the point of elimination. Maize and some Arabidopsis accessions lack active *CMT2* and therefore have no CHH methylation at heterochromatin transposons [73,75–78]. However, euchromatic transposons in these genomes retain CHH methylation due to RdDM [73,76,78]. The apparent success of these lineages raises the question of why *CMT2* has been retained as a distinct clade in angiosperms if heterochromatin and RdDM are sufficient for genome defense against transposons [38]. Unexpectedly, some *CMT2*-null accessions of Arabidopsis display robust CHH methylation at heterochromatic transposons, suggesting that additional redundancy and plasticity among methylation pathways remains to be discovered [75].

Spirodela polyrhiza, a small aquatic monocot that propagates through rapid asexual reproduction, provides an extreme example of loss of small RNA and methylation pathways [79]. *Spirodela* lacks *CMT2* and is deficient in expression of some RdDM components, resulting in essentially no CHH methylation and reduced levels of CHG and CG methylation [69,70,80]. Reduced CHH methylation has been observed in other clonally-propagated species, as well as at least one other aquatic plant, suggesting that either of these lifestyle factors might make CHH methylation dispensable [57,80].

Conclusion

The recent explosion in plant genome sequences allows us to trace the expansion and contraction of small RNA machinery across the plant family, uncovering diverse complements of proteins. Contrary to the assumption of ever-increasing complexity, there are also losses of conserved machinery, suggesting that genomes have evolved a wide range of gene silencing approaches. However, too often we lack direct evidence to conclude that homologous proteins function similarly across hundreds of millions of years of evolution. Functional studies in an expanding list of species are necessary to understand the diverse ways to manage a genome.

Acknowledgements

This work was supported by the National Science Foundation (MCB-1929678 to RAM).

Declaration of Interest

none

Figure Captions

Figure 1. Major components of prevalent plant small RNA pathways.
The core proteins or complexes responsible for small RNA biosynthesis and function are shown for the most common sources of plant small RNAs: RNA polymerases (Pol), RNA-dependent RNA Polymerase (RDR), Dicer-like endonuclease (DCL), and Argonaute effectors (AGO). With the exception of viral polymerases, which vary depending on the type of virus, all components in the same row are members of the same gene family. Ancestral genes duplicated and specialized for particular small RNA pathways, as listed at the top. MicroRNAs (miRNA) form dsRNA through single-molecule folding and do not require an enzyme for dsRNA synthesis. The Argonaute(s) associated with 24-nt phasiRNAs are unknown, but presumed to be members of the AGO4 clade. Only the most common Argonaute effectors are listed for each pathway.

Figure 2. Evolutionary history of small RNA machinery in plants.
Subunits of Pol II, IV, V, and VI are named NRPB, NRPD, NRPE, and NRPF, respectively, with the number indicating the subunit (1 for largest, 2 for second-largest, etc). Subunits other than the ones depicted are shared between Pol II, Pol IV, and Pol VI. Grey bars indicate the presence of presumed orthologous sequences in each plant lineages, with forks indicating the timing of gene duplication. The unfilled, dotted fork indicates ambiguity regarding the timing of

the DCL2/DCL4 divergence. Gene names are colored by their presumed function. Purple = RdDM, blue = microRNA, green = viral resistance, orange = other or unknown.

Figure 3: DNA methylation pathways in angiosperms.

(A) Methylation at CG sites is maintained by MET1. Semi-conservative DNA replication combines an older, methylated strand with a newly-synthesized unmethylated strand. MET1 recognizes the resulting hemi-methylated site and induces full methylation. (B) In euchromatin, asymmetric CHH sites do not result in hemi-methylated sites following replication and must be maintained by siRNA targeting of DRM methyltransferase (RdDM). (C, D) In heterochromatin, non-CG sites are maintained by the alternating actions of a histone methyltransferase (KYP/SUVH4, SUVH5, and/or SUVH6) and chromomethyltransferases (CMT2 and CMT3). The histone methyltransferase recognizes CHG or CHH methylation and induces dimethylation of lysine 9 on histone H3 (H3K9me2). This histone modification is recognized by CMT2 and CMT3, which cause CHH and CHG methylation, respectively.

