

Article

## Pol IV-Dependent siRNA Production is Reduced in *Brassica rapa*

Yi Huang<sup>1</sup>, Timmy Kendall<sup>1</sup> and Rebecca A. Mosher<sup>1,2,\*</sup>

<sup>1</sup> School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA;  
E-Mails: yih@email.arizona.edu (Y.H.); kendallt@email.arizona.edu (T.K.)

<sup>2</sup> Bio5 Institute, University of Arizona, Tucson, AZ 85721, USA

\* Author to whom correspondence should be addressed; E-Mail: rmosher@email.arizona.edu;  
Tel.: +1-520-626-4185; Fax: +1-520-621-7186.

Received: 27 August 2013; in revised form: 24 September 2013 / Accepted: 25 September 2013 /  
Published: 30 September 2013

---

**Abstract:** Plants produce a diverse array of small RNA molecules capable of gene regulation, including Pol IV-dependent short interfering (p4-si)RNAs that trigger transcriptional gene silencing. Small RNA transcriptomes are available for many plant species, but mutations affecting the synthesis of Pol IV-dependent siRNAs are characterized only in *Arabidopsis* and maize, leading to assumptions regarding nature of p4-siRNAs in all other species. We have identified a mutation in the largest subunit of Pol IV, *NRPDI*, that impacts Pol IV activity in *Brassica rapa*, an agriculturally important relative of the reference plant *Arabidopsis*. Using this mutation we characterized the Pol IV-dependent and Pol IV-independent small RNA populations in *B. rapa*. In addition, our analysis demonstrates reduced production of p4-siRNAs in *B. rapa* relative to *Arabidopsis*. *B. rapa* genomic regions are less likely to generate p4-siRNAs than *Arabidopsis* but more likely to generate Pol IV-independent siRNAs, including 24 nt RNAs mapping to transposable elements. These observations underscore the diversity of small RNAs produced by plants and highlight the importance of genetic studies during small RNA analysis.

**Keywords:** *Brassica rapa*; RNA Polymerase IV; RNA-directed DNA methylation; siRNA

---

## 1. Introduction

Most eukaryotes produce an array of small (20–30 nt) RNA molecules (sRNA) capable of gene regulation. In plants the primary classes of sRNAs include 21 nt micro (mi)RNAs and 24 nt short interfering (si)RNAs. MicroRNAs regulate gene expression post-transcriptionally through mRNA degradation or sequestration [1] while 24 nt siRNAs affect gene expression transcriptionally through chromatin modifications that inhibit Pol II transcription or through co-transcriptional degradation of nascent transcripts [2].

In the reference plant *Arabidopsis thaliana*, biosynthesis of 24 nt siRNAs is initiated by the plant-specific DNA-dependent RNA polymerase Pol IV [3], which generates a single-stranded RNA template for the RNA-dependent RNA polymerase RDR2 [4]. Double-stranded (ds)RNA produced by RDR2 is cleaved by the DICER-LIKE endonuclease DCL3 into 24 nt duplexes [5]. These duplexes are integrated into ARGONAUTE (AGO) proteins, which are recruited to specific genomic regions through Watson-Crick base pairing between the p4-siRNA and nascent transcripts generated by a second plant-specific DNA-dependent RNA polymerase, Pol V [6]. AGO/p4-siRNA/Pol V complexes enlists a number of proteins, including the DNA methyltransferase DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), to initiate RNA-directed DNA methylation and transcriptional silencing [7].

Pol IV-dependent (p4-)siRNAs are the largest class of *Arabidopsis* sRNAs—comprising 75% of the mass and over 99% of the complexity (unique sRNA sequences) in the sRNA transcriptome. P4-siRNAs are produced from thousands of genomic loci, compared to hundreds for all other classes of sRNA combined [3,8,9]. Most p4-siRNAs are produced from repetitive genomic elements [3,8,9] including transposable elements (TEs) [5,10–13]. P4-siRNAs are also required for *de novo* methylation of incoming transgenes [14] or new retrotransposon insertions [15,16], leading to the hypothesis that p4-siRNAs defend the genome against mobile genetic elements [2,17,18].

Small RNA sequencing from many species [19–26] indicates that 24 nt sRNAs are the most abundant class of sRNA in flowering plants, and these sRNAs frequently match TE sequences. However, because DCL3 also acts on dsRNA generated from Pol II-derived RNA hairpins [9,27–29] and dsRNA from other pathways [16], it is not possible to determine if these 24 nt sRNAs are Pol IV-dependent without loss-of-function Pol IV mutants. Currently, mutations affecting Pol IV activity are characterized only in *Arabidopsis* and maize. In maize, mutation of Pol IV and downstream components causes loss of 24 nt siRNAs [19,30,31]. However unlike *Arabidopsis*, 22 nt siRNAs of unknown function continue to be produced from many TE sequences when the maize Pol IV pathway is disrupted [19], suggesting that multiple sRNA pathways target TEs. Maize and *Arabidopsis* also differ in the developmental phenotypes associated with loss of Pol IV activity. Maize mutants display abnormal morphology [30] while *Arabidopsis* mutants are indistinguishable from wild type. With Pol IV mutants lacking in other species, generalizations regarding the synthesis of 24 nt siRNAs and the control of TEs cannot be drawn.

*Arabidopsis* is a member of the *Brassicaceae* family, which also includes important oil and vegetable species in the *Brassica* genus such as *Brassica oleracea* (broccoli, cauliflower, cabbage, kohlrabi, kale, and brussels sprouts), *Brassica rapa* (turnip, pak choi, Chinese cabbage, and yellow sarson), *Brassica juncea* (mustard), and *Brassica napus* (canola) [32]. Recently a number of genetic and

genomic tools were created to study these important species. In particular, the ~280 Mb genome of *B. rapa* was sequenced [33] and a large *B. rapa* TILLING (Targeting Induced Local Lesions IN Genomes) population created [34]. Within this extensively-mutated population of nearly 10,000 lines, approximately half of the GC base-pairs in the genome are mutated, making possible the selection of both loss-of-function and hypomorphic alleles. Small RNA transcriptomes were analyzed in *Brassica napus*, but Pol IV-dependent siRNAs could not be identified [26]. To further understand the role of Pol IV in production of siRNAs, we characterized sRNA populations from *B. rapa* mutants lacking functional Pol IV.

## 2. Experimental Section

*B. rapa* orthologs were identified through BLASTP search of the Brassica database [35] using *Arabidopsis* protein sequences (TAIR version 10) and confirmed through synteny analysis. The target region was selected based on output of the CoDDLe program (Codons to Optimized Discovery of Deleterious LESions) and screening of the TILLING population occurred at RevGenUK (Norwich, UK). Putative *brnrpd1* mutations were confirmed and homozygous mutants selected from segregating populations using CAPS/dCAPS markers (see Supplementary Figure S1). For *brnrpd1-1*, a PCR product was amplified with BrNRPD1.F1 (AAACGGGAGACAGCTTCTTACG) and BrNRPD1.R2 (TCAGAGGGTA-AGCGTACTCG) and cleaved with MboI resulting in 400 and 100 nt bands in wild type and 250, 150, and 100 nt bands in *brnrpd1-1*. For *brnrpd1-2*, a PCR product was amplified with BrNRPD1.F1 and BrNRPD1.R1 (AGGAACGTACCCGTGAAGACAGACT) and cleaved with HinF1, resulting in 175 and 25 nt bands in wild type and an uncleaved 200 nt band in *brnrpd1-2* (see Supplementary Figure S1).

