Title

Abundant expression of maternal siRNAs is a conserved feature of seed development

Journal

Proceedings of the National Academy of Sciences of the United States of America, 117(26)

ISSN

0027-8424

Authors

Grover, Jeffrey W
Burgess, Diane
Kendall, Timmy
et al.

Publication Date

2020-06-30

DOI

10.1073/pnas.2001332117

Peer reviewed

Abundant expression of maternal siRNAs is a conserved feature of seed development

Jeffrey W. Grover[a], Diane Burgess[b], Timmy Kendall[c], Abdul Baten[d,e], Suresh Pokhrel[f,g], Graham J. King[d], Blake C. Meyers[f,g], Michael Freeling[b,1], and Rebecca A. Mosher[c,h,1]

[a]Department of Molecular and Cellular Biology, The University of Arizona, Tucson, AZ 85721; [b]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720; [c]The School of Plant Sciences, The University of Arizona, Tucson, AZ 85721; [d]Southern Cross Plant Science, Southern Cross University, Lismore NSW 2480, Australia; [e]Grasslands Research Centre, AgResearch Ltd, 4410 Palmerston North, New Zealand; [f]Donald Danforth Plant Science Center, St. Louis, MO 63132; [g]Division of Plant Sciences, University of Missouri, Columbia, MO 65211; and [h]Bio5 Institute, The University of Arizona, Tucson, AZ 85721

Contributed by Michael Freeling, May 4, 2020 (sent for review January 23, 2020; reviewed by Julie A. Law and Michael Nodine)

Small RNAs are abundant in plant reproductive tissues, especially 24-nucleotide (nt) small interfering RNAs (siRNAs). Most 24-nt siRNAs are dependent on RNA Pol IV and RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) and establish DNA methylation at thousands of genomic loci in a process called RNA-directed DNA methylation (RdDM). In *Brassica rapa*, RdDM is required in the maternal sporophyte for successful seed development. Here, we demonstrate that a small number of siRNA loci account for over 90% of siRNA expression during *B. rapa* seed development. These loci exhibit unique characteristics with regard to their copy number and association with genomic features, but they resemble canonical 24-nt siRNA loci in their dependence on RNA Pol IV/RDR2 and role in RdDM. These loci are expressed in ovules before fertilization and in the seed coat, embryo, and endosperm following fertilization. We observed a similar pattern of 24-nt siRNA expression in diverse angiosperms despite rapid sequence evolution at siren loci. In the endosperm, siren siRNAs show a marked maternal bias, and siren expression in maternal sporophytic tissues is required for siren siRNA accumulation. Together, these results demonstrate that seed development occurs under the influence of abundant maternal siRNAs that might be transported to, and function in, filial tissues.

siRNA | RNA-directed DNA methylation | seed development | epigenetics

Seeds are critical for the worldwide food supply. Therefore, factors that influence seed development underpin global food security and represent important avenues for crop improvement. Seed development is complex, as seeds are composed of multiple genetically distinct tissues, including the embryo and endosperm, which are products of fertilization of the haploid egg and diploid central cell, respectively. The seed coat surrounds and protects the embryo and endosperm, is entirely maternal and sporophytic in origin, and is descended from the integument tissue surrounding the female gametes (1). Several studies indicate that 24-nucleotide (nt) small interfering RNAs (siRNAs) and RNA-directed DNA methylation (RdDM) are important during early seed development, although a biological role for RdDM during seed development has remained elusive (2–6).

A hallmark of RdDM is accumulation of 24-nt siRNAs produced by RNA Pol IV and RNA-DEPENDENT RNA POLYMERASE 2 (RDR2). RNA Pol IV and RDR2 together produce short double-stranded transcripts (7, 8). Pol IV/RDR2 products are then cleaved by DICER-LIKE 3 (DCL3) to produce 24-nt siRNA duplexes (9) that can be bound by ARGONAUTE 4 (AGO4), which removes the "passenger strand," leaving the "guide strand" bound (10, 11). The AGO4–siRNA complex interacts with the C-terminal domain of RNA Pol V and either the Pol V transcript or DNA within the Pol V transcription bubble (12–15), leading to the recruitment of the DNA methyltransferase DRM2 (DOMAINS REARRANGED METHYLTRANSFERASE) and methylation of cytosines (16). RdDM activity is associated with transcriptional silencing of transposable elements in euchromatin (17, 18) and can influence the expression of genes (19, 20).

Despite the abundance of 24-nt siRNAs in seeds, seed development proceeds normally in *Arabidopsis thaliana* lacking 24-nt siRNAs or RdDM machinery. However, loss of RdDM increases viability in paternal-excess crosses, suggesting that balancing the parental contributions of siRNAs might be important for seed development (2, 4, 5). In *Brassica rapa*, a close relative of *A. thaliana*, RdDM mutants display seed abortion ∼15 d postfertilization (19). This phenotype is controlled by maternal sporophyte genotype, further supporting the hypothesis that parental siRNA contributions are important for seed development.

Previous work has identified both maternal- and paternal-specific siRNAs in reproductive tissues. Male gametophytes (pollen grains) produce Pol IV-dependent epigenetically activated small interfering RNAs (easiRNAs) that are proposed to enter filial tissues during fertilization (4, 21, 22). However, the maternal genome is the source for the majority of 24-nt siRNAs in the developing seed, including a class of siRNAs that exclusively accumulate in reproductive tissues (23, 24). Maternal siRNA accumulation in the seed is linked to expression of developmental regulators in *A. thaliana* endosperm (3, 6). It has been hypothesized

Significance

Gene expression is regulated not only by DNA sequence, but also chemical modifications such as methylation. DNA methylation in plants is induced at thousands of genomic loci by 24-nucleotide (nt) small interfering RNAs (siRNAs). We show that during seed development, however, over 90% of 24-nt siRNAs are produced from fewer than 200 loci; this inequality of expression is conserved among plants. Rarely, chromosomal position of these loci can be conserved despite sequence evolution. siRNAs from these loci are abundant in seed coat and endosperm but not in leaves. Only maternal siRNAs are detected from these loci in the endosperm, suggesting that these siRNAs might move from maternal seed coat into filial tissues to establish DNA methylation in the next generation.

Author contributions: J.W.G., M.F., and R.A.M. designed research; J.W.G., D.B., T.K., and M.F. performed research; A.B., S.P., G.J.K., and B.C.M. contributed new reagents/analytic tools; J.W.G., D.B., M.F., and R.A.M. analyzed data; and J.W.G., D.B., M.F., and R.A.M. wrote the paper.

Reviewers: J.A.L., Salk Institute for Biological Studies; and M.N., Gregor Mendel Institute.

The authors declare no competing interest.

Data deposition: Sequencing data generated in this study have been deposited in the National Center for Biotechnology Information Sequence Read Archive under BioProject accession nos. PRJNA588293 and PRJNA592710. Custom scripts used to process data are available in the GitHub repository (https://github.com/The-Mosher-Lab/grover_et_al_sirens_2020).

[1]To whom correspondence may be addressed. Email: freeling@berkeley.edu or rmosher@email.arizona.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2001332117/-/DCSupplemental.

First published June 15, 2020.

that siRNAs are imprinted in the endosperm (6, 23, 25), although perturbation of known imprinting pathways did not change their apparent uniparental expression pattern (26), and this interpretation has been challenged (24). Rather than imprinted expression, maternal siRNAs could be loaded into developing filial tissues from the surrounding maternal sporophytic tissue, in a manner similar to maternal loading of piwi-associated RNAs (piRNAs) into *Drosophila melanogaster* eggs before fertilization (27, 28). However, such a mechanism has yet to be observed in plants.

Here, we demonstrate that the small RNA content of developing seeds is predominantly composed of a distinct category of maternally expressed siRNAs. These siRNAs accumulate at a small number of genomic loci we call "sirens," based on a category of small RNAs with characteristics similar to our own (25). Siren loci were originally described in rice endosperm (siRNA in endosperm), but we have found them to be highly expressed in other reproductive tissues. Siren siRNAs are highest expressed in the ovule before fertilization and the developing seed coat after fertilization and might move between maternal and filial tissues. Siren siRNAs trigger DNA methylation, uncovering a potential mechanism for maternal sporophytic control over seed development. Furthermore, we find siren loci in multiple species and rare conservation of siren position, but not sequence, among crucifers, indicating that siren loci are a general feature of angiosperm seed development. These results resolve conflicting reports regarding the origin of siRNAs in the seed and illuminate mechanisms for parental conflict and transgenerational inheritance.

Results

***B. rapa* Ovule siRNAs Overwhelmingly Accumulate within Siren Loci.** To better understand the role of siRNAs during seed development, we sequenced small RNAs from a variety of *B. rapa* tissues, organs, and organ systems: whole-seed samples collected throughout development (10, 14, 19, and 24 d after fertilization), the constituent parts of the seed (embryo, endosperm, and seed coat), and mature leaves (29). Together with our previously published *B. rapa* small RNA sequencing datasets and data from *B. rapa* anthers (30), we defined 84,468 small RNA loci covering a total of 102.5 Mb (29.6%) of the *B. rapa* R-o-18 genome. These loci were categorized by the most common size of small RNA reads mapping to them, and over 88% of small RNA loci contained primarily 24-nt siRNA (24 dominant) (*SI Appendix*, Fig. S1).

We began by comparing expression in cells from nonfertilized organs: ovules, which are composed almost entirely of diploid integument cells, and leaves. We examined the distribution of small RNA accumulation at all expressed small RNA loci in these two tissues and found that, despite a larger number of loci with detectable small RNA accumulation in leaves, the highest reads per kilobase mapped (RPKM) values were from ovules (Fig. 1*A*). This observation suggests that a few highly expressed loci might be obscuring representation of other loci in our ovule siRNA libraries.