Highlighted references

♦♦ Wang S, Liang H, Xu Y, Li L, Wang H, Sahu DN, Petersen M, Melkonian M, Sahu SK, Liu H: **Genome-wide analyses across Viridiplantae reveal the origin and diversification of small RNA pathway-related genes**. *Commun Biol* 2021, **4**:412.
This manuscript describes comprehensive phylogenetic analysis of the evolutionary history of small RNA biosynthesis proteins. The authors show that multiple RdDM pathway components, as well as founding members of AGO, DCL, and RDR clades are present in Charophytic green algae.

♦ Patel P, Mathioni SM, Hammond R, Harkess AE, Kakrana A, Arikit S, Dusia A, Meyers BC: **Reproductive phasiRNA loci and DICER-LIKE5, but not microRNA loci, diversified in monocotyledonous plants**. *Plant Physiol* 2021, **185**:1764–1782.
Using small RNA analysis from a wide range of taxa across angiosperms, this study showed that monocots have a specialized DCL paralog and expanded phasiRNA production.

♦♦ Zhang M, Ma X, Wang C, Li Q, Meyers BC, Springer NM, Walbot V: **CHH DNA methylation increases at 24-PHAS loci depend on 24-nt phased small interfering RNAs in maize meiotic anthers**. *New Phytol* 2021, **229**:2984–2997.

This study demonstrates that 24-nt phasiRNA trigger CHH methylation at PHAS loci during meiosis in maize anthers. While Pol IV dependent siRNAs are associated with de novo CHH methylation, this study shows that 24-nt phasiRNAs, produced by a different mechanism, can also establish CHH methylation.

♦ Domb K, Katz A, Harris KD, Yaari R, Kaisler E, Nguyen VH, Hong UVT, Griess O, Heskiau KG, Ohad N, et al.: **DNA methylation mutants in Physcomitrella patens elucidate individual roles of CG and non-CG methylation in genome regulation**. *Proc Natl Acad Sci U S A* 2020, **117**:33700–33710.
The authors create mutants lacking various methyltransferases to demonstrate the influence of DNA methylation pathways in *Physcomitrium*. They discover that DRM2 plays a minor role in CHH methylation, highlighting the diverse ways that lineages apply conserved machinery.

♦♦ Li Z, Li W, Guo M, Liu S, Liu L, Yu Y, Mo B, Chen X, Gao L: **Origin, evolution and diversification of plant ARGONAUTE proteins**. *Plant J* 2022, **109**:1086–1097.
Using phylogenetic analyses, the authors show that homologs of AGO 4/6/8/9, AGO 2/3/7, and AGO 1/5/10 originate in Charophytic green algal. They also identify the many duplications that create the large AGO families found in angiosperms.

♦ Zheng K, Wang L, Zeng L, Xu D, Guo Z, Gao X, Yang D-L: **The effect of RNA polymerase V on 24-nt siRNA accumulation depends on DNA methylation contexts and histone modifications in rice**. *Proc Natl Acad Sci U S A* 2021, **118**.
Although primarily focused on the function of RNA Pol V in rice, this manuscript also demonstrates that Pol V and its recently-duplicated paralog, Pol VI, are non-redundant. Along with Ref. 48, this finding is critical evidence that recent duplications of small RNA machinery are creating novel functional pathways.

♦ Teng C, Zhang H, Hammond R, Huang K, Meyers BC, Walbot V: **Dicer-like 5 deficiency confers temperature-sensitive male sterility in maize**. *Nat Commun* 2020, **11**:2912.
This manuscript assigns function to DCL5, a monocot-specific paralog of DCL3, demonstrating that it is involved in the production of 24-nt phasiRNAs in maize tapetal cells. Together with Ref. 45, this research supports the conclusion that functional novelty continues to arise throughout plant evolution.

