AGO4 is specifically required for heterochromatic siRNA accumulation at Pol V-dependent loci in *Arabidopsis thaliana*

Feng Wang[1,2] and Michael J. Axtell[1,2,*]

[1] Intercollege Plant Biology Ph.D. Program, Huck Institutes of the Life Sciences, Penn State University, University Park, PA 16802 USA.

[2] Department of Biology, Penn State University, University Park, PA 16802 USA

* Corresponding author: mja18@psu.edu

Running title: *AGO4*-dependent siRNA accumulation

Abstract

In plants, 24 nucleotide long heterochromatic siRNAs (het-siRNAs) transcriptionally regulate gene expression by RNA-directed DNA methylation (RdDM). The biogenesis of most het-siRNAs depends on the plant-specific RNA polymerase IV (Pol IV), and ARGONAUTE4 (AGO4) is a major het-siRNA effector protein. Through genome-wide analysis of sRNA-seq data sets, we found that *AGO4* is required for the accumulation of a small subset of het-siRNAs. The accumulation of *AGO4*-dependent het-siRNAs also requires several factors known to participate in the effector portion of the RdDM pathway, including RNA POLYMERASE V (POL V), DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) and SAWADEE HOMEODOMAIN HOMOLOG 1 (SSH1). Like many AGO proteins, AGO4 is an endonuclease that can 'slice' RNAs. We found that a slicing-defective AGO4 was unable to fully recover *AGO4*-dependent het-siRNA accumulation from *ago4* mutant plants. Collectively, our data suggest that *AGO4*-dependent siRNAs are secondary siRNAs dependent on the prior activity of the RdDM pathway at certain loci.

Keywords: AGO4 / heterochromatic siRNA accumulation / secondary heterochromatic siRNAs

Introduction

24 nucleotide (nt) heterochromatic small interfering RNAs (het-siRNAs) are usually loaded into ARGONAUTE4 (AGO4) to direct repressive chromatic modifications and subsequent transcriptional gene silencing via RNA-directed DNA Methylation (RdDM) [1,2]. Het-siRNA-induced transcriptional silencing plays important roles in transposable element silencing, stress responses and genome stability [3,4]. The production of het-siRNAs requires the plant-specific RNA polymerase IV (Pol IV) [5-8], RNA-dependent RNA polymerase 2 (RDR2) [9,10] and Dicer-like 3 (DCL3) [11]. A second plant-specific RNA polymerase, Pol V, generates scaffold RNAs targeted by het-siRNAs associated with AGO4 [12,13]. Despite its position at the effector portion of the RdDM pathway, Pol V is required for the accumulation of a subset of Pol IV-dependent het-siRNAs in *Arabidopsis* [14,15].

The *Arabidopsis* genome has 10 *AGO* genes. AGO4, AGO6, AGO8, and AGO9 form a monophyletic clade [16-18]. *AGO8* has been suggested as a pseudogene [18]. AGO4 and AGO6

both bind 24 nt het-siRNAs and contribute to the canonical RdDM pathway in a non-redundant fashion [19-21]. AGO6 also binds 21 nt siRNAs and act as a key effector of the non-canonical RDR6-RdDM pathway [22]. AGO9, which is primarily expressed in female gametes, interacts with het-siRNAs and silence TEs in female gametes [23]. Though AGO4, AGO6 and AGO9 are functionally related, the small RNA profile of an *ago4*/*ago6*/*ago9* triple mutant has not been reported yet.

According to the current model of the RNA-directed DNA methylation (RdDM) pathway, the biogenesis of het-siRNAs depends on Pol IV, RDR2, and DCL3 [3,4], while AGO4 is not directly required for the biogenesis of het-siRNAs. However, the accumulation of certain het-siRNAs was shown to be dependent on AGO4 in previous reports [2,19]. It has been hypothesized that the accumulation of a subset of het-siRNAs depends on AGO4-mediated target slicing [2]. All 10 Arabidopsis AGOs have a conserved Asp-Asp-His (DDH) or Asp-Asp-Asp (DDD) motif thought to form a catalytic center for cleavage of target RNA. The target-slicing ability of AGO1 and AGO7 has been confirmed *in vivo* [17,18]. AGO10 can slice miRNA target in vitro, but it is still unclear if the slicer-activity is required for its function in plants [24]. AGO4, which specifically binds 24nt het-siRNAs, can slice synthetic het-siRNA targets *in vitro* [2] as well as the passenger-strand of het-siRNA duplexes *in vivo* [25]. The *in vitro* and/or *in vivo* slicing ability of AGO4 is abolished by mutagenesis of the presumed catalytic triad [2,25]. However, the genome-wide effects of AGO4 slicing on global small RNA accumulation have not been previously reported.

Results and Discussion

Accumulation of a subset of 24 nt het-siRNAs depends on *AGO4* in *Arabidopsis*

To systematically study the profile of AGO4-dependent het-siRNAs and the effect of AGO4 catalytic activity on het-siRNA accumulation, we expressed wild-type *AGO4* (*pAGO4:FLAG-AGO4-DDH*, wt*AGO4* hereafter) or slicing-defective *AGO4* (*pAGO4:FLAG-AGO4-DAH*, *D742A* hereafter) driven by the native *AGO4* promoter in both the *ago4-4* single mutant background and the *ago4-4/6-2/9-1* triple mutant background in *Arabidopsis* (Fig EV1A). Three T3 transgenic plants with comparable level of protein accumulation (Fig EV1B) were used to prepare three biological replicate sRNA-seq libraries. It is worth noting that *ago4-4* is in Ws

background, while *ago6-2* and *ago9-1* are in Col-0 background. We therefore prepared three replicate control sRNA-seq libraries from both Ws and Col-0. We merged sRNA-seq libraries from the same genotype and aligned them to reference genome to study the overall small RNA size distribution in tested samples. Loci dominated by 24 nt small RNAs were the most abundant in all tested genotypes, and the fractions of small RNA from 24 nt small RNA-dominated loci were similar across different genotypes (Fig EV2). sRNA-seq libraries from all backgrounds were then merged, aligned to the reference genome, followed by *de novo* definition of expressed small RNA clusters. The 24 nt siRNA clusters that were *de novo* annotated are listed in Appendix Table S1.