To identify potential Pol IV/RdDM loci responsible for the largest share of small RNA abundance in ovules, and reduce the possibility that this shift in expression distribution was caused by highly expressed micro (mi)RNAs or other categories of small RNAs, we compared the cumulative expression distribution for 24-dominant loci in ovules vs. leaves (Fig. 1*B*). Comparing the contribution of each 24-dominant locus with the cumulative expression in leaf and ovule reveals a stark difference between the organs. Ninety percent of the expression in ovules was derived from only 191 loci (1.51% of ovule-expressed loci). In comparison, 34,424 loci (62.0%) are required to reach 90% of cumulative expression in leaves.

We next compared small RNA abundance at these highly expressed ovule loci in multiple organs (Fig. 1*C*). The 191 loci are highest expressed in ovule and seed coat. Ovules are predominantly composed of integument, which is the precursor of the seed coat, suggesting that expression of siren siRNAs begins in this maternal sporophytic tissue before fertilization and continues after fertilization. siRNAs from the 191 loci accumulate at negligible levels in leaves and anthers, demonstrating that their expression is associated with maternal reproductive structures before fertilization. However, siRNAs from the 191 loci also accumulate moderately in endosperm and to a low level in embryos and remain expressed throughout the latest stages of embryogenesis, including in mature seeds (Fig. 1*C*), indicating that they could affect targets throughout all stages of seed development.

The high expression of this small subset of ovule loci in endosperm indicates that they are similar to a class of small RNA loci identified in rice endosperm (25), termed siren loci (siRNA

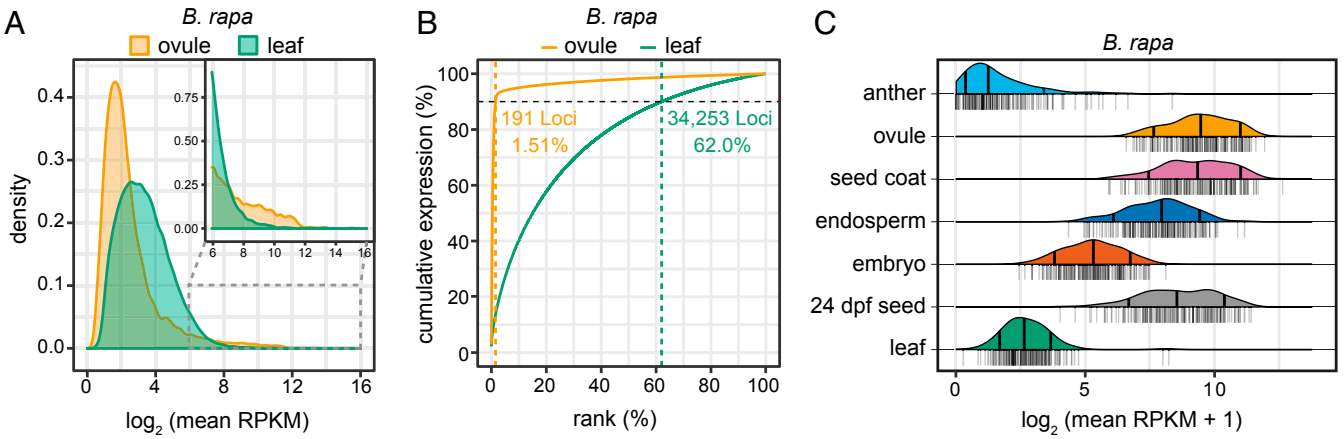

Fig. 1. Siren loci dominate the small RNA expression landscape in seeds. (*A*) Distribution of small RNA accumulation at each small RNA locus in ovule and leaf. The average of expression for loci with at least two replicates ≥2 RPKM is plotted (n = 11,149 in ovules, 61,721 in leaves). Despite the higher median expression in leaves, the highest expressed loci are found in ovules (*Inset*). (*B*) Cumulative expression plot of 24-dominant loci in ovule and leaf. Total expression at each locus was determined in reads per million (RPM) using combined replicates and ranked, and their cumulative expression was plotted. Loci were analyzed if they had expression ≥2 RPM in combined replicates (n = 12,636 in ovules, 55,228 in leaves). The 1.51% highest expressed 24-dominant loci are responsible for 90% of small RNA expression in ovules. In leaves, the highest 62% of 24-dominant loci account for 90% of expression. (*C*) Mean small RNA accumulation at the 191 siren loci in several tissues. Average expression level from replicates is shown, with individual data points plotted as a rug underneath the ridge plot. Quantile lines show 10th, 50th (median), and 90th percentiles. dpf, days post fertilization.

in endosperm). While our analysis indicates that siren loci are not specifically expressed in the endosperm, the previous study did not interrogate expression in ovules or seed coats and could not have observed accumulation of siren siRNAs in these tissues. We therefore refer to the 191 highly expressed ovule small RNA loci as siren loci to indicate their congruence with this previous work and suggest that the term siren might better indicate a prominent siRNA signal rising above all others.

***B. rapa* Siren Loci Have Unique Sequence Characteristics Compared with Most 24-Dominant siRNA Loci.** To further characterize the siren loci, we compared their size with other categories of small RNA loci. These included 24-dominant loci that were not classified as siren loci, 21- and 22-dominant loci, and 100 randomly sampled sets of 24-dominant loci ($n = 191$). The median size of siren loci was 4,238 bp and was significantly greater than all comparison groups (Fig. 2*A*). Siren loci account for 959.5 kb of genomic sequence (0.28%) and 0.94% of the total size of the small RNA-producing portion of the genome. However, we noted that siRNA accumulation in sirens was not uniform. Generally, most siRNAs mapped to a "core" region, with lower accumulation in flanking regions (*SI Appendix*, Fig. S2). These core regions are substantially smaller (Fig. 2*A*) but still account for 87% of siRNA accumulation at all sirens.

To assess the copy number of siren loci, we performed a Basic Local Alignment Search Tool (BLAST) search against the *B. rapa* genome using all 191 siren loci and the same comparison groups described above as queries. BLAST hits were collected if they were ≥432 nt (the first quartile for length of all *B. rapa* small RNA loci), were ≥50% of the query locus length, had ≥80% identity, and had an E value $\leq 1 \times 10^{-50}$. With these criteria, 90% of siren loci had only a single BLAST match compared with ∼50% of comparison groups (Fig. 2*B*). Siren loci also had fewer BLAST matches than the 191 random loci of equal sizes. However, the criteria for determining valid BLAST hits are biased against the nonsiren categories due to their smaller size. In order to address this caveat, we also determined the unique mapping rate for the reads aligning to siren loci and other categories of small RNA loci (Fig. 2*C*). Consistent with the uniqueness of siren loci determined by BLAST, the siRNAs mapping to these loci are more likely to map uniquely in the genome compared with other categories of loci. Combined, these analyses indicate that siren loci and the RNAs that accumulate at them are more often uniquely found in the genome than other categories of small RNA loci.

Next, we assessed overlap between small RNA-producing loci and a variety of genomic annotations (Table 1). Siren loci

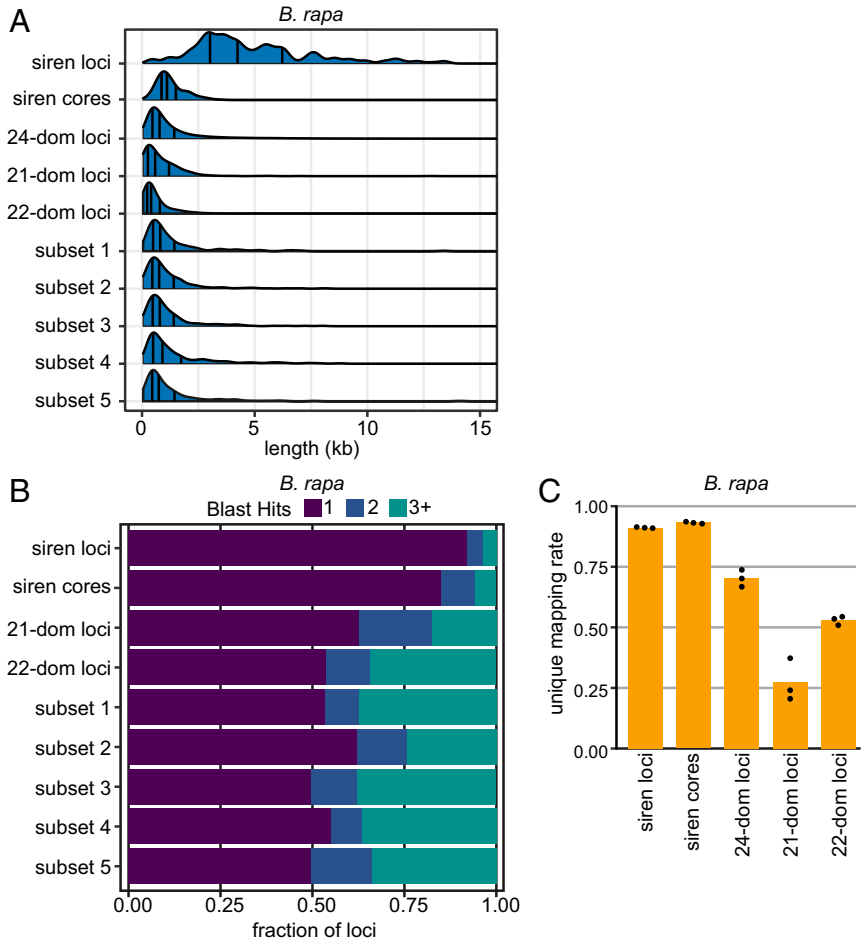

Fig. 2. Siren loci are large and unique. (*A*) Length distribution of small RNA loci. Siren loci are larger than other categories, including nonsiren 24 dominant ($n = 74{,}269$), 21 dominant ($n = 503$), 22 dominant ($n = 990$), and 100 random subsets of 24-dominant loci ($n = 191$ each, 5 shown). Quantile lines represent 10th, 50th (median), and 90th quantiles. (*B*) Valid blast hits by small RNA locus category. Most siren loci are present in single-copy genome wide, as determined by BLASTN search against the *B. rapa* R-o-18 genome (see *Methods*). (*C*) Fraction of uniquely mapping reads aggregated across all loci per category in ovule samples. Siren loci contain more uniquely mapping reads than nonsiren 24-dominant, 21-dominant, and 22-dominant loci. Bars represent the mean of three replicate libraries; individual values are shown.