References

1. Zhou M, Law JA: **RNA Pol IV and V in gene silencing: Rebel polymerases evolving away from Pol II's rules**. *Curr Opin Plant Biol* 2015, **27**:154–164.

2. Zhai J, Bischof S, Wang H, Feng S, Lee T-F, Teng C, Chen X, Park SY, Liu L, Gallego-Bartolome J, et al.: **A One Precursor One siRNA Model for Pol IV-Dependent siRNA Biogenesis**. *Cell* 2015, **163**:445–455.

3. Blevins T, Podicheti R, Mishra V, Marasco M, Wang J, Rusch D, Tang H, Pikaard CS: **Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis**. *Elife* 2015, **4**:e09591.

4. Li S, Vandivier LE, Tu B, Gao L, Won SY, Li S, Zheng B, Gregory BD, Chen X: **Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in Arabidopsis reveals features and regulation of siRNA biogenesis**. *Genome Res* 2015, **25**:235–245.

5. Singh J, Mishra V, Wang F, Huang H-Y, Pikaard CS: **Reaction Mechanisms of Pol IV, RDR2, and DCL3 Drive RNA Channeling in the siRNA-Directed DNA Methylation Pathway**. *Mol Cell* 2019, **75**:576-589.e5.

6. Huang K, Wu X-X, Fang C-L, Xu Z-G, Zhang H-W, Gao J, Zhou C-M, You L-L, Gu Z-X, Mu W-H, et al.: **Pol IV and RDR2: A two-RNA-polymerase machine that produces double-stranded RNA**. *Science* 2021, **374**:1579–1586.

7. Wang Q, Xue Y, Zhang L, Zhong Z, Feng S, Wang C, Xiao L, Yang Z, Harris CJ, Wu Z, et al.: **Mechanism of siRNA production by a plant Dicer-RNA complex in dicing-competent conformation**. *Science* 2021, **374**:1152–1157.

8. Loffer A, Singh J, Fukudome A, Mishra V, Wang F, Pikaard CS: **A DCL3 dicing code within Pol IV-RDR2 transcripts diversifies the siRNA pool guiding RNA-directed DNA methylation**. *Elife* 2022, **11**.

9. McCue AD, Panda K, Nuthikattu S, Choudury SG, Thomas EN, Slotkin RK: **ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation**. *EMBO J* 2015, **34**:20–35.

10. Böhmdorfer G, Sethuraman S, Rowley MJ, Krzyszton M, Rothi MH, Bouzit L, Wierzbicki AT: **Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin**. *Elife* 2016, **5**.

11. Lahmy S, Pontier D, Bies-Etheve N, Laudié M, Feng S, Jobet E, Hale CJ, Cooke R, Hakimi M-A, Angelov D, et al.: **Evidence for ARGONAUTE4-DNA interactions in RNA-directed DNA methylation in plants**. *Genes Dev* 2016, **30**:2565–2570.

12. Wang F, Axtell MJ: **AGO 4 is specifically required for heterochromatic si RNA accumulation at Pol V-dependent loci in Arabidopsis thaliana**. *Plant J* 2017, **90**:37–47.

13. Wendte JM, Haag JR, Singh J, McKinlay A, Pontes OM, Pikaard CS: **Functional Dissection of the Pol V Largest Subunit CTD in RNA-Directed DNA Methylation**. *Cell Rep* 2017, **19**:2796–2808.

14. Liu W, Duttke SH, Hetzel J, Groth M, Feng S, Gallego-Bartolome J, Zhong Z, Kuo HY, Wang Z, Zhai J, et al.: **RNA-directed DNA methylation involves co-transcriptional small-RNA-guided slicing of polymerase V transcripts in Arabidopsis**. *Nat Plants* 2018, **4**:181–188.

15. Sigman MJ, Panda K, Kirchner R, McLain LL, Payne H, Peasari JR, Husbands AY, Slotkin RK, McCue AD: **An siRNA-guided Argonaute protein directs RNA Polymerase V to initiate DNA methylation**. 2021, doi:10.1101/2021.08.19.457027.