We first examined siRNA accumulation in the Ws background, to compare *ago4-4* to the wild-type. A differential expression analysis was performed by comparing raw read counts from our *de novo* annotated small RNA loci for all libraries in Ws background. A principal component analysis (PCA) plot was prepared to visualize the overall differences between samples (Fig 1A). The biological replicates were grouped together, indicating good reproducibility (Fig 1A). *ago4-4*/*wtAGO4* grouped closely with the Ws wild-type, suggesting complementation of small RNA accumulation by expression of *wtAGO4* in the *ago4-4* background (Fig 1A). *ago4-4* and *ago4-4*/*D742A* were distinct from each other and from the wild-type and *ago4-4/wtAGO4* genotypes (Fig 1A).

Most differentially accumulated clusters were dominated by 24 nt siRNAs (Fig 1B). In *ago4-4*, 2,912 clusters were down-regulated relative to wild-type; we defined these as *AGO4*-dependent siRNA clusters (Fig 1B). Most of these (2,879) were dominated by 24 nt siRNAs. In contrast, only 121 clusters were down regulated in *ago4-4*/*wtAGO4*, indicating nearly full complementation of small RNA accumulation by *wtAGO4* (Fig 1B). Intriguingly, an intermediate amount of clusters (1,541, Fig 1B) was down-regulated in *ago4-4*/*D742A*, which suggested that slicing-defective AGO4 partially recovers the accumulation of *AGO4*-dependent small RNAs. Most 24 nt-dominated siRNA clusters are not *AGO4*-dependent. Only about 18% of the *de novo* annotated 24 nt siRNA clusters, which contained about 22% of small RNAs in Ws wild-type, were dependent on *AGO4* (Fig 1C).

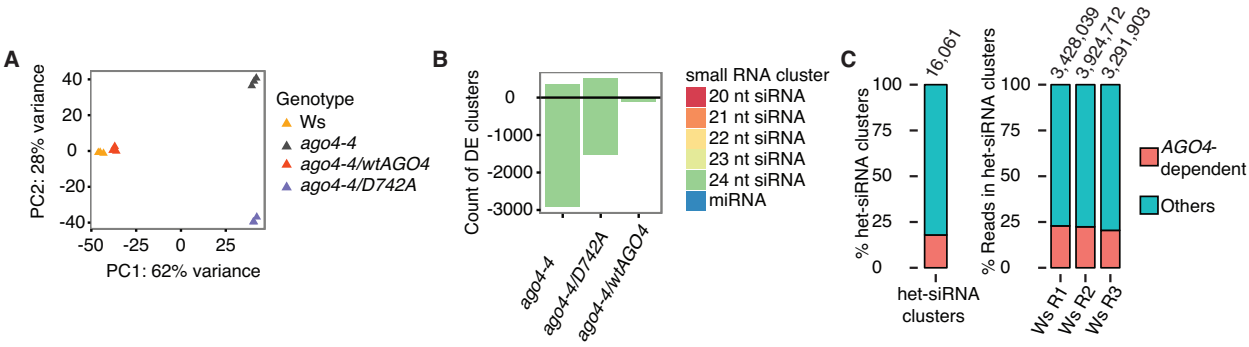

Figure 1. Identification of *AGO4*-dependent small RNA clusters in *Arabidopsis thaliana*

A. Principal component analysis demonstrating overall relationships between sRNA-seq libraries in Ws background.

B. Number of differentially expressed (DE) clusters in the indicated genotypes and small RNA clusters compared with Ws wild-type. DE clusters were defined as clusters with at least 2-fold change compared to wild-type at a false discovery rate of 1%.

C. Percentage of *AGO4*-dependent clusters and *AGO4*-dependent small RNAs in 24 nt siRNA loci. *AGO4*-dependent clusters were defined as clusters with at least 2-fold less accumulation in *ago4-4* compared with Ws.

Accumulation of small RNAs in *AGO4*-dependent clusters requires *NRPE1*, *DRM2* and *SSH1*

We classified the 16,061 *de novo* annotated 24 nt siRNA-dominated clusters into different groups based on *AGO4*-dependency (Appendix Table S1). As stated above 2,879 24 nt-dominated siRNA clusters were *AGO4*-dependent (FDR=0.01). We found another 1,359 24 nt-dominated clusters that were clearly *AGO4*-independent (FDR=0.01). Another 354 24 nt-dominated clusters were up-regulated in *ago4-4* (FDR=0.01), and the *AGO4*-dependency of the remaining 11,469 24 nt-dominated siRNA clusters could not be reliably inferred using our strict statistical tests, primarily due to low expression levels. We analyzed sRNA-seq accumulation from the *AGO4*-dependent and *AGO4*-independent clusters using data from *nrpd1-4*, *nrpe1-12*, *drm2-2*, and *ssh1-1* mutants [26], using accumulation of clusters overlapping high-confidence *MIRNA* loci as a control (Fig 2A). In *nrpd1*, siRNA accumulation was strongly down-regulated in both *AGO4*-dependent and *AGO4*-independent clusters (Fig 2A). In contrast, *AGO4*-

dependent clusters were much more strongly affected in the *nrpe1*, *drm2*, and *ssh1* backgrounds compared to AGO4-independent clusters (Fig 2A). We then normalized small RNA accumulation in *AGO4*-dependent and *AGO4*-independent het-siRNA clusters based on wild-type plants. We observed significantly reduced small RNA accumulation (Mann-Whitney test, $p<0.01$) in *AGO4*-dependent clusters relative to *AGO4*-independent clusters in all analyzed RdDM mutants except *nrpd1* (Fig 2B). Using small RNA-seq data from *nrpe1-1* plants [27], we defined 2,827 *NRPE1*-dependent small RNA clusters, the majority of which overlapped *AGO4*-dependent siRNA clusters (Fig 2C). This extent of overlap far exceeded the number expected by random chance (Fig 2D). Collectively, these data indicate that the subset of 24 nt dominated siRNA loci that depend on *AGO4* for accumulation are also generally dependent on the downstream RdDM factors *NRPE1*, *DRM2*, and *SSH1*.