Table 1. Overlap between siRNA loci and genomic features

	Genes	All transposons	Class I transposons	Class II transposons	Helitrons
24 Dominant (n = 74,269)					
Overlapping loci,* %	34.8	89.7	33.4	45.0	28.0
Fold enrichment vs. genome	**0.7**	**1.4**	**1.1**	**2**	**2.1**
Siren loci (n = 191)					
Overlapping loci,* %	75.9	96.3	34.0	67.0	62.8
Fold enrichment vs. genome	1.1	**1.1**	**0.7**	**1.5**	**2**
Fold enrichment vs. 24-nt dominant	**2.1**	1.1	1	**1.5**	**2.2**
Siren cores (n = 191)					
Overlapping loci,* %	39.8	70.7	7.85	27.7	40.8
Fold enrichment vs. genome	**0.8**	1.1	**0.3**	1.2	**2.8**
Fold enrichment vs. 24-nt dominant	1.1	**0.8**	**0.2**	**0.6**	**1.4**

Bold indicates $P < 0.0001$.
*Overlap of ≥1 nt required.

show neither depletion nor enrichment of genes compared with randomized coordinates of the same size and therefore, are enriched for genes relative to other 24-dominant siRNA loci, which show depletion for genes (Table 1). Siren loci are also enriched for DNA elements (class II transposons) and helitrons and depleted for retrotransposons (class I transposons). However, when only siren core sequences are considered, genes are depleted, and helitrons are the only transposon class that is enriched (Table 1). This further highlights the uniqueness of siren loci in the genome and a potential interaction, specifically, with helitron-type DNA transposons.

Because we categorized loci as 24 dominant when a plurality of their siRNAs were 24 nt, we also compared the fraction of RNAs between 19 and 26 nt in length that accumulated at siren loci (Fig. 3*A*). Both siren and nonsiren 24-dominant loci have a high proportion of 24-nt RNAs, whether assessed in aggregate (Fig. 3*A*) or individually (Fig. 3*B*).

Small RNAs function when bound to Argonaute proteins, which have distinct 5′ nucleotide preferences. The primary Argonaute associated with 24-nt siRNAs and RdDM is AGO4, which exhibits an adenine bias at the 5′ nucleotide (31, 32). Siren siRNAs have the same 5′ adenine bias as siRNAs from 24-dominant loci, suggesting that siren RNAs may interact with AGO4 or related Argonaute proteins (Fig. 3*C*).

The size of siren loci, their single-copy nature, and differences in associated genomic features together argue that siren loci are distinct from the majority of 24-dominant loci. However, accumulation of 24-nt siRNAs that begin with adenine suggests siren loci might interact with AGO4 and target de novo DNA methylation (10, 31).

Siren siRNAs Direct DNA Methylation. To determine the biogenesis of siren siRNAs, we compared small RNA accumulation at siren loci in *B. rapa nrpd1*, *rdr2*, and *nrpe1* mutants. To address oversampling that results from loss of most siRNAs in *nrpd1* and *rdr2* mutants, small RNA accumulation was normalized to mapped 21-nt RNAs, which are not affected by mutations in the RdDM machinery (33, 34). Median siren siRNA accumulation was reduced by ∼16-fold in *nrpd1* ovules and 8-fold in *rdr2* ovules, while *nrpe1* ovules maintain wild-type expression of siren RNA (Fig. 4*A*). This pattern of expression indicates that siren siRNAs are canonical Pol IV/RDR2 products that are not dependent on Pol V (35).

To determine, unequivocally, whether siren loci are associated with RdDM, we first assessed DNA methylation at six randomly selected siren loci using methylation-sensitive qPCR (Fig. 4*B*). This method assesses methylation of individual cytosines in the asymmetric CHH context (where H is A, T, or C), and the results are therefore indicative of de novo methylation (36, 37). The level of DNA methylation at these sites is lower in leaves compared with ovules, correlating with expression of siren siRNAs (Fig. 2*A*). At five of six loci tested, *rdr2* shows a significant reduction in DNA methylation in ovules, consistent with our expectation that siRNA accumulation enables DNA methylation at siren loci.

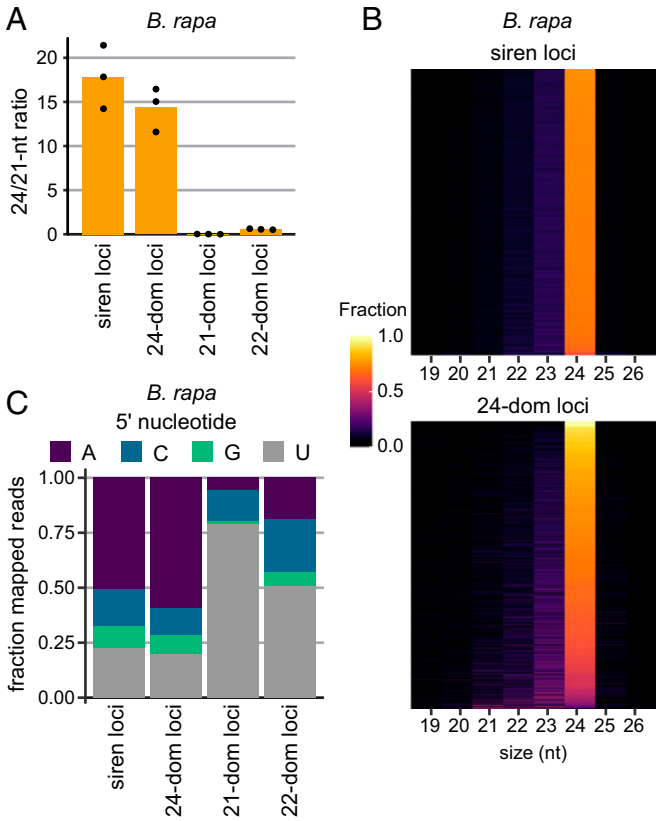

Fig. 3. Siren loci accumulate 24-nt small RNAs beginning with adenine. (*A*) The 24/21-nt ratio of small RNAs, aggregated across all loci of a category. Bars represent the mean of three ovule libraries; individual values are shown. (*B*) Heat map showing the fraction of each size class of RNA present at each locus (*y* axis). (*C*) The 5′ nucleotide bias of all RNAs in each category. *B* and *C* are from combined ovule libraries.

To investigate DNA methylation genome wide, and at all siren loci, we performed whole-genome bisulfite sequencing in wild-type and *rdr2* ovules and leaves (29). Consistent with the methylation-sensitive PCR assay, siren loci exhibit more CHH methylation in ovules than leaves, and this methylation is lost in *rdr2* ovules (Fig. 4*C*). Compared with nonsiren 24-dominant loci, siren loci are highly methylated in ovules, which might be a consequence of the substantial accumulation of siRNAs at these loci. This result, in combination with the small RNA sequencing, establishes siren loci as organ-specific sites of RdDM.

Siren Loci Are Present across Angiosperms. To determine whether siren loci are specific to *B. rapa*, or are present in more diverse species, we utilized public data from *A. thaliana* and rice (3, 25, 38–40). Using these datasets, we identified 16,422 small RNA loci in *A. thaliana* and 122,661 loci in rice. We then compared the contribution of each locus with the overall small RNA accumulation in that tissue as we previously did in *B. rapa*.

Similar to *B. rapa*, *A. thaliana* ovules were dominated by accumulation of siRNAs at a small number of loci (Fig. 5*A*). Similar profiles for the same tissue from different public datasets demonstrate that cumulative expression patterns are robust to factors such as library preparation kit, sequencing instrument, and which laboratory performed the experiment. Using the same 90% cumulative ovule expression threshold as in *B. rapa*, we identified 128 ovule siren loci in *A. thaliana* and assessed accumulation of siRNAs at each locus. As in *B. rapa*, accumulation at these loci remained high in developing seeds but was substantially reduced in leaves and in *nrpd1* and *rdr2* seeds (Fig. 5 *B* and *C*). Combined with the expression bias in ovules, this genetic and developmental expression pattern confirms that these are siren loci.

In rice endosperm, we identified a similarly skewed cumulative expression pattern (Fig. 5*D*), as we expected due to the previous observation of biased siRNA accumulation in this tissue (25). The unequal accumulation of siRNA in diverse plants indicates that extreme expression of siRNAs from a few siren loci is a conserved feature of angiosperm seed development.