Total nucleic acid from a mixture of floral and young silique tissue was prepared as described previously [36] and checked for integrity on a 1% agarose gel. Genomic DNA was removed with Ambion DNA-free according to manufacturer's instructions before submission to the University of Missouri Genomics Core for library preparation and sequencing.

After trimming adapters and parsing multiplexed data, sRNA reads were filtered to retain high quality reads of appropriate size (19–26 nt) and to remove reads matching the sense strand of structural RNAs (rRNA, tRNA, snRNA, and snoRNA). Reads were then aligned to the *B. rapa* genome using Bowtie (version 0.12.7) [37] and only reads with a perfect genomic match were retained for analysis. Because miRNAs and other 21 nt sRNA species are unlikely to be affected by mutation in p4-siRNA biosynthesis [3,5,8], size profiles of libraries were normalized to filtered, genome-matching 21 nt reads. For all other analyses, libraries were normalized to all filtered, genome-matching reads. Annotation of the *B. rapa* genome was as in [33]. Z scores were calculated by collecting annotation at randomly selected windows from the *B. rapa* genome (100 iterations).

## 3. Results and Discussion

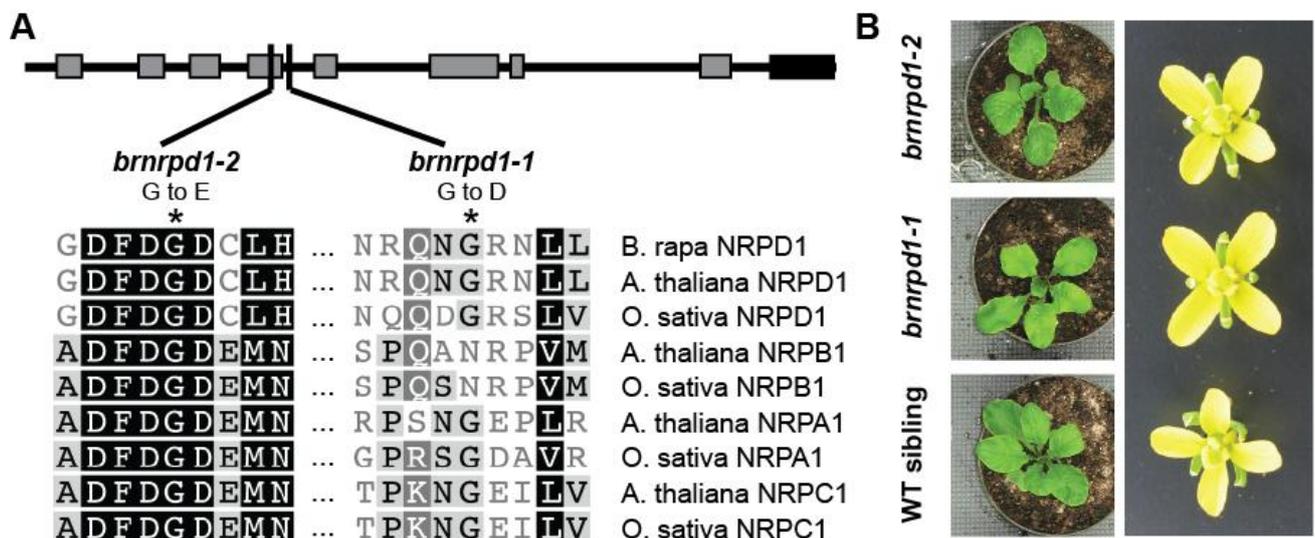
### 3.1. The Pol IV Pathway in Brassica Rapa

To analyze the synthesis of sRNAs in *Brassica* species, mutations were sought in p4-siRNA biosynthesis and function genes. The *B. rapa* genome underwent a whole genome triplication event

5–9 million years ago and many genes are still found in triplicate [33]. Homology searches of available *B. rapa* genome sequence revealed a single gene (Bra027611) with 78% amino acid identity to *Arabidopsis NRPD1*. Subsequent analysis of the completed *B. rapa* genome at the Brassica database [35] confirmed that this gene is syntenic to *AtNRPD1* and paralogs do not exist in *B. rapa*, therefore we refer to this gene as *BrNRPD1*. Similar analysis of other Pol IV components and associated proteins reveals that nearly all of these genes reduced to single copy after the genome triplication (Supplementary Table S1). 47% of *Arabidopsis* genes have more than one syntenic ortholog in *B. rapa* [33] and the gene dosage hypothesis [38] suggests that genes which physically interact in complexes are less likely to reduce copy number following whole genome duplication. Consistent with this, components of RNA polymerase II remain in multiple copies in the *B. rapa* genome. However, genes involved in other sRNA pathways, including those not known to function in complexes, such as DCLs, AGOs, and RDRs, also generally reduced to a single copy, suggesting that RNA silencing components might be susceptible to gene loss (Supplementary Table S1).

To identify a *B. rapa* mutant lacking Pol IV-dependent siRNAs, a *B. rapa* TILLING population [34] was screened for potential loss-of-function mutations in *BrNRPD1*. Within the target region eleven mutations were discovered; four silent mutations and five missense mutations in non-conserved residues were discarded. The remaining two mutations change amino acids that are highly conserved among RNA polymerase largest subunits (Figure 1A). These mutations were designated *brnrpd1-1* and *brnrpd1-2* and homozygous individuals were identified for subsequent analysis. The *brnrpd1-1* mutation is just outside the conserved D region, and the *brnrpd1-2* mutation is in the highly conserved Metal A binding site [39]. As in *Arabidopsis*, homozygous *brnrpd1* individuals are phenotypically indistinguishable from wild type or heterozygous siblings (Figure 1B).

**Figure 1.** Mutations in *Brassica rapa NRPD1*. (a) Location and nature of two missense mutations in Bra027611, BrNRPD1. The protein schematic depicts the eight conserved polymerase regions (gray boxes) and the DeCL domain (black box). Alignment shows conservation of the residues in *Arabidopsis* and rice DNA-dependent RNA polymerases (NRPB1 = Pol II, NRPA1 = Pol I, NRPC1 = Pol III). (b) Morphology of *brnrpd1-1* and *brnrpd1-2* leaves and flowers compared to wild type (R-o-18).