16. Matzke MA, Mosher RA: **RNA-directed DNA methylation: an epigenetic pathway of increasing complexity**. *Nat Rev Genet* 2014, **15**:394–408.

17. Luo J, Hall BD: **A multistep process gave rise to RNA polymerase IV of land plants**. *J Mol Evol* 2007, **64**:101–112.

18. Tucker SL, Reece J, Ream TS, Pikaard CS: **Evolutionary history of plant multisubunit RNA polymerases IV and V: subunit origins via genome-wide and segmental gene duplications, retrotransposition, and lineage-specific subfunctionalization**. *Cold Spring Harb Symp Quant Biol* 2010, **75**:285–297.

19. Huang Y, Kendall T, Forsythe ES, Dorantes-Acosta A, Li S, Caballero-Perez J, Chen X, Arteaga-Vázquez M, Beilstein MA, Mosher RA: **Ancient Origin and Recent Innovations of RNA Polymerase IV and V**. *Mol Biol Evol* 2015, **32**:1788–1799.

20. Wang Y, Ma H: **Step-wise and lineage-specific diversification of plant RNA polymerase genes and origin of the largest plant-specific subunits**. *New Phytol* 2015, **207**:1198–1212.

21. Coruh C, Cho SH, Shahid S, Liu Q, Wierzbicki A, Axtell MJ: **Comprehensive Annotation of Physcomitrella patens Small RNA Loci Reveals That the Heterochromatic Short Interfering RNA Pathway Is Largely Conserved in Land Plants**. *Plant Cell* 2015, **27**:2148–2162.

22. You C, Cui J, Wang H, Qi X, Kuo L-Y, Ma H, Gao L, Mo B, Chen X: **Conservation and divergence of small RNA pathways and microRNAs in land plants**. *Genome Biol* 2017, **18**.

23. Domozych DS, Popper ZA, Sørensen I: **Charophytes: Evolutionary giants and emerging model organisms**. *Front Plant Sci* 2016, **7**:1470.

24. de Vries J, Archibald JM: **Plant evolution: landmarks on the path to terrestrial life**. *New Phytol* 2018, **217**:1428–1434.

25. Timme RE, Bachvaroff TR, Delwiche CF: **Broad phylogenomic sampling and the sister lineage of land plants**. *PLoS One* 2012, **7**:e29696.

26. Wang S, Liang H, Xu Y, Li L, Wang H, Sahu DN, Petersen M, Melkonian M, Sahu SK, Liu H: **Genome-wide analyses across Viridiplantae reveal the origin and diversification of small RNA pathway-related genes**. *Commun Biol* 2021, **4**:412.

27. Willmann MR, Endres MW, Cook RT, Gregory BD: **The Functions of RNA-Dependent RNA Polymerases in Arabidopsis**. *Arabidopsis Book* 2011, **9**:e0146.

28. Borges F, Martienssen RA: **The expanding world of small RNAs in plants**. *Nat Rev Mol Cell Biol* 2015, **16**:727–741.

29. Bélanger S, Zhan J, Meyers BC: **Genome-wide analysis of small RNA biogenesis proteins refine the evolution of Dicer-like and Argonaute gene families in flowering plants**. *bioRxiv* 2022, doi:10.1101/2022.01.18.476847.

30. Valli AA, Santos BACM, Hnatova S, Bassett AR, Molnar A, Chung BY, Baulcombe DC: **Most microRNAs in the single-cell alga Chlamydomonas reinhardtii are produced by Dicer-like 3-mediated cleavage of introns and untranslated regions of coding RNAs**. *Genome Res* 2016, **26**:519–529.

31. Jiao C, Sørensen I, Sun X, Sun H, Behar H, Alseekh S, Philippe G, Palacio Lopez K, Sun L, Reed R, et al.: **The Penium margaritaceum Genome: Hallmarks of the Origins of Land Plants**. *Cell* 2020, **181**:1097-1111.e12.