Siren Loci Exhibit High Rates of Evolution. To investigate the evolutionary rate of siren loci, we searched for homology between *B. rapa* and *A. thaliana* siren loci. Because *B. rapa* experienced a

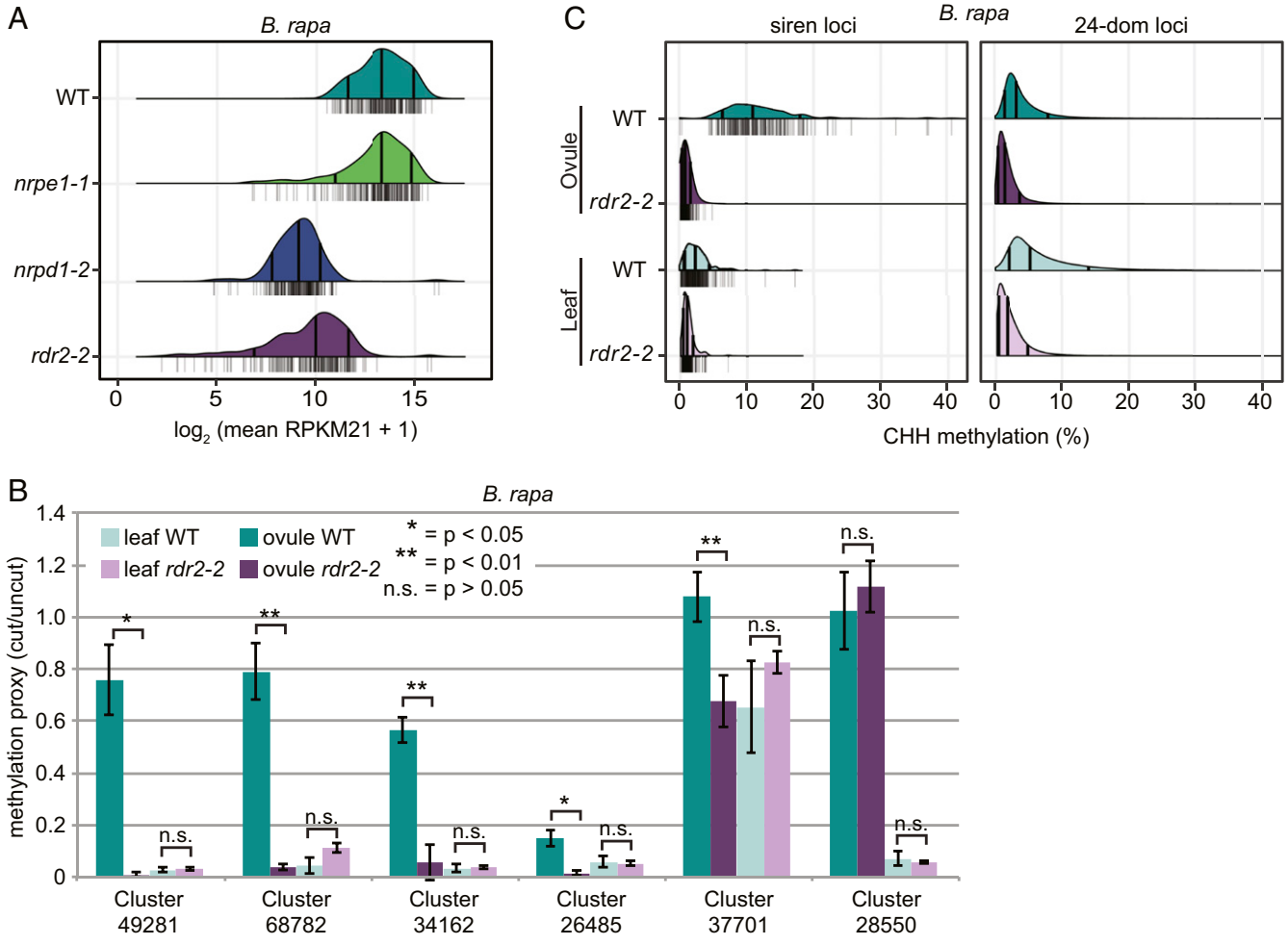

Fig. 4. Siren loci are organ-specific RdDM targets. (*A*) Small RNA accumulation at siren loci in ovules. Small RNA abundance at each locus is normalized by locus length, in kilobases, and library size, which is measured by the number of mapping 21-nt small RNAs (RPKM21). Mean of three replicates shown. Quantile lines show 10th, 50th (median), and 90th percentiles. Individual measurements are shown as the rug below each density. (*B*) Relative methylation from chop-qPCR. Bar is the mean of three replicates; error bars represent SD. Five of six tested siren loci are significantly differentially methylated at the assayed nucleotide based on a two-tailed Student's *t* test. (*C*) CHH methylation at siren loci and nonsiren 24-dominant loci. Percent methylation calculated based on methylated and unmethylated cytosine calls over each locus in the category and sample. Minimum depth of five reads required to calculate a percentage. n.s., not significant; WT, wild type.

recent whole-genome triplication ~22 Mya, *A. thaliana* loci are homologous with up to three loci in the *B. rapa* genome (41, 42). Using homology of protein-coding genes as a guide, syntenic locations in the *B. rapa* genome were identified for the 64 highest-expressed *A. thaliana* sirens. None of these locations showed sequence similarity to the *A. thaliana* siren loci. However, nine (14%) had a siren locus at one or more *B. rapa* syntenic positions (*SI Appendix*, Table S1). One *A. thaliana* siren was conserved at all three *B. rapa* orthologous positions, despite lack of nucleotide sequence conservation among these positions (Fig. 6). Seven additional *B. rapa* siren loci were conserved at a homeologous position in *B. rapa*, and with the exception of one pair, these also had no conserved sequence. These data demonstrate that over the 40 My since the divergence of *A. thaliana* and *B. rapa* from a common ancestor (42), a siren's position is only rarely conserved, and sequence is not conserved. In addition, comparison of homeologous sequences within *B. rapa* indicates that nucleotide sequence at siren loci has diverged rapidly in the ~22 My since the whole-genome triplication.

Siren RNAs Are Maternal in Origin. The primary function of RdDM is hypothesized to be silencing of transposons (16). Because the seed coat is destined for programmed cell death, and therefore control of transposons is not necessary, it is counterintuitive that siren siRNAs would express highly in this tissue. Alternatively, siRNAs might be produced within the seed coat in order to transit into the seed where they are required for development of embryo or endosperm. If siren siRNAs that accumulate in filial tissues are produced in the maternal sporophyte, they should map specifically to maternal chromosomes.

To determine the parental origin of siren RNAs, we sequenced small RNAs from reciprocal crosses between two *B. rapa* varieties, R-o-18 and R500 (29). siRNA accumulation in whole seeds from these crosses is maternally biased (19), and similar bias in whole-seed small RNAs is reported in *A. thaliana* (23). However, it is unclear whether this bias is due to high expression in the maternal seed coat or represents a widespread maternal bias. We therefore determined the parental contribution of siRNAs arising from the siren loci in isolated seed coat, endosperm, and embryo by counting reads that map specifically to only one variety and fall within a siren locus. In R-o-18 × R500 endosperm, most small RNAs map equally well to both parental genomes and are therefore uninformative. However, ~4.5% of siren RNAs mapped specifically to R-o-18, a level similar to R-o-18 × R-o-18 endosperm and significantly higher than R500 × R-o-18 endosperm (Fig. 7*A*). Similarly, ~2.3% of siren siRNAs in R500 × R-o-18 endosperm were R500 specific, a fraction that was unchanged when R-o-18 genomes were contributed from the

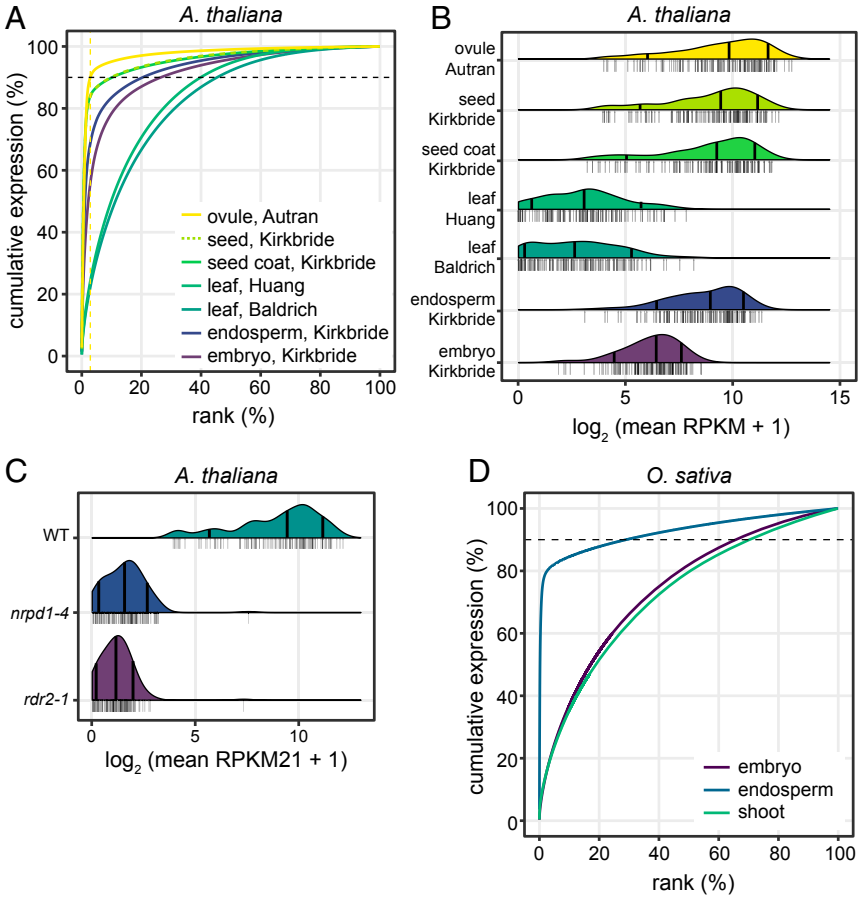

Fig. 5. Siren loci are present in diverse angiosperms. (*A*) Cumulative expression plot as in Fig. 1*B* in publicly available *A. thaliana* datasets. *A. thaliana* siren loci are defined by 90% of cumulative expression in the ovule dataset. Expressed 24-dominant loci ≥2 reads per million (RPM) in combined replicates are shown. (*B*) Ridge plot showing mean small RNA accumulation at *A. thaliana* siren loci in several tissues. (*C*) Ridge plot showing mean small RNA accumulation at *A. thaliana* siren loci in seeds from the Kirkbride et al. (3) dataset. Small RNA abundance at each locus is normalized by locus length, in kilobases, and library size, which is measured by the number of mapping 21-nt small RNAs (RPKM21). For all ridge plots, quantile lines show 10th, 50th (median), and 90th percentiles. Individual data points are the mean of available replicates, which are plotted as a rug underneath each ridge plot. (*D*) Cumulative expression plot for *O. sativa* small RNA data. Expressed 24-dominant loci ≥2 RPM in combined replicates are shown. n.s., not significant; WT, wild type.