### 3.2. Small RNA Production in *Brassica rapa*

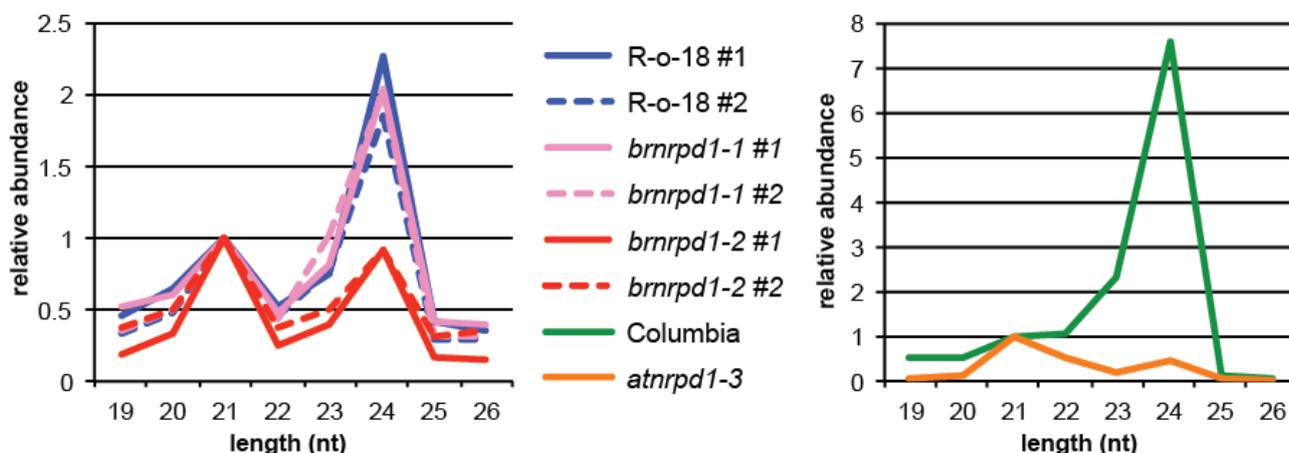
To characterize the *B. rapa* small RNA population, sRNA-seq libraries were generated from R-o-18 (wild type), *brnrpd1-1*, and *brnrpd1-2* reproductive tissue and sequenced on the Illumina platform. The resulting reads were filtered and matched to the genome before analysis (Table 1). Identical filters were applied to published *Arabidopsis* Columbia-0 and *atnrpd1-3* datasets [40] as a control. Although the R-o-18 strain differs from the sequenced reference strain, the percentage of sRNAs matching the genome is not appreciably different between *B. rapa* and *Arabidopsis* libraries, suggesting few sRNAs were lost due to overlap with SNPs.

**Table 1.** Small RNA-seq libraries.

Dataset	Total reads	Filtered	Genome matching	Non-redundant
R-o-18 #1	6,382,869	4,188,806	65.63%	544,690
R-o-18 #2	10,801,569	7,019,933	64.99%	708,041
<i>brnrpd1-1</i> #1	11,043,788	7,183,286	65.04%	943,725
<i>brnrpd1-1</i> #2	11,103,304	7,265,692	65.44%	774,524
<i>brnrpd1-2</i> #1	6,967,097	4,676,438	67.12%	481,686
<i>brnrpd1-2</i> #2	10,687,391	7,233,704	67.68%	553,872
Columbia	11,777,408	6,452,321	54.79%	1,525,519
<i>atnrpd1-3</i>	10,531,269	3,873,131	36.78%	365,031

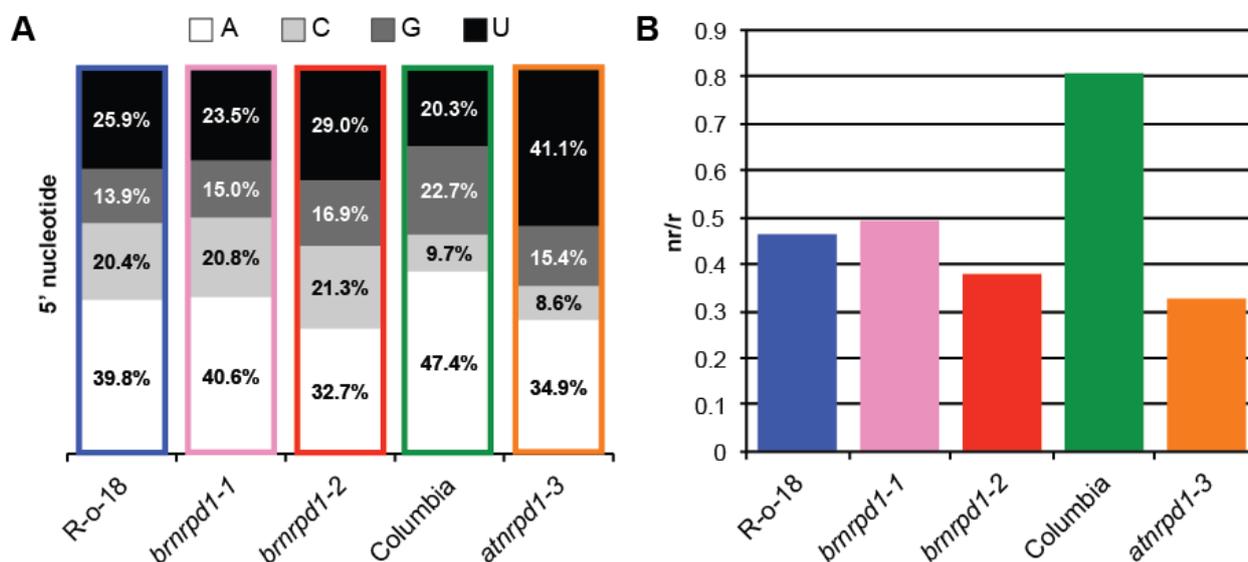
Length of the sRNA is an important factor in determining sRNA function, because it determines which Argonaute effector protein the sRNA will direct. In *Arabidopsis* AGO1, AGO2, AGO7, and AGO10 prefer to bind 21 nt sRNAs, which are predominantly microRNAs triggering post-transcriptional silencing [41]. In contrast, AGO4, AGO6, and AGO9 associate primarily with 24 nt sRNAs and target nascent RNAs to induce DNA methylation [41,42]. In *B. rapa*, the most abundant sRNA sizes are 24 nt and 21 nt, consistent with other sequenced plant small RNA transcriptomes [19–26]. Unlike its close relative *B. napus*, 22 and 23 nt sRNAs are infrequent in *B. rapa* [26]. Twenty-four nt sRNA are approximately 2-fold more abundant than 21 nt siRNAs, whereas this ratio is 7.5 to 1 in *Arabidopsis* (Figure 2). MiRNAs comprise a higher proportion of 21mers in *B. rapa* than *Arabidopsis* (44% versus 22%), suggesting that the reduced ratio of 24:21 nt sRNAs is not due to expression of additional 21 nt siRNAs in *B. rapa*, but rather decreased expression of 24 nt siRNAs relative to *Arabidopsis*. Loss of Pol IV activity through the *atnrpd1-3* mutation reduces the number of 24 nt siRNAs by over 15-fold, reflecting the importance of Pol IV in synthesis of 24 nt siRNAs in *Arabidopsis* (Figure 2). In *brnrpd1-2* reduction in 24 nt siRNAs is a more moderate 2-fold, bringing the ratio of 24:21nt sRNAs to 0.9:1 (0.43:1 in *atnrpd1-3*). The higher level of 24 nt sRNAs in *brnrpd1-2* compared to *atnrpd1-3* suggests that there are more Pol IV-independent 24 nt sRNAs in *B. rapa*. Alternatively, the *brnrpd1-2* mutation might retain some activity, although this is unlikely given the mutation occurs in the Metal A binding site, which is required for activity [43]. There is no change in size profile of sRNAs in the *brnrpd1-1* mutation, suggesting that this allele is functional or hypomorphic.