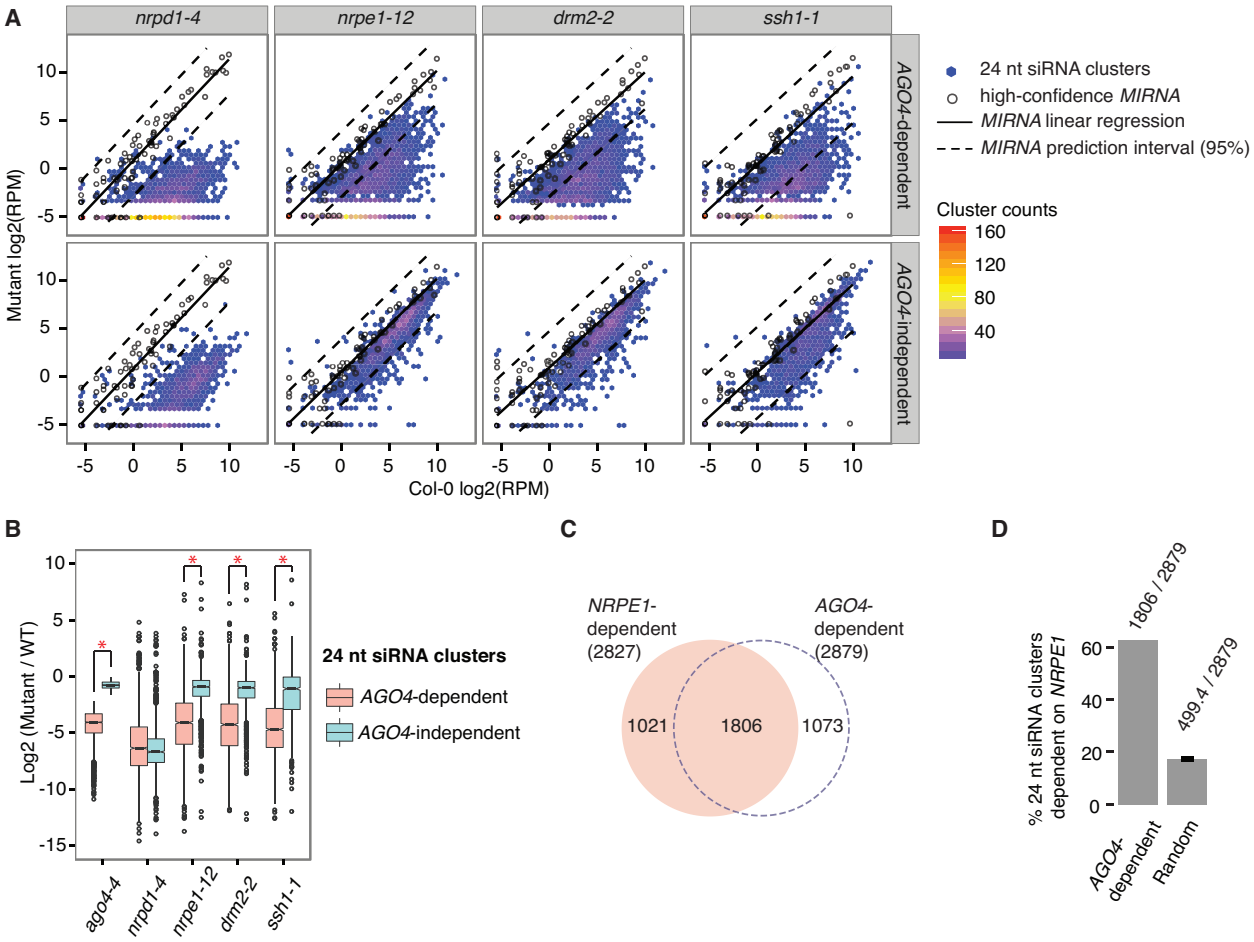

Figure 2. *AGO4*-dependent and *AGO4*-independent 24 nt siRNA clusters in other RdDM mutants

A. Small RNA accumulation from *AGO4*-dependent and *AGO4*-independent 24 nt siRNA clusters in indicated RdDM mutants. Log2 transformed reads per million (RPM) in indicated genotypes were plotted. A linear regression (solid line) and associated 95% prediction interval (dashed lines) was plotted based upon accumulation from clusters overlapping high-confidence *MIRNA* loci.

B. Normalized small RNA accumulation in *AGO4*-dependent and *AGO4*-independent clusters in indicated RdDM mutants. Boxplots show medians (horizontal lines), the 1st-3rd quartile range (boxes), 95% confidence of medians (notches), other data out to 1.5 times the interquartile range (whiskers) and outliers (dots). Asterisks indicate significant differences (Mann-Whitney U test, $p<0.01$) between *AGO4*-dependent and *AGO4*-independent clusters in the indicated mutant.

C. Venn diagram showing the overlap of *AGO4*-dependent and *NRPE1*-dependent 24 nt siRNA clusters.

D. Percentage of overlap between *AGO4*-dependent and *NRPE1*-dependent clusters. Overlaps expected by random chance were estimated by by randomly choosing 2827 and 2879 clusters from all 24 nt siRNA clusters. The mean and standard deviation (n=10) of randomly overlapping percentages are shown.