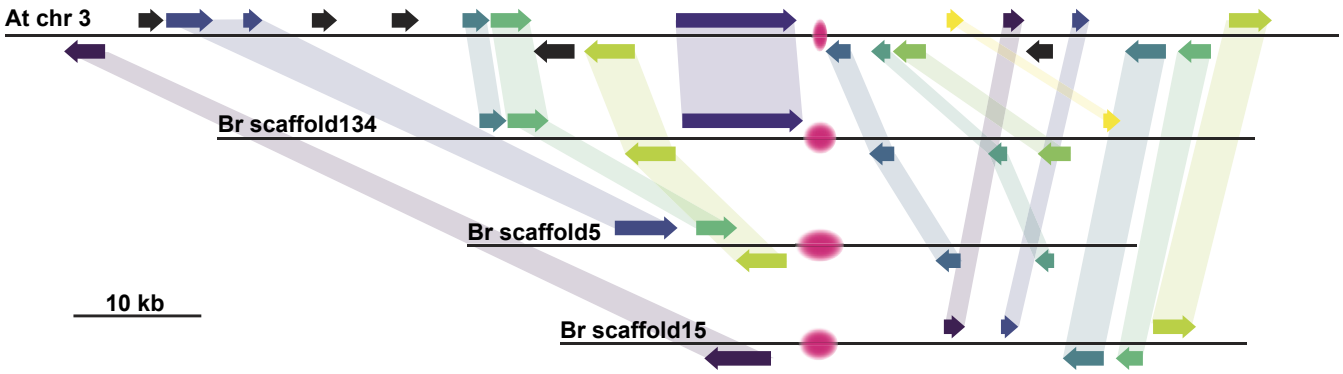

Fig. 6. Siren loci can be conserved between *B. rapa* (Br) and *A. thaliana* (At). Depiction of synteny between three homologous regions of the *B. rapa* genome and a homologous region of *A. thaliana* chromosome 3. All three homologous regions in *B. rapa* and the orthologous region of the *A. thaliana* genome contain a siren locus, depicted as pink fuzzy ovals. Homologous genes are shown with identically colored arrows and shadows.

pollen. Siren siRNAs are strongly maternal in hybrid endosperms, as the paternal genotype does not influence the percentage of maternal-specific siren siRNAs.

We further assessed the parental origin of siren siRNAs in the endosperm by measuring maternal bias at individual siren loci (Fig. 7*B*). Maternal bias in the endosperm was indistinguishable from the maternal sporophytic seed coat, suggesting either that siren loci are expressed only from matrigenic (maternally inherited) alleles in the endosperm or that siRNAs produced in the developing seed coat accumulate in the endosperm. In contrast to the marked maternal bias in endosperm, siRNAs were biallelically expressed in embryos (Fig. 7*B*), indicating that siren siRNAs are expressed from both matrigenic and patrigenic alleles in the embryo.

To reduce the possibility that maternal bias of siren siRNAs in the endosperm was due to contamination by maternal seed coat during our dissection (43), we also measured siren accumulation in publicly available *A. thaliana* libraries, including laser-capture microdissected endosperm, which was shown to have little such contamination (3). Reciprocal crosses between *nrpd1* or *rdr2* and the wild type demonstrate that siRNA accumulation at siren loci in whole seeds or seed coats is dependent on maternal alleles of *NRPD1* and *RDR2* (Fig. 7*C*). In endosperm, siRNA accumulation at siren loci was dependent only on maternal *NRPD1*, indicating that the maternal nature of siren RNAs is unlikely to be an artifact of contamination by maternal seed coats.

Together, these results suggest that, although siren loci express siRNA from both matrigenic and patrigenic alleles in the embryo, endosperm accumulates only maternal siren siRNAs. Because the majority of the siRNA accumulation in developing seeds is from siren loci in endosperm and seed coat, developing seeds are therefore dominated by maternal-specific siRNAs.

Discussion

Here, we have shown that siRNA accumulation in seeds is overwhelmingly due to siren loci, a small number of highly expressed siRNA loci. Siren siRNAs are most abundant in ovules before fertilization and comprise 90% of siRNA accumulation in that tissue. Strong expression of siren loci in maternal sporophytic tissue after fertilization explains conflicting reports regarding maternal siRNA bias in whole seeds (23, 24), as these sporophytic siRNAs mask the maternal bias from the same loci in endosperm.

An alternative explanation for the abundant siren siRNAs in ovules is that all other siRNA loci are down-regulated, leaving only siren loci remaining. Two observations argue against this hypothesis. First, abundant 24-nt siRNA accumulation is routinely detected in reproductive tissues from a variety of plants, indicating that widespread down-regulation of 24-dominant loci is not a general phenomenon (44–47). Second, siren siRNAs are abundant in *B. rapa* whole seeds at the mature green stage (Fig. 1*C*), despite the fact that these seeds are composed mostly of embryo, which expresses a more diverse population of siRNAs. The abundance of 24-nt siRNAs in reproductive tissues and accumulation of siren siRNAs in late-stage seeds argue that skewed expression patterns in ovules and seed coats result from specific up-regulation of siren loci rather than down-regulation of all other loci.

Exactly what distinguishes siren loci from most RdDM loci is not clear. While other RdDM loci are enriched for all classes of transposons, siren loci are enriched only for helitrons sequences. Unlike other RdDM loci, siren loci are specifically expressed in ovules and developing seeds and likely represent the "Type I" loci previously described as reproductive-specific in *A. thaliana* (23). How this specific and strong expression is established is unclear since the only known mechanism of Pol IV recruitment is via the chromatin reader SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1), which recognizes the combination of unmethylated Lysine 4 and methylated Lysine 9 on Histone H3 (48). However, SHH1-mediated recruitment only accounts for about half of known RdDM loci, suggesting that additional mechanisms of Pol IV recruitment remain to be identified. Other examples of organ-specific RdDM suggest that regulated expression of Pol IV activity, perhaps through tissue-specific epigenetic modification, is an underexplored developmental mechanism (40, 49).

We were surprised by our result that, occasionally, the syntenic chromosomal position of a siren in *A. thaliana* was conserved in *B. rapa* but always in the absence of sequence conservation (Fig. 6). We consider two possible explanations for this "position but not sequence conservation" result. First, siren character (i.e., abundant reproductive-specific siRNA production) might be determined by a conserved epigenetic signature, much like centromere position is determined by specific chromatin (50). Alternatively, a siren control sequence might exist linked to, but not overlapping, the siren itself. This could account for conservation of siren character without conservation of siren sequence. If this siren control sequence has a "high birth and death" rate (51), siren character would also disappear at a high rate. These two explanations are not mutually exclusive.

The siren siRNAs detected in endosperm are strongly maternally biased (Fig. 7). Although we cannot eliminate the possibility, this pattern is unlikely to result from trace seed coat contamination, given the abundant accumulation of siren siRNAs and nearly complete maternal bias. Our observation is also consistent with reports of parentally biased 24-nt siRNA accumulation in endosperm or rice and *Arabidopsis* (2, 25). This bias