**Figure 2.** Size profile of small RNAs. Filtered, genome-matching sRNA reads from reproductive tissue of *B. rapa* (left) and *Arabidopsis* (right). R-o-18 and Columbia are wild type. Biological replicates are denoted #1 and #2.



The 5' nucleotide of a sRNA is also important to direct that molecule into the correct Argonaute effector protein. *Arabidopsis* sRNAs beginning with A are bound by AGO2, AGO4, AGO6, and AGO9, sRNAs beginning with U associate with AGO1, and sRNAs beginning with C are bound by AGO5 [41,42]. Given the deep conservation of microRNA sequence and function, it is likely that these specificities are conserved among plants. In *B. rapa* first nucleotide frequencies are similar to *Arabidopsis*, with fewer A and G 5' nucleotides and correspondingly more C and U (Figure 3A). As expected, loss-of-function *nprpd1* mutations cause a reduction in the level of sRNAs beginning with an A in both species, likely due to a reduction in 24 nt sRNAs associated with AGO4, AGO6, or AGO9.

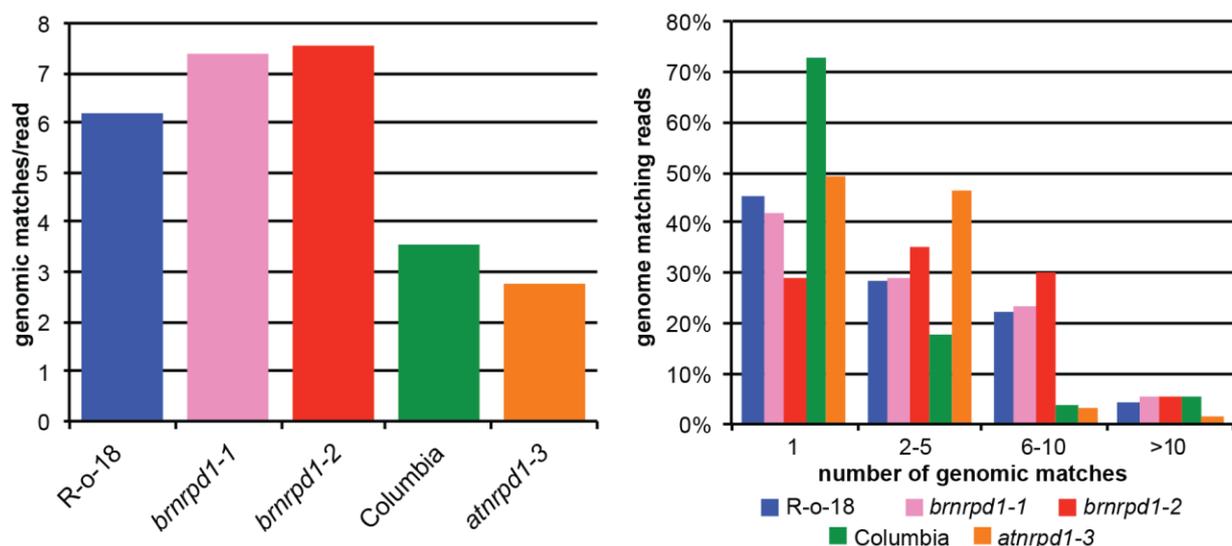
**Figure 3.** Characteristics of small RNA transcriptomes in *B. rapa* and *Arabidopsis*. (a) 5' nucleotide composition of sRNAs. (b) Complexity of sRNA populations as measured by the nr/r ratio between sRNA sequences (non-redundant) and reads (redundant). Because the complexity of an sRNA library is inversely proportional to the number of reads in the library, complexity was assessed with 100,000 randomly selected reads from each library.



Different sRNA classes also display different population complexities. MiRNAs are precisely cleaved from relatively small precursors, resulting in many copies of the exact sRNA sequence. In comparison, p4-siRNAs are randomly cleaved from larger double-stranded precursors, generating a highly diverse mixture of sRNA sequences. The *Arabidopsis* sRNA transcriptome is much more diverse than *B. rapa* and this diversity is due to Pol IV-dependent siRNAs, because complexity drops sharply in the *atnrpd1-3* mutant (Figure 3B). In *brnrpd1-2* complexity is also lower than R-o-18, however it is higher than *atnrpd1-3*, indicating that Pol IV-independent sRNAs are more complex in *B. rapa* than in *Arabidopsis*.

Pol IV-dependent siRNAs are frequently generated from repetitive genomic elements such as TEs and therefore have several perfect matches in the genome. Measuring the average number of genomic matches per sRNA is therefore one measure of the repetitive content of the sRNA population. Wild-type *B. rapa* sRNAs match the genome more than 6 times on average, compared to approximately 3.5 times for wild-type *Arabidopsis* sRNAs (Figure 4 left). This difference is due primarily to increases in moderately repetitive sequences (matching the genome 2–10 times), which are more frequent in *B. rapa* than in *Arabidopsis* (Figure 4 right). Reads matching the genome more than 10 times are more abundant in *Arabidopsis* than *B. rapa* and these reads are strongly diminished in *atnrpd1-3* (from 5.5% to 1.5%). Surprisingly, this class is more abundant in the *brnrpd1-2* mutant compared to R-o-18 (5.2% versus 4.3%). This suggests that while highly repetitive DNA produces Pol IV-dependent siRNAs in *Arabidopsis*, in *B. rapa* repetitive DNA generates siRNAs through a Pol IV-independent pathway.

**Figure 4.** Genomic repetitiveness of small RNA transcriptomes in *B. rapa* and *Arabidopsis*. Average genomic matches per read (left) and distribution of genomic repetitiveness (right).