An AGO4 catalytic residue is required for full accumulation of most *AGO4*-dependent 24 nt siRNAs

We then compared complementation of siRNA accumulation from *AGO4*-dependent clusters between the *wtAGO4* and *AGO4-D742A* transgenic lines. Small RNA accumulation was recovered in *wtAGO4* from nearly all *AGO4*-dependent clusters, but only from a small subset of loci in the slicing-defective *AGO4-D742A* plants (Fig 3A). We defined *AGO4-D742A* complemented loci as those that were significantly down-regulated in the *ago4-4* background but not in the *ago4-4/AGO4-D742A* transgenic plants (Fig 3B). Conversely, *AGO4-D742A* non-complemented loci were defined as those that were significantly down-regulated in both *ago4-4* and *ago4-4/AGO4-D742A* (Fig 3B). By this measure, half (49.9%) of the *AGO4*-dependent

siRNA loci required AGO4 catalytic activity for their accumulation. Detailed examination of accumulation levels revealed that recovery was generally not to full wild-type levels at loci designated as complemented by *AGO4-D742A* (Fig 3C). We conclude that that the catalytic ability of AGO4 is important for full accumulation of most *AGO4*-dependent 24 nt siRNAs, but to varying degrees at different loci.

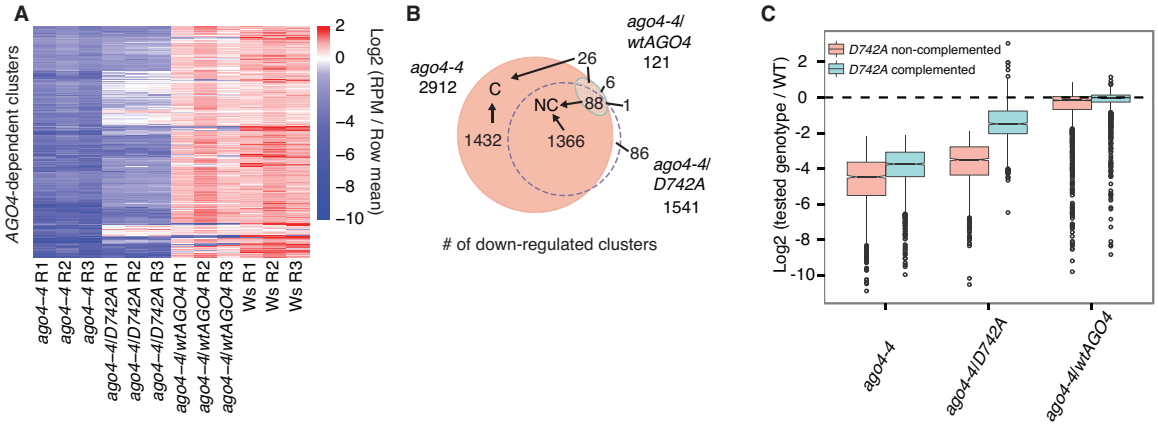

Figure 3. Slicing-defective *AGO4-D742A* partially complements small RNA accumulation from *AGO4*-dependent siRNA loci

A. Heatmap showing normalized (log2-transformed and mean-centered) small RNA accumulation from *AGO4*-dependent clusters in indicated genotypes and replicates.

B. Euler diagram showing overlaps between down-regulated small RNA clusters (FDR=0.01) in the indicated genotypes compared to Ws wild-type. C: Complemented by the *AGO4-D742A* transgene; NC: Not complemented by the *AGO4-D742A* transgene.

C. Normalized small RNA accumulation levels from *AGO4*-dependent loci that were complemented or not complemented by the *AGO4-D742A* transgene. The ratio of small RNA accumulation in indicated genotypes over that in the Ws wild-type was computed and then log2-transformed. Boxplots show medians (horizontal lines), the 1st-3rd quartile range (boxes), 95% confidence of medians (notches), other data out to 1.5 times the interquartile range (whiskers) and outliers (dots).

Slicing-defective AGO4 partially complements small RNA accumulation in the *ago4-4/ago6-2/ago9-1* triple mutant

AGO4, *AGO6*, and *AGO9* have related but non-redundant functions in gene silencing, and all three can bind 24 nt siRNAs [19]. We obtained the triple mutant *ago4-4/ago6-2/ago9-1* and analyzed small RNA expression levels from inflorescence tissue. Significant ecotype-specific changes in small RNA accumulation levels were observed between Ws (the parental background of the *ago4-4* allele) and Col-0 (the parental background of the *ago6-2* and *ago9-1* alleles) (Figure EV3A). About 15% of the small RNA clusters had significant differential accumulation (FDR = 0.01) when comparing Ws and Col-0 (Figure EV3B). Different small RNA accumulation in these DE clusters was presumably caused by the different genetic backgrounds. We therefore excluded these loci from our analyses.

When analyzing the remaining, ecotype-insensitive small RNA clusters, we observed that the Col-0, Ws, and *ago4-4*/*wtAGO4* samples were tightly grouped (Fig 4A). This demonstrates both the effective removal of clusters that have ecotype-specific differences in accumulation, as well as strong complementation by the *wtAGO4* transgene. While *ago4-4*/*6-2*/*9-1*/*wtAGO4* strongly diverged from *ago4-4*/*6-2*/*9-1*, *ago4-4*/*6-2*/*9-1*/*D742A* showed only minimal differences from *ago4-4*/*6-2*/*9-1* (Fig 4A). This implies that introduction of wild-type *AGO4,* but not a slicing-defective *AGO4,* can rescue much of the small RNA accumulation defects of the triple mutant. Full elimination of AGO4-clade AGOs didn't affect accumulation of the majority of 24 nt siRNA clusters: About 22% (3005/13602) of the 24 nt siRNA clusters in Col-0 were *AGO4*/*AGO6*/*AGO9*-dependent, and these clusters contributed only about 15% of the small RNA reads (Fig 4B). Only about 24% (719/3005) of the *AGO4/AGO6/AGO9*-dependent clusters were not complemented by *wtAGO4* (Fig 4C), indicating that *AGO6* and/or *AGO9* are required for accumulation from relatively few clusters. Similar to the single-mutant analysis (Fig 3), many of the *AGO4/AGO6/AGO9*-dependent clusters were not complemented by *AGO4-D742A* (Fig 4C). In addition, even the set of loci that were designated as complemented by *AGO4-D742A* still generally showed less accumulation then observed with the *wtAGO4* transgene (Fig 4D). Overall, these analyses demonstrate that *AGO4* is required for the accumulation of a much larger number of siRNAs compared to *AGO6* and *AGO9* in inflorescences, and that the slicing activity of AGO4 is required for full accumulation of most of these siRNAs.