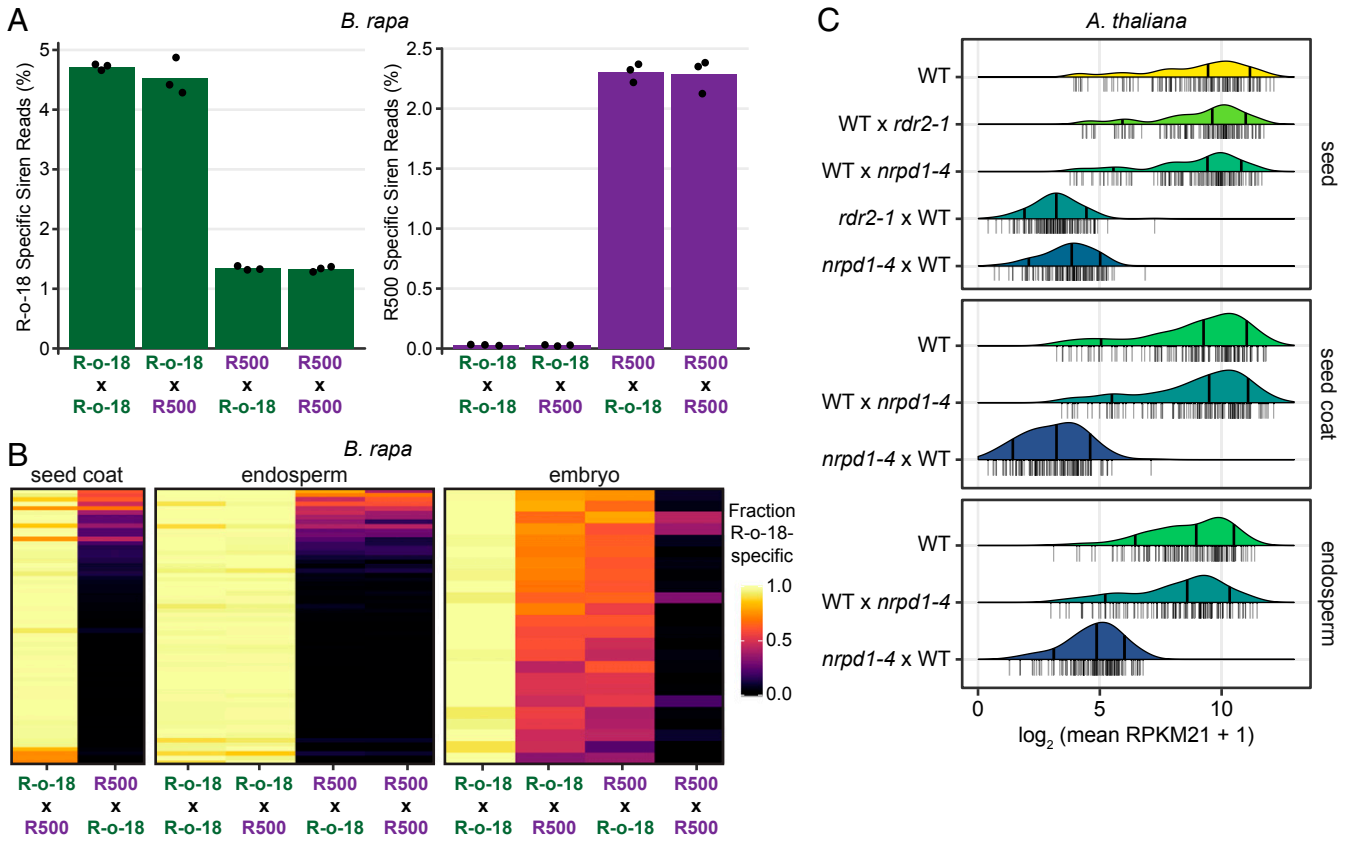

Fig. 7. Siren RNAs are maternal in origin. (*A*) Percentage R-o-18–specific and R500-specific siren reads in endosperm of R-o-18 × R500 reciprocal crosses. Bars represent mean of three replicates; individual replicates are plotted as dots. The 1% of R-o-18–specific siren reads in the R500 × R500 cross likely derive from regions of the R500 genome that produce siRNAs but are not present in the current R500 genome assembly. (*B*) Heat map showing the fraction of R-o-18–specific reads at each *B. rapa* siren locus in reciprocal R-o-18 × R500 crosses in seed coat ($n = 64$), endosperm ($n = 62$), and embryo ($n = 24$). Loci were analyzed if they had ≥10 reads across all samples in a tissue and an R-o-18 read fraction <0.7 in R500 × R500 (R500 × R-o-18 for seed coat), as R-o-18–specific reads in those samples represent reads derived from regions also present in R500 but absent from the current version of the R500 genome assembly. (*C*) Ridge plot of small RNA accumulation at *A. thaliana* siren loci in endosperm, seed coat, and whole seeds from the Kirkbride et al. (3) dataset. Small RNA abundance at each locus is normalized by locus length, in kilobases, and library size, which is measured by the number of mapping 21-nt small RNAs (RPKM21). Quantile lines show 10th, 50th (median), and 90th percentiles. Individual data points are the mean of available replicates, plotted as a rug underneath each ridge plot. WT, wild type.

might result from imprinted expression in the endosperm (matrigenic expression) or movement of siren siRNAs into the endosperm from surrounding maternal sporophytic tissue where siren siRNAs are abundantly expressed (maternal expression) (Fig. 8). In *A. thaliana*, Pol IV is required maternally for accumulation of siren siRNAs in filial tissues (Fig. 7*C*). This observation supports the maternal expression model; however, it remains possible that Pol IV is required during female gametophyte development to establish the epigenetic marks necessary for imprinted matrigenic expression from siren loci (52). How siRNAs might move between the developing seed coat and filial tissues is currently unknown.

The developmental function of siren expression is unknown, but RdDM activity at siren loci might be required for successful seed development. In *B. rapa*, loss of maternal RdDM causes severe seed development defects and high rates of seed abortion regardless of the genotype of the filial tissues (19), suggesting that siren siRNA production in the developing seed coat might be required for seed development. However, silencing by siren RNAs is not required in *A. thaliana*, as *nrpd1* mutant mothers have only subtle developmental phenotypes (3, 19). Furthermore, rapid evolution of siren loci argues against a conserved role in regulation of developmental genes. Experiments to determine whether siren siRNAs impact developmental regulators, including imprinted expression in the endosperm, are needed to ascertain any developmental role of siren expression.

The maternal bias of siren siRNAs and their production in maternal sporophytic tissues is reminiscent of piRNAs in *D. melanogaster*. In *D. melanogaster*, piRNAs are produced in ovaries before fertilization and trigger additional piRNA production in the embryo following fertilization (53). In the *Drosophila* genome, piRNA precursor transcripts arise from roughly 140 loci containing clusters of transposon fragments, and the resulting piRNAs silence transposons (53, 54). Any transposon that is active and mobile in the genome may be "trapped" by a piRNA cluster, thereby causing the production of piRNAs capable of silencing the active element. Although most siren siRNAs perfectly match only one genomic location, it is possible that siRNAs can tolerate a few mismatches at their target loci or that spreading of siren expression to neighboring transposons is a mechanism to produce siRNAs capable of suppressing mobile elements during embryogenesis.

The production of abundant 24-nt siRNAs from the integuments surrounding the maternal gametophyte is also reminiscent of phased 24-nt siRNA production from the tapetum surrounding the male gametophyte (55). Phased siRNAs do not require Pol IV and are instead produced following cleavage of a long Pol II transcript. The biological role of these phased siRNAs is unknown, but it is striking that they are produced from sporophytic

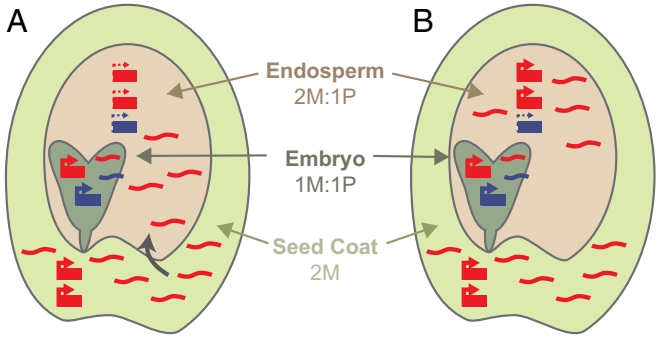

Fig. 8. Possible mechanisms for siren RNA accumulation in the seed. Siren RNAs are produced abundantly in seed coat, a maternal sporophytic tissue surrounding the endosperm and embryo, the products of fertilization. Embryos produce siren siRNA from both matrigenic (maternally inherited) and patrigenic (paternally inherited) chromosomes. Endosperm siren siRNA accumulation might occur via movement of maternally expressed siRNAs from the seed coat (*A*) or via matrigenic-specific expression in the endosperm (*B*). The M and P ratios indicate the ratio of maternal to paternal genomic content in each tissue.

tissue surrounding the developing gamete, the same pattern we see from siren siRNAs. We detect elevated phasing scores at siren loci; however, highly expressed siRNA loci can be misidentified by common phasing algorithms (56). Because Pol IV produces transcripts only slightly longer than a single siRNA (7, 8), it is unlikely that 24-nt siren RNAs are phased. Despite this difference, siren siRNAs might represent a maternal counterpoint to paternal phased 24-nt siRNAs or paternal gametophyte-derived 21/22-nt easiRNAs (4).

Siren loci are a conserved class of maternal, reproductive, small RNA loci that are distinguished by their tissue-specific expression. The existence of siren loci provides context to understand uniparentally expressed siRNAs and dynamic remodeling of the epigenome within the seed. The biological function of siren loci is currently unknown; however, the correlation between siren expression and seed development in *B. rapa* opens new avenues to investigate the role of tissue-specific RdDM in reproduction and to enhance yield of grain crops.

Methods

Plant Material and Growth Conditions. *B. rapa* plants were grown in a greenhouse at 18 °C with at least 16 h of light. Developing seeds were manually dissected at torpedo stage (17 to 20 d postfertilization) for the collection of embryo, endosperm, and seed coat samples. Ovules and anthers were collected by dissecting pistils prior to anthesis. *B. rapa* RdDM mutants were genotyped as described previously (19). RdDM mutant lines *braA.nrpd1.a-2*, *braA.rdr2.a-2*, and *braA.nrpe1.a-1* are referred to as *nrpd1-2*, *rdr2-2*, and *nrpe1-1*, respectively.

Small RNA Sequencing and Analysis. Small RNA sequencing libraries were prepared using the NEBNext Small RNA kit (NEB; E7300L) according to the manufacturer's recommendations and sequenced by either the University of Missouri DNA Core or The University of Arizona Genetics Core with either an Illumina HiSeq2500 or NextSeq500, depending on the sample (Datasets S1 and S2).

Small RNA reads were obtained from the sequencing facility and quality checked with FastQC (57). Adapters were removed using Trim Galore (58) when necessary, with options –length 10 and –quality 20. Following trimming, structural and noncoding RNAs and reads mapping to the *B. rapa*, *A. thaliana*, or *Oryza sativa* chloroplast and mitochondria genomes were removed by aligning to the rfam database v14 (59, 60), excluding miRNAs and miRNA precursors, and the respective species' chloroplast and mitochondrial genomes using Bowtie (61). Following these steps, reads shorter than 19 nt and longer than 26 nt were removed with a Python script (fastq_length_filter.py). The remaining reads were aligned to the *B. rapa* R-o-18 (v2; provided by A.B. and G.J.K.), TAIR10 (62), or MSU7 (63) genomes, depending on species, with Bowtie, and genome-mapping reads were retained. Options for all Bowtie steps were -v 0, -m 50, –best, -a, and –nomaqround, –norc was used for rfam filtering only. Detailed information on all data produced in this study, as well as previously available data, is available in Datasets S1–S3.

Small RNA loci were annotated with ShortStack (64, 65) using filtered reads from all libraries. Options used for ShortStack were –mismatches 0, –mmap u, –mincov 0.5 rpm, and –pad 75. Replicates were checked for consistency by principal component analysis (*SI Appendix*, Fig. S3) on vst-transformed counts using R and the DESeq2 package (66) and by comparison of the size profiles of mapped small RNAs between replicates. A SnakeMake workflow is available for the small RNA analysis pipeline (67).