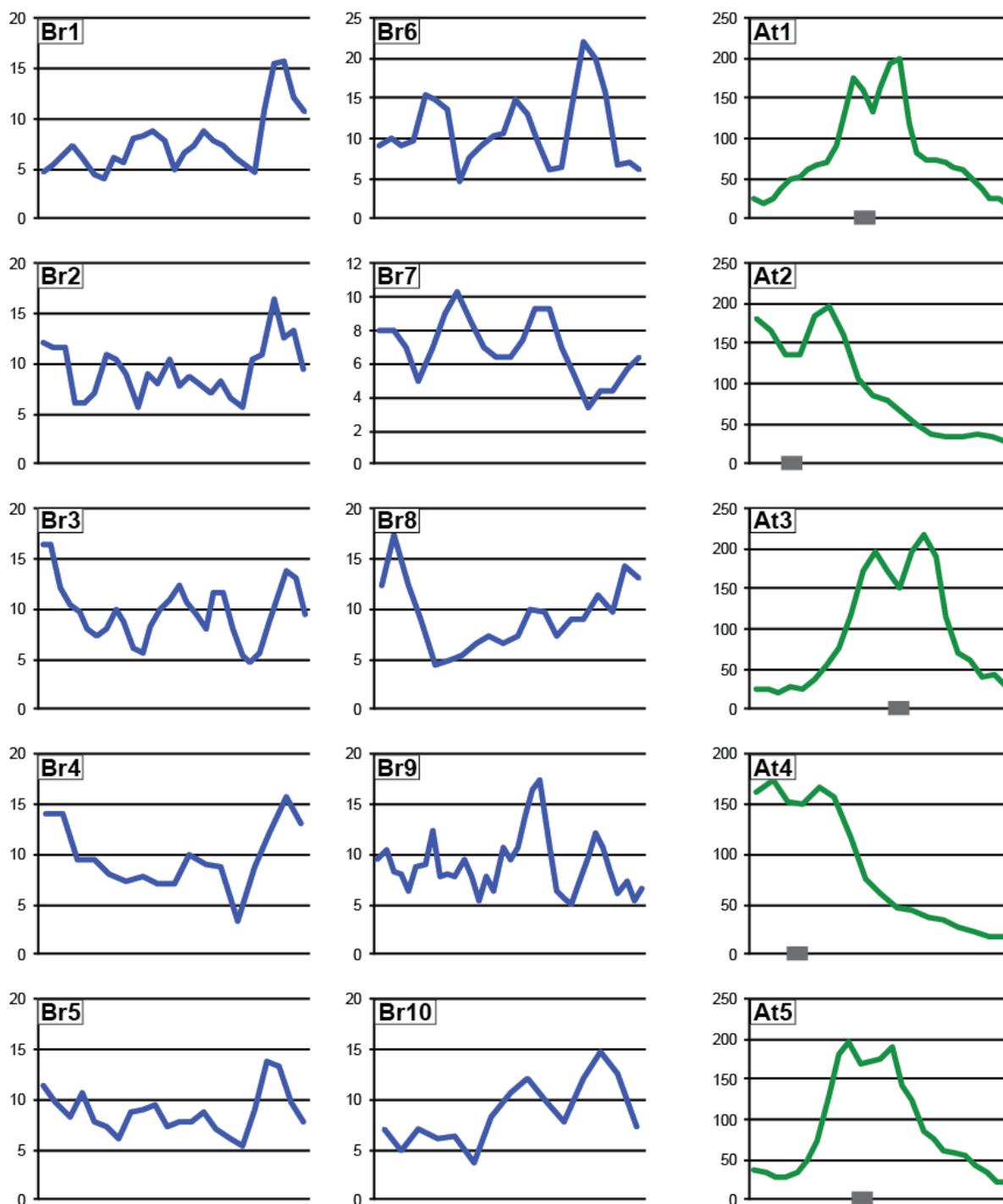


### 3.3. Pol IV-Dependent Loci in Brassica rapa

In *Arabidopsis*, Pol IV-dependent siRNAs are produced from thousands of genomic loci [3,8,9]. To assess the genomic regions producing p4-siRNAs in *B. rapa*, sRNAs were matched to the reference genome and quantified in static 500 base windows. Abundance in each window was normalized to total library size with reads matching to multiple genomic sites counted fractionally at each location and uniquely-mapping sRNAs flagged. Of the 512,523 total windows in the *B. rapa* genome,

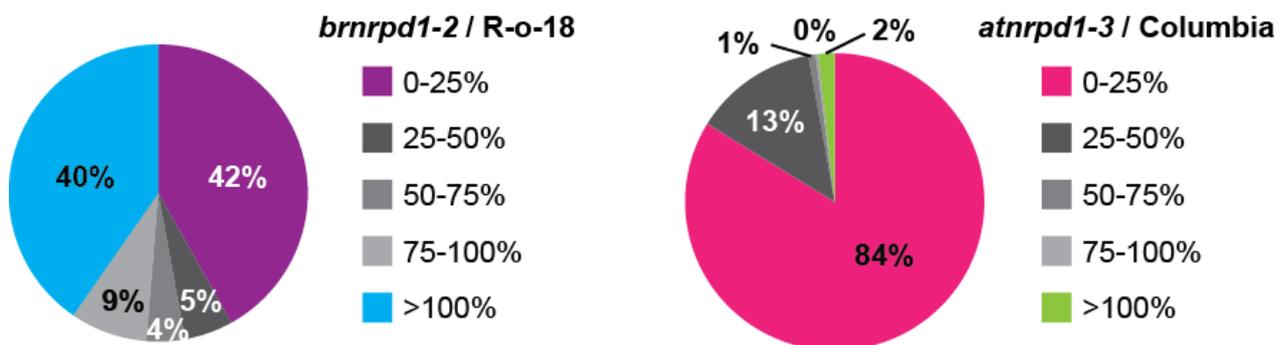
2,317 windows (0.45%) had at least one uniquely mapping read and an HNA (Hits Normalized Abundance)  $\geq 15$  reads per million, a level of sRNA production that allows confident analysis. Similar investigation of *Arabidopsis* sRNAs identified 10,305 windows (4.3% of 238,296 windows) passing these thresholds. Small RNA loci in *B. rapa* are present across each chromosome, while sRNA loci are most abundant in pericentromeric regions of *Arabidopsis* chromosomes (Figure 5).

**Figure 5.** Chromosomal distribution of sRNA-producing loci. Number of siRNA-producing windows in a 1Mb bin (rolling average of 3 bins) across the ten *B. rapa* and 5 *Arabidopsis* chromosomes. Location of the *Arabidopsis* centromeres (as in [44]) are noted with grey boxes.