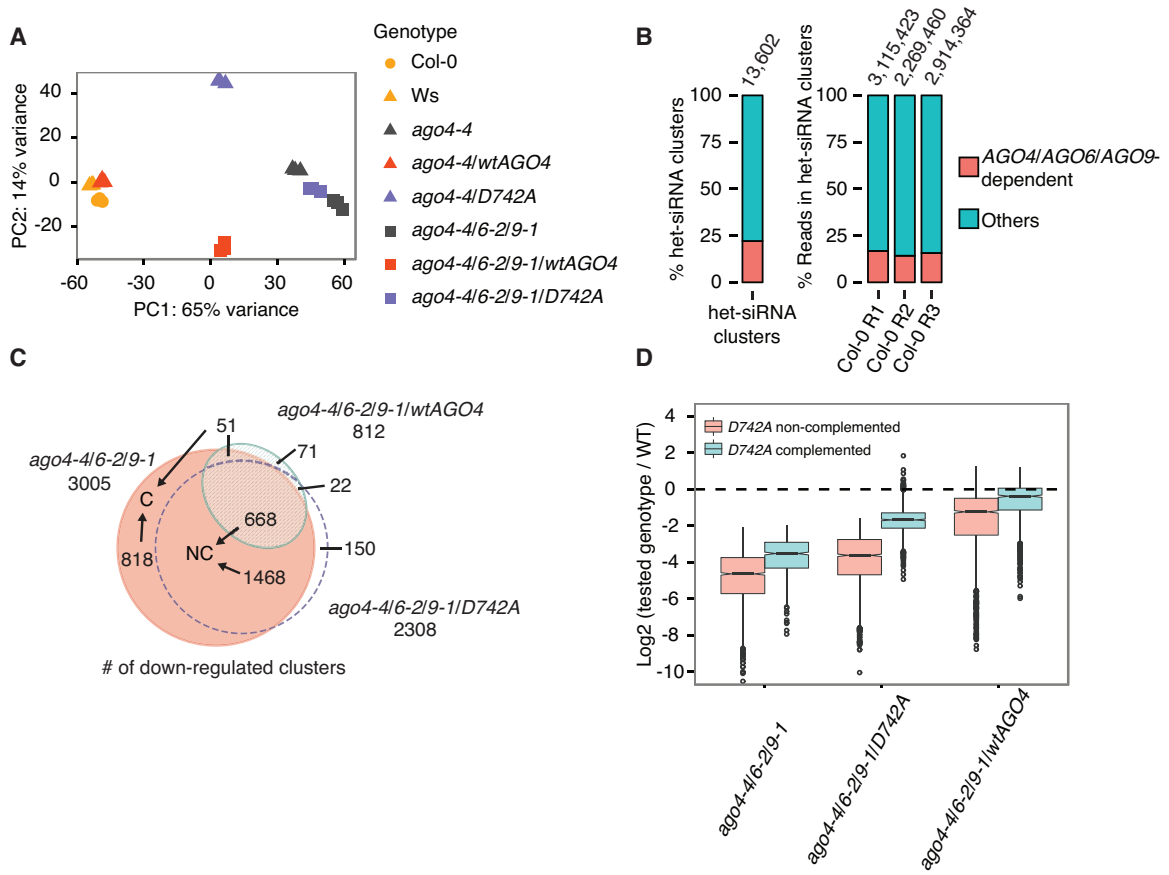

Figure 4. Slicing-defective *AGO4-D742A* partially complements small RNA accumulation in the *ago4-4/6-2/9-1* background

A. Principal component analysis demonstrating overall relationships between all sRNA-seq libraries from ecotype-insensitive small RNA clusters. The first two principal components are shown.

B. Euler diagram showing the number of significantly down-regulated small RNA clusters (FDR=0.01) in indicated genotypes compared with Ws wild-type. C: *AGO4-D742A* complemented clusters in *ago4-4/6-2/9-1* background; NC: *AGO4-D742A* non-complemented clusters in *ago4-4/6-2/9-1* background.

C. Normalized small RNA accumulation in *AGO4-D742A* complemented and non-complemented clusters. The ratio of small RNA accumulation in indicated genotypes over that in the Ws wild-type was computed and then log2-transformed. Boxplots show medians (horizontal lines), the 1st-3rd quartile range (boxes), 95% confidence of medians (notches), other data out to 1.5 times the interquartile range (whiskers) and outliers (dots).

Most 24 nt siRNAs do not require *AGO4, AGO6, or AGO9* for accumulation

AGO4 is required for 24 nt small RNA at some loci, but not others [1,2]. Our genome-wide analysis confirms this observation, and quantifies the extent of the dichotomy: Most 24 nt siRNA loci are unaffected by loss of *AGO4*, while only a small subset have siRNA accumulation defects. Even when all three functional members of the *AGO4* clade [18] are removed in the *ago4-4/6-2/9-1* triple mutant, accumulation of 24 nt siRNAs from most loci is unaffected. This situation seems to contrast to the relationship between the major *Arabidopsis* miRNA-binding Argonaute AGO1 and miRNAs: In the null mutant *ago1-3*, accumulation of the majority of miRNAs is decreased [28,29]. Why might the majority of 24 nt siRNAs maintain stable accumulation levels in the absence of AGO4, AGO6, and AGO9? One possibility is that they are stabilized by AGO3. Despite not being a member of the AGO4 clade, *Arabidopsis* AGO3 is primarily associated with 24 nt siRNAs, and can partially complement the DNA methylation defects seen in the *ago4* mutant [30]. Alternatively, many 24 nt siRNAs might be stabilized by association with non-AGO RNA binding proteins, or perhaps not require protein binding at all.