Expression comparisons between wild-type samples were conducted by converting small RNA counts in each locus, per sample, into RPKM using the number of genome-mapping filtered reads and the size of the ShortStack predicted locus. Comparisons between the wild type and mutants were normalized per million mapped 21-nt reads, rather than all reads, to address oversampling of non–24-nt reads in RdDM mutants.

Siren cores were identified by repeating ShortStack's clustering step using –pad 1 and –mincov 2rpm with only the wild-type R-o-18 ovule small RNA alignments. The resulting small RNA loci were overlapped with identified siren loci using bedtools intersect, and the overlapping locus that accounted for the largest share of each siren's expression was taken as that siren's core region.

Enrichment for genomic features within small RNA loci was determined using BEDTools intersect (68). Features were considered to overlap a locus if there was at least 1 nt shared. Shuffling of locus coordinates was performed using BEDTools shuffle, and random 24-mer loci were selected using R software. Fold enrichment was calculated compared with the intersecting features from the random sets of loci, and *Z* scores were calculated, where $Z = (\text{observed} - \text{mean}_{\text{random}})/\text{SD}_{\text{random}}$. *Z* scores were converted to *P* values using R.

Parent-specific small RNAs were determined by first aligning small RNA reads to a concatenated R-o-18 + R500 genome (v1.4; provided by Kathleen Greenham and C. Robertson McClung, Dartmouth University, Dartmouth, MA) with bowtie to calculate the number of all mapped reads. Then, uniquely aligning (and hence, parent-specific) reads were remapped with ShortStack. ShortStack's –locifile option was used to tabulate small RNA counts at previously defined small RNA loci in each sample.

We identified homologous siren loci in the R500 genome by BLAST search, retaining the longest BLAST hit with the highest percent identity. Using this strategy, we were able to identify a homologous R500 siren locus for all 191 R-o-18 siren loci.

Whole-Genome Bisulfite Sequencing and Analysis. Genomic DNA was extracted with the GeneJET Plant Genomic DNA Purification Kit (Thermo Fisher Scientific; K0791), and whole-genome bisulfite libraries were prepared as previously described (69). Unmethylated lambda phage DNA (Promega; D1521) was included in the library as a bisulfite conversion control. Paired end libraries were sequenced at The University of Arizona Genetics Core on an Illumina NextSeq500.

Reads were checked for quality with FastQC and trimmed with Trim Galore using options –trim-n and –quality 20. Trimmed reads were aligned to the *B. rapa* R-o-18 genome (v2; provided by A.B. and G.J.K.) with bwa-meth (70). Picard Tools (71) were used to mark PCR duplicates, and the properly paired alignment rate was determined with samtools (72) using options -q 10, -c, -F 3840, -f 66. Genomic coverage was determined using mosdepth (73) with options -x and -Q 10 and a custom Python script (bed_coverage_to_x_coverage.py). Per-cytosine methylation was extracted with MethylDackel (74) in two passes, first to determine the inclusion bounds based on methylation bias per read position using MethylDackel mbias and then with MethylDackel extract. MethylDackel options were –CHG, –CHH, and –methylKit. Bisulfite conversion rates (Dataset S4) were determined by alignment to the bacteriophage lambda (NCBI GenBank accession no. J02459.1) and *B. rapa* var. *pekinesis* chloroplast (NCBI GenBank accession no. NC_015139.1) genomes and a custom Python script (bedgraph_bisulfite_conv_calc.py). All conversion rates were greater than 99%. Methylation over siren loci was determined with BEDTools intersect and a custom Python script (bedgraph_methylation_by_bed.py). Replicates were checked for consistency by plotting aggregate genome-wide methylation levels in each sequence context (*SI Appendix*, Fig. S4).

A snakemake workflow for the bisulfite sequencing analysis pipeline is available (75).

Methylation-Sensitive qPCR. Genomic DNA was extracted from *B. rapa* ovules and leaves with the GeneJET Plant Genomic DNA Purification Kit (Thermo Fisher Scientific; K0791). DNA quality and concentration were determined

on a Nanodrop instrument before diluting samples to equal concentrations. For each locus tested, 100 ng of genomic DNA was incubated with its respective restriction enzyme. qPCR was performed with 4 μL of digest product on a Bio-Rad MyIQ2. Cycling conditions were 10 min at 95 °C (enzyme activation) followed by 40 cycles of 95 °C for 10 s, 60 °C for 30s, and 72 °C for 30 s and data collection. Specificity of each reaction was confirmed by the presence of a single-peak melting temperature on the derivative melting curve plot and the presence of the expected size band through gel electrophoresis of PCR products. The resulting data were analyzed relative to an undigested DNA control prepared in parallel for each locus. Primers and restriction enzymes for each siren locus tested are listed in Dataset S5.

Evolutionary Analysis of Small RNA Loci. A list of potential *B. rapa* syntenic orthologs of *A. thaliana* was made from the combined outputs of Synfind (76) and Synmap (77) from CoGe (https://genomevolution.org/coge/). Synfind was run under standard conditions (40-gene window, minimum of 4 anchoring genes). Synmap was run using Last (-D 25, -A 4), and syntenic blocks were merged with Quota Align using a 1:3 ratio of coverage depth and an overlap distance of 40. The initial list included erroneous pairs resulting from pairing with incomplete genes or from paralogous pairs.

To remove spurious orthologs, the coverage and percent nucleotide identity for each pair of coding sequences were determined using strand-specific lastz, and putative orthologs were retained if they had an overall coverage of at least 80% and identity of at least 70%. If a homologous *Arabidopsis* gene exists, it is possible that the putative orthologous pair actually has a paralogous relationship. *A. thaliana–B. rapa* pairs were therefore removed when the lastz percentage identity, and coverage scores were higher for the homologous *Arabidopsis* gene (78). *B. rapa* genes were also removed when their percentage identity fell below a certain threshold relative to the percentage identity of the highest-scoring rapa ortholog in each group ($<\max - (0.1 \times \max)$). Tandems (orthologs identified within a 10-gene window on either side) were collapsed, keeping the tandem with the highest percentage id. Since a maximum of three orthologs would be expected from the triplicated *B. rapa* genome, only the three highest-scoring genes were kept.

B. rapa orthologs were sorted to subgenome as described previously (79). In the few cases where two orthologs were assigned to the same subgenome, the out-of-order ortholog was removed. Out-of-order genes in a subgenome were also removed when all 4 genes on either side were in order (within 30 genes of each other), but the gene was farther than 60 genes away from any of these genes.

Sirens were assigned to the closest *B. rapa* or *A. thaliana* genes using bedtools closest (68). Orthologous regions with *B. rapa* and *A. thaliana* sirens were manually inspected on CoGe GEvo panels for sirens lying between the same flanking *B. rapa–A. thaliana* syntenic gene pair using genomes annotated with sirens. In particularly TE-rich regions, it was not always possible to identify one or more *B. rapa* syntenic regions.

Data Availability. Sequencing data generated in this study have been deposited in the NCBI SRA under BioProject accession number PRJNA588293. A full listing of SRA accessions by sample is available in Datasets S1–S4. Custom scripts used to process data are available in the GitHub repository (https://github.com/The-Mosher-Lab/grover_et_al_sirens_2020). Sequences of siren loci are available in fasta format (Dataset S6).

ACKNOWLEDGMENTS. We thank Dr. C. Robertson McClung and Dr. Kathleen Greenham from Dartmouth College for sharing the *B. rapa* R500 genome. We also thank Dr. Robert Schmitz and his laboratory at the University of Georgia for technical advice regarding whole-genome bisulfite library preparation. Additionally, we thank The University of Arizona Genetics Core for sequencing the samples analyzed in this study and Maya Bose at The University of Arizona for assistance in developing the small RNA analysis pipeline. This work was supported by NSF Grant IOS-1546825. J.W.G. was supported in part by NIH Grant T32-GM008659. M.F. is grateful to the US Department of Agriculture for support of the Rausser College of Natural Resources, University of California, Berkeley.