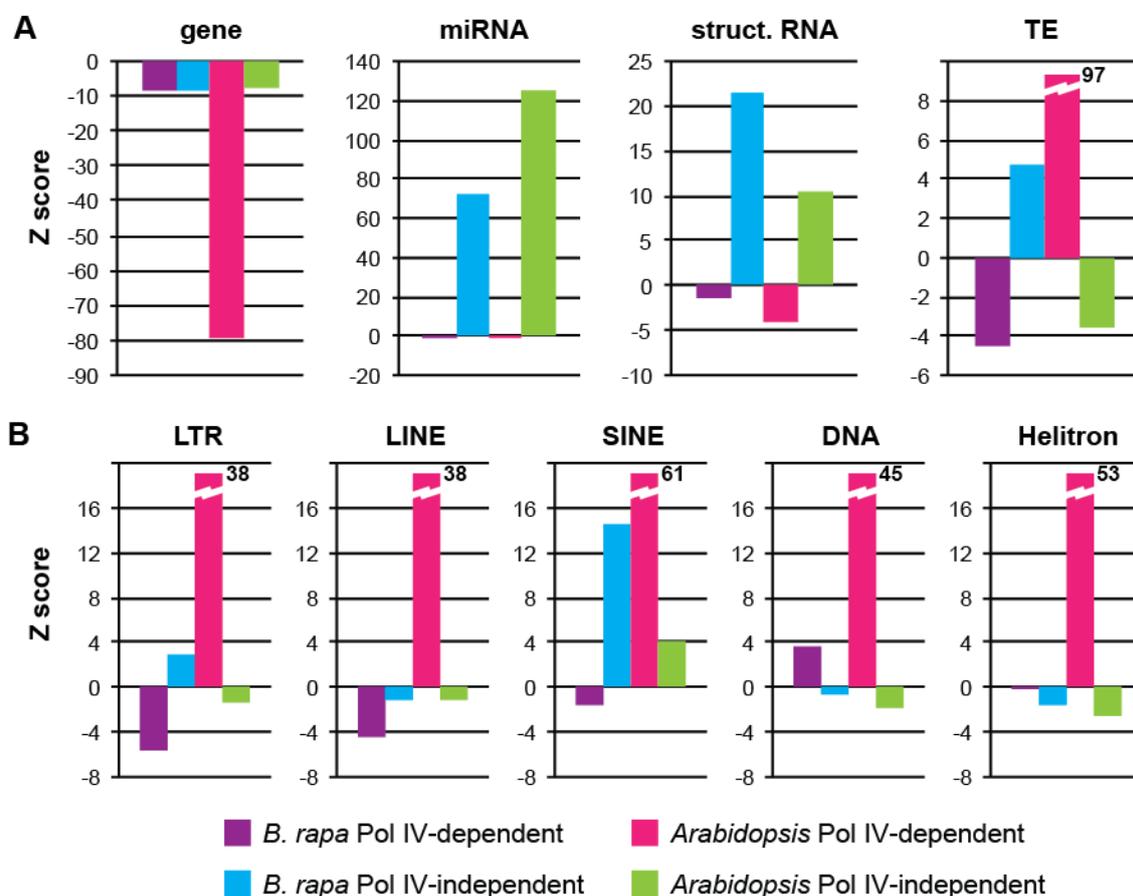
To determine the genomic loci producing p4-siRNAs, sRNA-producing windows were categorized based on the HNA in *brnrpd1-2* compared to R-o-18. 42% (971 windows) display strong depletion of sRNAs in *brnrpd1-2*, including 472 windows (20% of sRNA-producing windows) with <5% of R-o-18 HNA. A similar number of windows (935, 40%) show increased sRNA accumulation in the *brnrpd1-2* mutant, although this is mostly likely due to oversampling of unchanged Pol IV-independent sRNAs [3,19,45] (Figure 6). This pattern demonstrates that the *brnrpd1-2* mutation has a strong effect on sRNA biogenesis at specific loci rather than a weak effect at all loci, suggesting it is a null allele. In *Arabidopsis*, 84% (8637) of sRNA-producing windows are Pol IV-dependent and only 179 windows (1.7%) are Pol IV-independent. When the entire genome is considered, a window in *Arabidopsis* is almost 20 times more likely to produced Pol IV-dependent siRNAs than a window in the *B. rapa* genome (3.6% versus 0.19%), but a window in the *B. rapa* genome is almost 2.5 times as likely to produce Pol IV-independent sRNAs (0.18% versus 0.075%). The increased number of Pol IV-dependent sRNA loci in *Arabidopsis* compared to *B. rapa* might explain the difference in chromosomal distribution, as most of the loci present in the pericentromere of *Arabidopsis* are Pol IV-dependent (Supplementary Figure S2). Although the sRNA size profile in *brnrpd1-1* is very similar to R-o-18, analysis of genomic windows suggests this allele is hypomorphic, due to the correlation between sRNA accumulation in *brnrpd1-1* and *brnrpd1-2* mutants (Supplementary Figure S3).

**Figure 6.** Pol IV-dependent siRNA loci in *B. rapa* and *Arabidopsis*. Small RNA-generating windows were divided into five groups based on the degree of sRNA accumulation in the *brnrpd1-2* or *atnrpd1-3* mutants compared to respective wild types. Pie charts display the proportion of sRNA-generating windows falling into each Pol IV-dependency group ( $n = 2,317$  for *B. rapa*, 10,305 for *Arabidopsis*).



Annotations from each window were collected to assess which genomic features are associated with Pol IV-dependent and Pol IV-independent sRNA production in *B. rapa* (Supplementary Figure S4). Genes are moderately underrepresented in both Pol IV-dependent and Pol IV-independent loci, while microRNAs and windows overlapping structural RNA genes (rRNA, tRNA, snRNA) are overrepresented only among Pol IV-independent loci (Figure 7A). Similar patterns are seen at *Arabidopsis* Pol IV-independent and Pol IV-dependent loci, although genes are more highly depleted in Pol IV-dependent loci. Surprisingly, TEs are slightly underrepresented among *B. rapa* Pol IV-dependent loci and slightly over-represented among Pol IV-independent loci. This is opposite of the pattern detected in *Arabidopsis* and predicted by models of Pol IV function.

**Figure 7.** Genomic features at sRNA-generating windows. Over- and under-representation of annotation classes for Pol IV-dependent windows (0–25% class) and Pol IV-independent windows (>100% class). (a) General genome features. (b) Transposable element classes.



Further analysis of the types of TEs associated with each group of sRNA loci demonstrates mild but significant depletion of LTR and LINE retrotransposons among *B. rapa* Pol IV-dependent loci, while DNA transposons are slightly overrepresented (Figure 7B). Among Pol IV-independent loci LTR retrotransposons are barely over-represented while SINE elements are significantly overrepresented. In *Arabidopsis*, all categories of TEs display strong overrepresentation specifically at Pol IV-dependent loci. These differences suggest that biogenesis of sRNAs from TEs is distinct in *Brassica* and *Arabidopsis* and that the concordance between p4-siRNA production and TE presence in *Arabidopsis* might not be conserved among all plants.

#### 4. Conclusions

Although *B. rapa* and *Arabidopsis* are close relatives, their sRNA transcriptomes are surprisingly different. In *Arabidopsis*, the Pol IV pathway produces abundant 24 nt sRNAs matching thousands of genomic locations while in *B. rapa*, Pol IV produces a moderate level of 24 nt sRNA from several hundred genomic locations. Pol IV activity is focused on pericentromeric sequences in *Arabidopsis* and is distributed across the chromosomes in *B. rapa*. Although it is possible that the *brnrpd1-2* allele retains some function, the large number of loci that lose all sRNA production and the small number of loci with intermediate sRNA production in *brnrpd1-2* make it likely that this is a null mutation.