***AGO4*-dependent siRNAs are likely secondary siRNAs**

Two models have been proposed to explain why some 24 nt siRNAs are dependent on *AGO4*. Qi *et al*. [2] hypothesized that *AGO4*-dependent siRNAs might reflect target slicing-dependent secondary siRNA biogenesis similar to that which is sometimes observed from miRNA targets [31]. In this model, double-stranded RNA could be synthesized from using AGO4-sliced primary transcripts, which are then further processed into 24 nt secondary siRNAs by DCL3. Because Pol V makes chromatin-associated, long non coding RNAs that are targeted by AGO4 [32], the sliced-secondary siRNA model predicts that *AGO4*-dependent siRNAs would also be *NRPE1*-dependent (recall that *NRPE1* encodes the largest sub-unit of Pol V). Our analysis shows that this prediction is supported by the data: most *NRPE1*-dependent siRNA clusters are also *AGO4*-dependent, and *vice-versa*. However, we also found that *AGO4*-dependent siRNAs also tend to be *DRM2*-dependent. This isn't an obvious prediction of the sliced-secondary siRNA model because DRM2, a *de novo* DNA methyltransferase, is thought to be recruited to chromatin in the vicinity of an AGO4-Pol V interaction. An alternative model proposed that an initial wave of *de novo* AGO4/Pol V-dependent DNA methylation at a locus could subsequently recruit Pol IV and

thus produce secondary siRNAs in a self-reinforcing loop [14]. Our observation that *ago4*, *nrpe1*, *drm2,* and *ssh1* were all required for accumulation of the same subsets of 24 nt siRNA loci is fully consistent with the self-reinforcing loop model for secondary het-siRNAs. Intriguingly, much stronger reduction of het-siRNA accumulation was observed in *nrpd1-4* than in *ssh1-1*, suggesting that SSH1 may be specifically required for guiding Pol IV to the regions targeted by self-reinforcing silencing.

On the role of AGO4-catalyzed slicing

A full description of the functions of AGO4-catalyzed endonuclease activity (*e.g.* slicing) remains elusive. In other systems, two general functions of AGO-catalyzed slicing have been described: Slicing of passenger strands during AGO-loading of a small RNA duplex [33], and slicing of target RNAs [34]. For *Arabidopsis* AGO1, both *in vitro* and *in vivo* experiments demonstrate that AGO1-catalyzed slicing is not required for miRNA loading, but is required for many aspects of target regulation [29,35-37]. In contrast, *in vitro* and *in vivo* data have demonstrated that AGO4-catalyzed slicing is required for passenger strand removal during siRNA loading and subsequent nuclear localization of the AGO4-siRNA complex [25]. Although AGO4 can slice a free target RNA *in vitro* [2], to our knowledge there is no direct evidence of AGO4-catalyzed slicing of Pol V target RNAs *in vivo*. Our analysis showed that the catalytic capability of AGO4 is critical for the full accumulation of nearly all *AGO4*-dependent siRNAs. Many siRNAs were not rescued at all by slicing-defective AGO4, and even those that showed some degree of complementation almost never recovered to the extent allowed by complementation with the wild-type AGO4. The dependency of *AGO4*-dependent siRNAs upon AGO4-catalyzed slicing could be fully explained by defects in siRNA loading [25]. In either the sliced-secondary siRNA or self-reinforcement secondary siRNA models, lack of proper loading and subsequent nuclear localization of the 'primary' siRNAs would prevent accumulation of the *AGO4*-dependent sub-population.

Whether or not AGO4-catalyzed slicing occurs at the targeting stage (*e.g.* in the nucleus upon targeted Pol V transcripts) remains unclear. If so, it would seem to present difficulties for the current model of RdDM, which supposes that a stable tethering of AGO4-siRNA complexes to nascent RNAs is required to recruit DRM2 to the vicinity. If slicing is not used at the targeting

stage, the challenge becomes understanding how it is prevented *in vivo*, given that *in vitro* AGO4-siRNA complexes are perfectly competent to direct target cleavage [2]. Resolution of these questions is an important goal for the future that will further illuminate the mechanisms of RdDM.

Materials and Methods

Plant materials and growth condition

All *Arabidopsis thaliana* plants were grown at 21°C with 16 h light/8 h dark. *ago4-4* (FLAG_216G02) was from INRA T-DNA transformants in the Wassilevskija (Ws) ecotype. *ago6-2* (SALK_031553) and *ago9-1* (SALK_127358) were from Salk T-DNA transformants in the Columbia-0 (Col-0) ecotype. The *ago4-4/ago6-2/ago9-1*triple mutant was generated by crossing *ago4-4* to *ago6-2* first, and crossing the *ago4-4/6-2* double mutant to *ago9-1*. Homozygous mutants were selected by genotyping using primers that specifically amplify T-DNA inserted alleles. All the genotyping primers are listed in Table EV1.

Cloning of wild-type and slicing-defective AGO4

cDNA encoding *AGO4* (*AT2G27040*) was amplified from *Arabidopsis thaliana* cDNA in Col-0 ecotype. A *FLAG* tag was inserted at the 5' of *AGO4* cDNA right after start codon by PCR. The *FLAG*-tagged *AGO4* sequence was sub-cloned into the pGII0179 vector. A ~ 2 kb DNA sequence located upstream of the start codon of *AGO4* in Col-0, and a ~ 500 bp DNA sequence downstream of stop codon of *AGO4* in Col-0, were further sub-cloned into *AGO4* expression vector as native promoter and terminator (*pAGO4:FLAG-AGO4*). Mutagenesis of the catalytic motif of *AGO4* was performed by overlapping extension PCR. Primers with desired changes, which encode alanine instead of aspartic acid at the 742th amino acid position of AGO4, were used to introduce slicing defective mutation. The wild-type *AGO4* sequence in *AGO4* expression vector was then swapped by mutagenized *AGO4* to generate slicing-defective *AGO4* expression vector (*pAGO4:FLAG-AGO4-D742A*). The hygromycin-B phosphotransferase gene was inserted into both wild-type and slicing-defective *AGO4* expression vectors for hygromycin resistance selection in transgenic plants. All primers used for subcloning are listed in Table EV1.