1. R. Yadegari, G. N. Drews, Female gametophyte development. *Plant Cell* **16**, S133–S141 (2004).
2. R. M. Erdmann, P. R. V. Satyaki, M. Klosinska, M. Gehring, A small RNA pathway mediates allelic dosage in endosperm. *Cell Rep.* **21**, 3364–3372 (2017).
3. R. C. Kirkbride *et al*., Maternal small RNAs mediate spatial-temporal regulation of gene expression, imprinting, and seed development in *Arabidopsis*.*Proc. Natl. Acad. Sci. U.S.A.* **116**, 2761–2766 (2019).
4. G. Martinez *et al*., Paternal easiRNAs regulate parental genome dosage in *Arabidopsis*. *Nat. Genet.* **50**, 193–198 (2018).
5. P. R. V. Satyaki, M. Gehring, Paternally acting canonical RNA-directed DNA methylation pathway genes sensitize *Arabidopsis* endosperm to paternal genome dosage. *Plant Cell* **31**, 1563–1578 (2019).
6. J. Lu, C. Zhang, D. C. Baulcombe, Z. J. Chen, Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of *Arabidopsis* seeds. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5529–5534 (2012).
7. T. Blevins *et al*., Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in *Arabidopsis*. *eLife* **4**, e09591 (2015).
8. J. Zhai *et al*., A one precursor one siRNA model for pol IV-dependent siRNA biogenesis. *Cell* **163**, 445–455 (2015).
9. J. Singh, V. Mishra, F. Wang, H.-Y. Huang, C. S. Pikaard, Reaction mechanisms of pol IV, RDR2, and DCL3 drive RNA channeling in the siRNA-directed DNA methylation pathway. *Mol. Cell* **75**, 576–589.e5 (2019).
10. E. R. Havecker *et al*., The *Arabidopsis* RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell* **22**, 321–334 (2010).
11. F. Wang, M. J. Axtell, AGO4 is specifically required for heterochromatic siRNA accumulation at Pol V-dependent loci in *Arabidopsis thaliana*. *Plant J.* **90**, 37–47 (2017).
12. M. El-Shami *et al*., Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev.* **21**, 2539–2544 (2007).
13. S. Lahmy *et al*., Evidence for ARGONAUTE4-DNA interactions in RNA-directed DNA methylation in plants. *Genes Dev.* **30**, 2565–2570 (2016).
14. C. F. Li *et al*., An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell* **126**, 93–106 (2006).
15. W. Liu *et al*., RNA-directed DNA methylation involves co-transcriptional small-RNA-guided slicing of polymerase V transcripts in *Arabidopsis*. *Nat. Plants* **4**, 181–188 (2018).
16. M. A. Matzke, R. A. Mosher, RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15**, 394–408 (2014).
17. H. Stroud, M. V. C. Greenberg, S. Feng, Y. V. Bernatavichute, S. E. Jacobsen, Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152**, 352–364 (2013).
18. A. Zemach *et al*., The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**, 193–205 (2013).
19. J. W. Grover *et al*., Maternal components of RNA-directed DNA methylation are required for seed development in *Brassica rapa*. *Plant J.* **94**, 575–582 (2018).
20. D. Pontier *et al*., Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev.* **19**, 2030–2040 (2005).
21. G. Martínez, K. Panda, C. Köhler, R. K. Slotkin, Silencing in sperm cells is directed by RNA movement from the surrounding nurse cell. *Nat. Plants* **2**, 16030 (2016).
22. R. K. Slotkin *et al*., Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**, 461–472 (2009).
23. R. A. Mosher *et al*., Uniparental expression of PolIV-dependent siRNAs in developing endosperm of *Arabidopsis*. *Nature* **460**, 283–286 (2009).
24. D. Pignatta *et al*., Natural epigenetic polymorphisms lead to intraspecific variation in *Arabidopsis* gene imprinting. *eLife* **3**, e03198 (2014).
25. J. A. Rodrigues *et al*., Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7934–7939 (2013).
26. R. A. Mosher *et al*., An atypical epigenetic mechanism affects uniparental expression of Pol IV-dependent siRNAs. *PLoS One* **6**, e25756 (2011).
27. J. Brennecke *et al*., An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**, 1387–1392 (2008).
28. C. Klattenhoff, W. Theurkauf, Biogenesis and germline functions of piRNAs. *Development* **135**, 3–9 (2008).
29. J. W. Grover, R. A. Mosher, *Brassica rapa* seed epigenome profiling. National Center for Biotechnology Information Sequence Read Archive. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA588293. Deposited 7 November 2019.
30. S. Pokhrel, B. C. Meyers, Deep small RNA profiling of anthers of Brassica rapa R-o-18 and rdr2 mutant (field mustard). National Center for Biotechnology Information Sequence Read Archive. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA592710. Deposited 30 November 2019.
31. S. Mi *et al*., Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5′ terminal nucleotide. *Cell* **133**, 116–127 (2008).
32. F. Frank, J. Hauver, N. Sonenberg, B. Nagar, *Arabidopsis* Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. *EMBO J.* **31**, 3588–3595 (2012).
33. A. J. Herr, M. B. Jensen, T. Dalmay, D. C. Baulcombe, RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**, 118–120 (2005).
34. K. D. Kasschau *et al*., Genome-wide profiling and analysis of *Arabidopsis* siRNAs. *PLoS Biol.* **5**, 0479–0493 (2007).
35. R. A. Mosher, F. Schwach, D. Studholme, D. C. Baulcombe, PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3145–3150 (2008).
36. J. A. Law, S. E. Jacobsen, Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220 (2010).
37. H. Zhang *et al*., Protocol: A beginner's guide to the analysis of RNA-directed DNA methylation in plants. *Plant Methods* **10**, 18 (2014).

38. D. Autran *et al*., Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. *Cell* **145**, 707–719 (2011).
39. P. Baldrich *et al*., Plant extracellular vesicles contain diverse small RNA species and are enriched in 10- to 17-nucleotide "Tiny" RNAs. *Plant Cell* **31**, 315–324 (2019).
40. J. Huang *et al*., Meiocyte-specific and AtSPO11-1-dependent small RNAs and their association with meiotic gene expression and recombination. *Plant Cell* **31**, 444–464 (2019).
41. X. Wang *et al*.; *Brassica rapa* Genome Sequencing Project Consortium, The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* **43**, 1035–1039 (2011).
42. M. A. Beilstein, N. S. Nagalingum, M. D. Clements, S. R. Manchester, S. Mathews, Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18724–18728 (2010).
43. M. A. Schon, M. D. Nodine, Widespread contamination of arabidopsis embryo and endosperm transcriptome data sets. *Plant Cell* **29**, 608–617 (2017).
44. Y. Huang, T. Kendall, R. A. Mosher, Pol IV-dependent siRNA production is reduced in *Brassica rapa*. *Biology (Basel)* **2**, 1210–1223 (2013).
45. Y. Liu *et al*., Genome-wide comparison of microRNAs and their targeted transcripts among leaf, flower and fruit of sweet orange. *BMC Genomics* **15**, 695 (2014).
46. H. Hu *et al*., The complexity of posttranscriptional small RNA regulatory networks revealed by in silico analysis of *Gossypium arboreum* L. leaf, flower and boll small regulatory RNAs. *PLoS One* **10**, e0127468 (2015).
47. B. Nystedt *et al*., The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**, 579–584 (2013).
48. J. A. Law *et al*., Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498**, 385–389 (2013).
49. J. Walker *et al*., Sexual-lineage-specific DNA methylation regulates meiosis in *Arabidopsis*. *Nat. Genet.* **50**, 130–137 (2018).
50. M. J. Mendiburo, J. Padeken, S. Fülöp, A. Schepers, P. Heun, *Drosophila* CENH3 is sufficient for centromere formation. *Science* **334**, 686–690 (2011).
51. J. Nam *et al*., Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1910–1915 (2004).
52. R. A. Mosher, Maternal control of Pol IV-dependent siRNAs in *Arabidopsis* endosperm. *New Phytol.* **186**, 358–364 (2010).
53. Y. W. Iwasaki, M. C. Siomi, H. Siomi, PIWI-interacting RNA: Its biogenesis and functions. *Annu. Rev. Biochem.* **84**, 405–433 (2015).
54. S. Hirakata, M. C. Siomi, piRNA biogenesis in the germline: From transcription of piRNA genomic sources to piRNA maturation. *Biochim. Biophys. Acta* **1859**, 82–92 (2016).
55. S. Ono *et al*., EAT1 transcription factor, a non-cell-autonomous regulator of pollen production, activates meiotic small RNA biogenesis in rice anther tapetum. *PLoS Genet.* **14**, e1007238 (2018).
56. S. Polydore, A. Lunardon, M. J. Axtell, Several phased siRNA annotation methods can frequently misidentify 24 nucleotide siRNA-dominated *PHAS* loci. *Plant Direct* **2**, e00101 (2018).
57. S. Anders, FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 1 June 2020.
58. F. Krueger, Trim Galore. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/. Accessed 1 June 2020.
59. I. Kalvari *et al*., Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* **46**, D335–D342 (2018).
60. I. Kalvari *et al*., Non-coding RNA analysis using the rfam database. *Curr. Protoc. Bioinformatics* **62**, e51 (2018).
61. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
62. T. Z. Berardini *et al*., The *Arabidopsis* information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* **53**, 474–485 (2015).
63. Y. Kawahara *et al*., Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice (N. Y.)* **6**, 4 (2013).
64. M. J. Axtell, ShortStack: Comprehensive annotation and quantification of small RNA genes. *RNA* **19**, 740–751 (2013).
65. N. R. Johnson, J. M. Yeoh, C. Coruh, M. J. Axtell, Improved placement of multi-mapping small RNAs. *G3 (Bethesda)* **6**, 2103–2111 (2016).
66. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
67. M. Bose, J. W. Grover, sRNA Snakemake Workflow. https://doi.org/10.5281/zenodo.3539531. Accessed 1 June 2020.
68. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
69. M. A. Urich, J. R. Nery, R. Lister, R. J. Schmitz, J. R. Ecker, MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat. Protoc.* **10**, 475–483 (2015).
70. B. S. Pedersen, K. Eyring, S. De, I. V Yang, D. A. Schwartz, Fast and accurate alignment of long bisulfite-seq reads. arXiv:1401.1129 (6 January 2014).
71. Broad Institute, Picard Tools. http://broadinstitute.github.io/picard/. Accessed 1 June 2020.
72. H. Li *et al*.; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
73. B. S. Pedersen, A. R. Quinlan, Mosdepth: Quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).
74. D. Ryan, MethylDackel. https://github.com/dpryan79/MethylDackel. Accessed 1 June 2020.
75. J. W. Grover, WGBS Snakemake Workflow. https://doi.org/10.5281/zenodo.3539516. Accessed 1 June 2020.
76. H. Tang *et al*., SynFind: Compiling syntenic regions across any set of genomes on demand. *Genome Biol. Evol.* **7**, 3286–3298 (2015).
77. E. Lyons, B. Pedersen, J. Kane, M. Freeling, The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the rosids. *Trop. Plant Biol.* **1**, 181–190 (2008).
78. B. C. Thomas, B. Pedersen, M. Freeling, Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res.* **16**, 934–946 (2006).
79. F. Cheng *et al*., Deciphering the diploid ancestral genome of the Mesohexaploid *Brassica rapa*. *Plant Cell* **25**, 1541–1554 (2013).