Most flowering plants possess more 24 nt than 21 nt sRNAs [19–25], however the balance between these sizes is unusually high in *Arabidopsis* floral tissue, suggesting that *Arabidopsis* might be an outlier. It is possible that the Pol IV pathway has been recently recruited to pericentromeric TEs in *Arabidopsis* and is hyperactive at these sequences. Abundant p4-siRNA production from the pericentromere would boost the 24:21 nt ratio and might mask “normal” p4-siRNA production from genomic sites distributed across the chromosomes. Alternatively, the floral-specific expression of many Pol IV loci in *Arabidopsis* [46] might be absent in *B. rapa*, resulting in a “vegetative” expression pattern throughout development. Further analysis of sRNA transcriptomes in the *Brassicaceae* is needed to unravel these possibilities.

In *Arabidopsis*, Pol IV transcribes all classes of TE [3] and is required for transcriptional silencing of some elements [10,11,47]. Although 13% of identified Pol IV-dependent loci overlap annotated TEs in *B. rapa* (Supplementary Figure S4), TE sequences are not enriched at Pol IV sites compared to the entire genome. Instead, some highly repetitive sequences in *B. rapa* produce Pol IV-independent siRNAs, reminiscent of the RDR2-independent 22 nt siRNAs produced from highly repetitive TE sequences in maize [19]. Twenty-four nt sRNAs from many species are enriched in TE sequences compared to 21 nt sRNAs, however it is unclear whether they are enriched compared to the genome as a whole. Additional analysis of sRNA transcriptomes from sequenced genomes is needed to strengthen the assumption that 24 nt sRNAs are primarily TE in origin.

Our analysis of a *B. rapa* Pol IV mutant highlights the diverse populations of sRNAs produced in plants and the hazards of drawing models based on a single species. *B. rapa* displays limited Pol IV activity relative to *Arabidopsis*, which might have allowed Pol IV-independent pathways to proliferate. Our research demonstrates that Pol IV-independent 24 nt sRNAs might be as abundant as 21 nt sRNAs in many plant species. It will be interesting to unravel the biogenesis of Pol IV-independent 24 nt sRNAs and discover whether they integrate into AGO4/AGO6/AGO9 and target RNA-directed DNA methylation or participate in other silencing pathways.

## Acknowledgments

The authors thank Fran Robson at RevGenUK for assistance in screening the *B. rapa* TILLING population. This research was funded by in part by a University of Arizona Foundation Faculty Seed Grant to RAM.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

1. Valencia-Sanchez, M.A.; Liu, J.; Hannon, G.J.; Parker, R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Gene Dev.* **2006**, *20*, 515–524.
2. Moazed, D. Small RNAs in transcriptional gene silencing and genome defence. *Nature* **2009**, *457*, 413–420.

3. Mosher, R.A.; Schwach, F.; Studholme, D.; Baulcombe, D.C. PolIVb influences RNA-directed DNA-methylation independently of its role in siRNA biogenesis. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 3145–3150.
4. Haag, J.R.; Ream, T.S.; Marasco, M.; Nicora, C.D.; Norbeck, A.D.; Pasa-Tolic, L.; Pikaard, C.S. *In vitro* transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. *Mol. Cell* **2012**, *48*, 811–818.
5. Xie, Z.; Johansen, L.K.; Gustafson, A.M.; Kasschau, K.D.; Lellis, A.D.; Zilberman, D.; Jacobsen, S.E.; Carrington, J.C. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2004**, *2*, 642–652.
6. Wierzbicki, A.T.; Ream, T.S.; Haag, J.R.; Pikaard, C.S. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat. Genet.* **2009**, *41*, 630–634.
7. Law, J.A.; Jacobsen, S.E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **2010**, *11*, 204–220.
8. Kasschau, K.D.; Fahlgren, N.; Chapman, E.J.; Sullivan, C.M.; Cumbie, J.S.; Givan, S.A.; Carrington, J.C. Genome-wide profiling and analysis of *Arabidopsis* siRNAs. *PLoS Biol.* **2007**, *5*, e57.
9. Zhang, X.; Henderson, I.R.; Lu, C.; Green, P.J.; Jacobsen, S.E. Role of RNA polymerase IV in plant small RNA metabolism. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 4536–4541.
10. Herr, A.J.; Jensen, M.B.; Dalmay, T.; Baulcombe, D. RNA polymerase IV directs silencing of endogenous DNA. *Science* **2005**, *308*, 118–120.
11. Kanno, T.; Huettel, B.; Mette, M.F.; Aufsatz, W.; Jaligot, E.; Daxinger, L.; Kreil, D.P.; Matzke, M.; Matzke, A.J. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* **2005**, *37*, 761–765.
12. Onodera, Y.; Haag, J.R.; Ream, T.; Nunes, P.C.; Pontes, O.; Pikaard, C.S. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **2005**, *120*, 613–622.
13. Pontier, D.; Yahubyan, G.; Vega, D.; Bulski, A.; SaezVasquez, J.; Hakimi, M.-A.; Lerbs-Mache, S.; Colot, V.; Lagrange, T. Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Gene Dev.* **2005**, *19*, 2030–2040.
14. Chan, S.W.; Zilberman, D.; Xie, Z.; Johansen, L.K.; Carrington, J.C.; Jacobsen, S.E. RNA silencing genes control de novo DNA methylation. *Science* **2004**, *303*, 1336.
15. Perez-Hormaeche, J.; Potet, F.; Beauclair, L.; le Masson, I.; Courtial, B.; Bouche, N.; Lucas, H. Invasion of the *Arabidopsis* genome by the tobacco retrotransposon Tnt1 is controlled by reversible transcriptional gene silencing. *Plant Physiol.* **2008**, *147*, 1264–1278.
16. Mari-Ordóñez, A.; Marchais, A.; Etcheverry, M.; Martin, A.; Colot, V.; Voinnet, O. Reconstructing de novo silencing of an active plant retrotransposon. *Nat. Genet.* **2013**, *45*, 1029–1039.
17. Matzke, M.; Kanno, T.; Daxinger, L.; Huettel, B.; Matzke, A.J. RNA-mediated chromatin-based silencing in plants. *Curr. Opin. Cell Biol.* **2009**, *21*, 367–376.
18. Pikaard, C.S.; Haag, J.R.; Ream, T.; Wierzbicki, A.T. Roles of RNA polymerase IV in gene silencing. *Trends Plant Sci.* **2008**, *13*, 390–397.