Plant transformation and transgenic plant selection

Wild-type or slicing-defective *AGO4* expression vector was introduced into *ago4-4* or the *ago4-4*/*ago6-2*/*ago9-1* background by floral dip with *Agrobacterium tumefaciens* strain *GV3101* bearing the pSOUP plasmid and designated expression vectors. Transgenic plants were selected on 1/2 strength Murashige-Skoog plates supplemented with 15mg/L Hygromycin-B. Independent transgenic lines with single insertion were selected in the T2 generation. Homozygous lines with comparable wild-type or slicing-defective AGO4 protein accumulation in the T3 generation were further selected to prepare sRNA-seq libraries.

sRNA-seq library preparation

Libraries were constructed by using 1µg total RNA extracted from *Arabidopsis* immature inflorescence tissue as described in Wang *et al*. (2016). Three biological replicates from each genotype were prepared. Raw data have been deposited at NCBI GEO under accession number GSE79119 (Col-0 samples) and GSE87333 (all other samples). Details for sRNA-seq libraries are listed in Table EV2

Differential expression analysis

sRNA-seq data sets, including libraries from wild-type AGO4 and slicing-defective AGO4 transgenic lines in *ago4-4* and *ago4-4/ago6-2/ago9-1* background, mutant controls of *ago4-4* and *ago4-4/ago6-2/9-1*, wild-type controls of Col-0 and Ws, were merged and run with ShortStack 3.3 [38] with options --adapter TGGAATTC --mincov 50. All sRNA-seq libraries were aligned to the *Arabidopsis* TAIR10 reference genome.

A matrix of raw read counts from *de novo* annotated small RNA clusters in all three biological replicates of different genotypes were used for differential expression analysis with the R package DESeq2 [39]. Clusters with at least a 2-fold change relative wild-type at a 1% false discovery rate were defined as differentially expressed.

To identify differentially expressed clusters in *nrpe1* compared to Col-0, sRNA-seq data sets from a previous study [27] with three biological replicates of *nrpe1-1* and three biological replicates of Col-0 were analyzed with the same pipeline as described above, except that small

RNA clusters were previously annotated by analyzing the AGO4-related data sets. sRNA-seq libraries used in this analysis are listed in Table EV2.

Heatmap of small RNA accumulation in AGO4-dependent clusters

To generate the heatmap for small RNA accumulation visualization, we first transformed read per million (RPM) data in AGO4-dependent cluters with the equation $E = \log_2(R_i/R_m)$, where E is the input for heatmap, R_i is the RPM of a cluster in a sRNA-seq library, R_m is the mean RPM of a cluster across different sRNA-seq libraries been analyzed for the heatmap. The matrix of transformed RPM was then used for heatmap preparation with the R package pheatmap [40].

Euler diagrams

All Euler diagrams in this study were prepared with eulerAPE 3.0 [41].

Small RNA accumulation in *nrpd1-4*, *nrpe1-12*, *drm2-2* and *ago4-4*

sRNA-seq libraries from a study [26] containing samples from *nrpd1-4*, *nrpe1-12*, *drm2-2, ssh1-1* and Col-0 were aligned to the *Arabidopsis* TAIR10 genome using ShortStack 3.3 [38] with a locifile specifying small RNA clusters which were defined in the AGO4 sRNA-seq data sets. The 3' adapters were removed with the option --adapter TGGAATTC. Before log2 transformation, a value of 0.5 was added to all raw counts. Log2 transformed RPMs of 24 nt siRNA clusters from *nrpd1-4*, *nrpe1-12, drm2-2* and *ssh1-1* as well as Col-0 were plotted to illustrate small RNA accumulation in 24 nt siRNA clusters. Log2 transformed RPMs of high-confidence miRNA genes were also plotted. The linear regression and 95% predicted intervals were calculated based on the distribution of high-confidence miRNA genes. Small RNA accumulation at 24 nt siRNA loci in indicated RdDM mutants was then normalized to corresponding wild-type plants, with equation $N = \log2 (RPM_{mutant}/RPM_{WT})$. Statistical differences between AGO4-dependent and AGO4-independent clusters were tested using the Mann-Whitney U test.

Author contributions

MJA conceived of the project. FW generated transgenic plants, constructed small RNA-seq libraries and performed data analysis. MJA and FW wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

The authors thank Penn State University Genomic Core Facility for small RNA-seq services, and all members of the Axtell Lab for constructive comments.

Funding

US National Science Foundation [1121438 to M.J.A.]; purchase of the Illumina HiSeq2500 used for small RNA-seq was funded by a major research instrumentation award from the US National Science Foundation [1229046 to M.J.A.].

References Cited

1. Zilberman D, Cao XF, Jacobsen SE (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**: 716–719.
2. Qi Y, He X, Wang X-J, Kohany O, Jurka J, Hannon GJ (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**: 1008–1012.
3. Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**: 204–220.
4. Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet* **15**: 394–408.
5. Onodera Y, Haag JR, Ream T, Nunes PC, Pontes O, Pikaard CS (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**: 613–622.
6. Herr AJ, Jensen MB, Dalmay T, Baulcombe DC (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**: 118–120.
7. Blevins T, Podicheti R, Mishra V, Marasco M, Tang H, Pikaard CS (2015) Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. *Elife* **4**: e09591.
8. Zhai J, Bischof S, Wang H, Feng S, Lee T-F, Teng C, Chen X, Park SY, Liu L, Gallego-Bartolome J, et al. (2015) A One Precursor One siRNA Model for Pol IV-Dependent siRNA Biogenesis. *Cell* **163**: 445–455.
9. Xie ZX, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol* **2**: 642–652.
10. Kasschau KD, Fahlgren N, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Carrington JC (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS Biol* **5**: 479–