32. Zhang H, Xia R, Meyers BC, Walbot V: **Evolution, functions, and mysteries of plant ARGONAUTE proteins**. *Curr Opin Plant Biol* 2015, **27**:84–90.

33. Li Z, Li W, Guo M, Liu S, Liu L, Yu Y, Mo B, Chen X, Gao L: **Origin, evolution and diversification of plant ARGONAUTE proteins**. *Plant J* 2022, **109**:1086–1097.

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

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

36. Olmedo-Monfil V, Durán-Figueroa N, Arteaga-Vázquez M, Demesa-Arévalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada J-P: **Control of female gamete formation by a small RNA pathway in Arabidopsis**. *Nature* 2010, **464**:628–632.

37. Hernández-Lagana E, Rodríguez-Leal D, Lúa J, Vielle-Calzada J-P: **A multigenic network of ARGONAUTE4 clade members controls early megaspore formation in Arabidopsis**. *Genetics* 2016, **204**:1045–1056.

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

39. de Mendoza A, Lister R, Bogdanovic O: **Evolution of DNA methylome diversity in eukaryotes**. *J Mol Biol* 2019, **432**:1687–1705.

40. Zhang H, Lang Z, Zhu J-K: **Dynamics and function of DNA methylation in plants**. *Nat Rev Mol Cell Biol* 2018, **19**:489–506.

41. Yaari R, Katz A, Domb K, Harris KD, Zemach A, Ohad N: **RdDM-independent de novo and heterochromatin DNA methylation by plant CMT and DNMT3 orthologs**. *Nat Commun* 2019, **10**:1613.

42. Domb K, Katz A, Harris KD, Yaari R, Kaisler E, Nguyen VH, Hong UVT, Griess O, Heskiau KG, Ohad N, et al.: **DNA methylation mutants in Physcomitrella patens elucidate individual roles of CG and non-CG methylation in genome regulation**. *Proc Natl Acad Sci U S A* 2020, **117**:33700–33710.

43. Trujillo JT, Seetharam AS, Hufford MB, Beilstein MA, Mosher RA: **Evidence for a Unique DNA-Dependent RNA Polymerase in Cereal Crops**. *Mol Biol Evol* 2018, **35**:2454–2462.

44. Haag JR, Brower-Toland B, Krieger EK, Sidorenko L, Nicora CD, Norbeck AD, Irsigler A, LaRue H, Brzeski J, McGinnis K, et al.: **Functional diversification of maize RNA polymerase IV and V subtypes via alternative catalytic subunits**. *Cell Rep* 2014, **9**:378–390.

45. Zheng K, Wang L, Zeng L, Xu D, Guo Z, Gao X, Yang D-L: **The effect of RNA polymerase V on 24-nt siRNA accumulation depends on DNA methylation contexts and histone modifications in rice**. *Proc Natl Acad Sci U S A* 2021, **118**.

46. Johnson C, Kasprzewska A, Tennessen K, Fernandes J, Nan G-L, Walbot V, Sundaresan V, Vance V, Bowman LH: **Clusters and superclusters of phased small RNAs in the developing inflorescence of rice**. *Genome Res* 2009, **19**:1429–1440.

47. Wei L, Gu L, Song X, Cui X, Lu Z, Zhou M, Wang L, Hu F, Zhai J, Meyers BC, et al.: **Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice**. *Proc Natl Acad Sci U S A* 2014, **111**:3877–3882.

48. Teng C, Zhang H, Hammond R, Huang K, Meyers BC, Walbot V: **Dicer-like 5 deficiency confers temperature-sensitive male sterility in maize**. *Nat Commun* 2020, **11**:2912.

49. Patel P, Mathioni SM, Hammond R, Harkess AE, Kakrana A, Arikit S, Dusia A, Meyers BC: **Reproductive phasiRNA loci and DICER-LIKE5, but not microRNA loci, diversified in monocotyledonous plants**. *Plant Physiol* 2021, **185**:1764–1782.