19. Nobuta, K.; Lu, C.; Shrivastava, R.; Pillay, M.; de Paoli, E.; Accerbi, M.; Arteaga-Vazquez, M.; Sidorenko, L.; Jeong, D.H.; Yen, Y.; *et al.* Distinct size distribution of endogeneous siRNAs in maize: Evidence from deep sequencing in the mop1–1 mutant. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 14958–14963.
20. Szittyá, G.; Moxon, S.; Santos, D.M.; Jing, R.; Fevereiro, M.P.; Moulton, V.; Dalmay, T. High-throughput sequencing of *Medicago truncatula* short RNAs identifies eight new miRNA families. *BMC Genomics* **2008**, *9*, 593.
21. Mohorianu, I.; Schwach, F.; Jing, R.; Lopez-Gomollon, S.; Moxon, S.; Szittyá, G.; Sorefan, K.; Moulton, V.; Dalmay, T. Profiling of short RNAs during fleshy fruit development reveals stage-specific sRNAome expression patterns. *Plant. J.* **2011**, *67*, 232–246.
22. Jeong, D.H.; Park, S.; Zhai, J.; Gurazada, S.G.; de Paoli, E.; Meyers, B.C.; Green, P.J. Massive analysis of rice small RNAs: Mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. *Plant Cell* **2011**, *23*, 4185–4207.
23. Aryal, R.; Yang, X.; Yu, Q.; Sunkar, R.; Li, L.; Ming, R. Asymmetric purine-pyrimidine distribution in cellular small RNA population of papaya. *BMC Genomics* **2012**, *13*, 682.
24. Ortiz-Morea, F.A.; Vicentini, R.; Silva, G.F.; Silva, E.M.; Carrer, H.; Rodrigues, A.P.; Nogueira, F.T. Global analysis of the sugarcane microtranscriptome reveals a unique composition of small RNAs associated with axillary bud outgrowth. *J. Exp. Bot.* **2013**, *64*, 2307–2320.
25. Sun, F.; Guo, W.; Du, J.; Ni, Z.; Sun, Q.; Yao, Y. Widespread, abundant, and diverse TE-associated siRNAs in developing wheat grain. *Gene* **2013**, *522*, 1–7.
26. Zhao, Y.T.; Wang, M.; Fu, S.X.; Yang, W.C.; Qi, C.K.; Wang, X.J. Small RNA profiling in two *Brassica napus* cultivars identifies microRNAs with oil production- and development-correlated expression and new small RNA classes. *Plant Physiol.* **2012**, *158*, 813–823.
27. Henderson, I.R.; Zhang, X.Y.; Lu, C.; Johnson, L.; Meyers, B.C.; Green, P.J.; Jacobsen, S.E. Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat. Genet.* **2006**, *38*, 721–725.
28. Daxinger, L.; Kanno, T.; Bucher, E.; van der Winden, J.; Naumann, U.; Matzke, A.J.; Matzke, M. A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. *EMBO J.* **2009**, *28*, 48–57.
29. Dunoyer, P.; Brosnan, C.A.; Schott, G.; Wang, Y.; Jay, F.; Alioua, A.; Himber, C.; Voinnet, O. An endogenous, systemic RNAi pathway in plants. *EMBO J.* **2010**, *29*, 1699–1712.
30. Erhard, K.F.; Stonaker, J.L.; Parkinson, S.E.; Lim, J.P.; Hale, C.J.; Hollick, J.B. RNA polymerase IV functions in paramutation in *zea mays*. *Science* **2009**, *323*, 1201–1205.
31. Stonaker, J.L.; Lim, J.P.; Erhard, K.F., Jr.; Hollick, J.B. Diversity of Pol IV function is defined by mutations at the maize *rnr7* locus. *PLoS Genet.* **2009**, *5*, e1000706.
32. Warwick, S.I. Brassicaceae in Agriculture. *Plant. Genet. Genomics* **2011**, *9*, 33–65.
33. Wang, X.; Wang, H.; Wang, J.; Sun, R.; Wu, J.; Liu, S.; Bai, Y.; Mun, J.H.; Bancroft, I.; Cheng, F.; *et al.* The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* **2011**, *43*, 1035–1039.
34. Stephenson, P.; Baker, D.; Girin, T.; Perez, A.; Amoah, S.; King, G.J.; Ostergaard, L. A rich TILLING resource for studying gene function in *Brassica rapa*. *BMC Plant Biol.* **2010**, *10*, 62.

35. Cheng, F.; Liu, S.; Wu, J.; Fang, L.; Sun, S.; Liu, B.; Li, P.; Hua, W.; Wang, X. BRAD, the genetics and genomics database for Brassica plants. *BMC Plant Biol.* **2011**, *11*, 136. Available online: <http://www.brassicadb.org/> (accessed on 1 August 2013).
36. Mosher, R.A.; Tan, E.H.; Shin, J.; Fischer, R.L.; Pikaard, C.S.; Baulcombe, D.C. An atypical epigenetic mechanism affects uniparental expression of Pol IV-dependent siRNAs. *PLoS One* **2011**, *6*, e25756.
37. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **2009**, *10*, R25.
38. Birchler, J.A.; Veitia, R.A. The gene balance hypothesis: From classical genetics to modern genomics. *Plant Cell* **2007**, *19*, 395–402.
39. Haag, J.R.; Pikaard, C.S. Multisubunit RNA polymerases IV and V: Purveyors of non-coding RNA for plant gene silencing. *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 483–492.
40. Wierzbicki, A.T.; Cocklin, R.; Mayampurath, A.; Lister, R.; Rowley, M.J.; Gregory, B.D.; Ecker, J.R.; Tang, H.; Pikaard, C.S. Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. *Gene Dev.* **2012**, *26*, 1825–1836.
41. Mi, S.J.; Cai, T.; Hu, Y.G.; Chen, Y.; Hodges, E.; Ni, F.R.; Wu, L.; Li, S.; Zhou, H.; Long, C.Z.; *et al.* Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell* **2008**, *133*, 116–127.
42. Havecker, E.R.; Wallbridge, L.M.; Hardcastle, T.J.; Bush, M.S.; Kelly, K.A.; Dunn, R.M.; Schwach, F.; Doonan, J.H.; Baulcombe, D.C. The arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell* **2010**, *22*, 321–334.
43. Haag, J.R.; Pontes, O.; Pikaard, C.S. Metal A and metal B sites of nuclear RNA polymerases Pol IV and Pol V are required for siRNA-dependent DNA methylation and gene silencing. *PLoS One* **2009**, *4*, e4110.
44. Pereira, V. Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. *Genome Biol.* **2004**, *5*, R79.
45. Lu, C.; Kulkarni, K.; Souret, F.F.; MuthuValliappan, R.; STEj, S.S.; Poethig, R.S.; Henderson, I.R.; Jacobsen, S.E.; Wang, W.; Green, P.J.; *et al.* MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. *Genome Res.* **2006**, *16*, 1276–1288.
46. Mosher, R.A.; Melnyk, C.W.; Kelly, K.A.; Dunn, R.M.; Studholme, D.J.; Baulcombe, D.C. Uniparental expression of PolIV-dependent siRNAs in developing endosperm of *Arabidopsis*. *Nature* **2009**, *460*, 283–286.
47. Huettel, B.; Kanno, T.; Daxinger, L.; Aufsatz, W.; Matzke, A.J.M.; Matzke, M. Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*. *EMBO J.* **2006**, *25*, 2828–2836.