50. Xia R, Chen C, Pokhrel S, Ma W, Huang K, Patel P, Wang F, Xu J, Liu Z, Li J, et al.: **24-nt reproductive phasiRNAs are broadly present in angiosperms**. *Nat Commun* 2019, **10**:627.

51. Zhang M, Ma X, Wang C, Li Q, Meyers BC, Springer NM, Walbot V: **CHH DNA methylation increases at 24-PHAS loci depend on 24-nt phased small interfering RNAs in maize meiotic anthers**. *New Phytol* 2021, **229**:2984–2997.

52. Carbonell A: **Plant ARGONAUTEs: Features, functions, and unknowns**. *Methods Mol Biol* 2017, **1640**:1–21.

53. Jullien PE, Schröder JA, Bonnet DMV, Pumplin N, Voinnet O: **Asymmetric expression of Argonautes in reproductive tissues**. *Plant Physiol* 2022, **188**:38–43.

54. Zhai J, Zhang H, Arikit S, Huang K, Nan G-L, Walbot V, Meyers BC: **Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phasiRNAs in maize anthers**. *Proc Natl Acad Sci U S A* 2015, **112**:3146–3151.

55. Das S, Swetha C, Pachamuthu K, Nair A, Shivaprasad PV: **Loss of function of Oryza sativa Argonaute 18 induces male sterility and reduction in phased small RNAs**. *Plant Reprod* 2020, **33**:59–73.

56. Takuno S, Ran J-H, Gaut BS: **Evolutionary patterns of genic DNA methylation vary across land plants**. *Nat Plants* 2016, **2**:15222.

57. Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, Rohr NA, Rambani A, Burke JM, Udall JA, et al.: **Widespread natural variation of DNA methylation within angiosperms**. *Genome Biol* 2016, **17**:194.

58. Lang D, Ullrich KK, Murat F, Fuchs J, Jenkins J, Haas FB, Piednoel M, Gundlach H, Van Bel M, Meyberg R, et al.: **The Physcomitrella patens chromosome-scale assembly reveals moss genome structure and evolution**. *Plant J* 2018, **93**:515–533.

59. Ikeda Y, Nishihama R, Yamaoka S, Arteaga-Vazquez MA, Aguilar-Cruz A, Grimanelli D, Pogorelcnik R, Martienssen RA, Yamato KT, Kohchi T, et al.: **Loss of CG Methylation in Marchantia polymorpha Causes Disorganization of Cell Division and Reveals Unique DNA Methylation Regulatory Mechanisms of Non-CG Methylation**. *Plant Cell Physiol* 2018, **59**:2421–2431.

60. Aguilar-Cruz A, Grimanelli D, Haseloff J, Arteaga-Vázquez MA: **DNA methylation in Marchantia polymorpha**. *New Phytol* 2019, **223**:575–581.

61. Lee EK, Cibrian-Jaramillo A, Kolokotronis S-O, Katari MS, Stamatakis A, Ott M, Chiu JC, Little DP, Stevenson DW, McCombie WR, et al.: **A functional phylogenomic view of the seed plants**. *PLoS Genet* 2011, **7**:e1002411.

62. Ma L, Hatlen A, Kelly LJ, Becher H, Wang W, Kovarik A, Leitch IJ, Leitch AR: **Angiosperms are unique among land plant lineages in the occurrence of key genes in the RNA-directed DNA methylation (RdDM) pathway**. *Genome Biol Evol* 2015, **7**:2648–2662.

63. Lunardon A, Johnson NR, Hagerott E, Phifer T, Polydore S, Coruh C, Axtell MJ: **Integrated annotations and analyses of small RNA-producing loci from 47 diverse plants**. *Genome Res* 2020, **30**:497–513.

64. Wan L-C, Wang F, Guo X, Lu S, Qiu Z, Zhao Y, Zhang H, Lin J: **Identification and characterization of small non-coding RNAs from Chinese fir by high throughput sequencing**. *BMC Plant Biol* 2012, **12**:146.

65. Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, et al.: **The Norway spruce genome sequence and conifer genome evolution**. *Nature* 2013, **497**:579–584.

66. Zhang J, Wu T, Li L, Han S, Li X, Zhang S, Qi L: **Dynamic expression of small RNA populations in larch (Larix leptolepis)**. *Planta* 2013, **237**:89–101.

67. Ausin I, Feng S, Yu C, Liu W, Kuo HY, Jacobsen EL, Zhai J, Gallego-Bartolome J, Wang L, Egertsdotter U, et al.: **DNA methylome of the 20-gigabase Norway spruce genome**. *Proc Natl Acad Sci U S A* 2016, **113**:E8106–E8113.

68. Nakamura M, Köhler C, Hennig L: **Tissue-specific transposon-associated small RNAs in the gymnosperm tree, Norway spruce**. *BMC Genomics* 2019, **20**:997.

69. Michael TP, Bryant D, Gutierrez R, Borisjuk N, Chu P, Zhang H, Xia J, Zhou J, Peng H, El Baidouri M, et al.: **Comprehensive definition of genome features in Spirodela polyrhiza by high-depth physical mapping and short-read DNA sequencing strategies**. *Plant J* 2017, **89**:617–635.

70. An D, Zhou Y, Li C, Xiao Q, Wang T, Zhang Y, Wu Y, Li Y, Chao D-Y, Messing J, et al.: **Plant evolution and environmental adaptation unveiled by long-read whole-genome sequencing of Spirodela**. *Proc Natl Acad Sci U S A* 2019, **116**:18893–18899.

71. Kawakatsu T, Huang S-SC, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y, et al.: **Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions**. *Cell* 2016, **166**:492–505.

72. Sasaki E, Kawakatsu T, Ecker JR, Nordborg M: **Common alleles of CMT2 and NRPE1 are major determinants of CHH methylation variation in Arabidopsis thaliana**. *PLoS Genet* 2019, **15**:e1008492.

73. Dubin MJ, Zhang P, Meng D, Remigereau M-S, Osborne EJ, Paolo Casale F, Drewe P, Kahles A, Jean G, Vilhjálmsson B, et al.: **DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation**. *Elife* 2015, **4**:e05255.

74. Szittya G, Silhavy D, Molnár A, Havelda Z, Lovas A, Lakatos L, Bánfalvi Z, Burgyán J: **Low temperature inhibits RNA silencing-mediated defence by the control of siRNA generation**. *EMBO J* 2003, **22**:633–640.

75. Shen X, De Jonge J, Forsberg SKG, Pettersson ME, Sheng Z, Hennig L, Carlborg Ö: **Natural CMT2 variation is associated with genome-wide methylation changes and temperature seasonality**. *PLoS Genet* 2014, **10**:e1004842.

76. Zemach A, Kim MY, Hsieh P-H, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D: **The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin**. *Cell* 2013, **153**:193–205.

77. Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE: **Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis**. *Nat Struct Mol Biol* 2014, **21**:64–72.

78. West PT, Li Q, Ji L, Eichten SR, Song J, Vaughn MW, Schmitz RJ, Springer NM: **Genomic distribution of H3K9me2 and DNA methylation in a maize genome**. *PLoS One* 2014, **9**:e105267.

79. Wang W, Haberer G, Gundlach H, Gläßer C, Nussbaumer T, Luo MC, Lomsadze A, Borodovsky M, Kerstetter RA, Shanklin J, et al.: **The Spirodela polyrhiza genome reveals**

insights into its neotenous reduction fast growth and aquatic lifestyle. *Nat Commun* 2014, **5**:3311.

80. Harkess A, Bewick AJ, Lu Z, Fourounjian P, Messing J, Michael TP, Schmitz RJ, Meyers BC: **Unusual predominance of maintenance DNA methylation in Spirodela polyrhiza**. *bioRxiv* 2020, doi:10.1101/2020.12.03.410332.