493.
11. Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, Jacobsen SE (2006) Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet* **38**: 721–725.
12. Wierzbicki AT, Haag JR, Pikaard CS (2008) Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. *Cell* **135**: 635–648.
13. Wierzbicki AT, Ream TS, Haag JR, Pikaard CS (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* **41**: 630–634.
14. Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, Hakimi MA, Lerbs-Mache S, Colot V, Lagrange T (2005) 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.
15. Mosher RA, Schwach F, Studholme D, Baulcombe DC (2008) PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc Natl Acad Sci USA* **105**: 3145–3150.
16. Mallory A, Vaucheret H (2010) Form, Function, and Regulation of ARGONAUTE Proteins. *Plant Cell* **22**: 3879–3889.
17. Fang X, Qi Y (2016) RNAi in Plants: An Argonaute-Centered View. *Plant Cell* **28**: 272–285.
18. Vaucheret H (2008) Plant ARGONAUTES. *Trends in Plant Science* **13**: 350–358.
19. Havecker ER, Wallbridge LM, Hardcastle TJ, Bush MS, Kelly KA, Dunn RM, Schwach F, Doonan JH, Baulcombe DC (2010) The Arabidopsis RNA-Directed DNA Methylation Argonautes Functionally Diverge Based on Their Expression and Interaction with Target Loci. *Plant Cell* **22**: 321–334.
20. Zheng X, Zhu J, Kapoor A, Zhu J-K (2007) Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *EMBO J* **26**: 1691–1701.
21. Duan C-G, Zhang H, Tang K, Zhu X, Qian W, Hou Y-J, Wang B, Lang Z, Zhao Y, Wang X, et al. (2015) Specific but interdependent functions for Arabidopsis AGO4 and AGO6 in RNA-directed DNA methylation. *EMBO J* **34**: 581–592.
22. McCue AD, Panda K, Nuthikattu S, Choudury SG, Thomas EN, Slotkin RK (2015) ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. *EMBO J* **34**: 20–35.
23. Olmedo-Monfil V, Duran-Figueroa N, Arteaga-Vazquez M, Demesa-Arevalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada J-P (2010) Control of female gamete formation by a small RNA pathway in Arabidopsis. *Nature* **464**: 628–U200.
24. Zhu H, Hu F, Wang R, Zhou X, Sze S-H, Liou LW, Barefoot A, Dickman M, Zhang X (2011) Arabidopsis Argonaute10 Specifically Sequesters miR166/165 to Regulate Shoot Apical Meristem Development. *Cell* **145**: 242–256.
25. Ye R, Wang W, Iki T, Liu C, Wu Y, Ishikawa M, Zhou X, Qi Y (2012) Cytoplasmic Assembly and Selective Nuclear Import of Arabidopsis ARGONAUTE4/siRNA Complexes. *Mol Cell* **46**: 859–870.
26. Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, Strahl BD, Patel DJ, Jacobsen SE (2013) Polymerase IV occupancy at RNA-directed DNA methylation sites

requires SHH1. *Nature* **498**: 385–.
27. Lee T-F, Gurazada SGR, Zhai J, Li S, Simon SA, Matzke MA, Chen X, Meyers BC (2012) RNA polymerase V-dependent small RNAs in Arabidopsis originate from small, intergenic loci including most SINE repeats. *Epigenetics* **7**: 781–795.
28. Vaucheret H, Vazquez F, Crété P, Bartel DP (2004) The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Dev* **18**: 1187–1197.
29. Arribas-Hernández L, Kielpinski LJ, Brodersen P (2016) mRNA Decay of Most Arabidopsis miRNA Targets Requires Slicer Activity of AGO1. *Plant Physiol* **171**: 2620–2632.
30. Zhang Z, Liu X, Guo X, Wang X-J, Zhang X (2016) Arabidopsis AGO3 predominantly recruits 24-nt small RNAs to regulate epigenetic silencing. *Nature Plants* **2**: 16049.
31. Fei Q, Xia R, Meyers BC (2013) Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell* **25**: 2400–2415.
32. Wierzbicki AT (2012) The role of long non-coding RNA in transcriptional gene silencing. *Curr Opin Plant Biol* **15**: 517–522.
33. Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**: 607–620.
34. Qi Y, Denli AM, Hannon GJ (2005) Biochemical specialization within Arabidopsis RNA silencing pathways. *Mol Cell* **19**: 421–428.
35. Iki T, Yoshikawa M, Nishikiori M, Jaudal MC, Matsumoto-Yokoyama E, Mitsuhara I, Meshi T, Ishikawa M (2010) In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Mol Cell* **39**: 282–291.
36. Carbonell A, Fahlgren N, Garcia-Ruiz H, Gilbert KB, Montgomery TA, Nguyen T, Cuperus JT, Carrington JC (2012) Functional analysis of three Arabidopsis ARGONAUTES using slicer-defective mutants. *Plant Cell* **24**: 3613–3629.
37. Arribas-Hernández L, Marchais A, Poulsen C, Haase B, Hauptmann J, Benes V, Meister G, Brodersen P (2016) The Slicer Activity of ARGONAUTE1 Is Required Specifically for the Phasing, Not Production, of Trans-Acting Short Interfering RNAs in Arabidopsis. *Plant Cell* **28**: 1563–1580.
38. Johnson NR, Yeoh JM, Coruh C, Axtell MJ (2016) Improved Placement of Multi-mapping Small RNAs. *G3 (Bethesda)* **6**: 2103–2111.
39. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:.
40. Kolde R (2015) *pheatmap: Pretty Heatmaps. R package version 1.0. 2*.
41. Micallef L, Rodgers P (2014) eulerAPE: Drawing Area-Proportional 3-Venn Diagrams Using Ellipses. *PLoS ONE* **9**:.

Expanded view

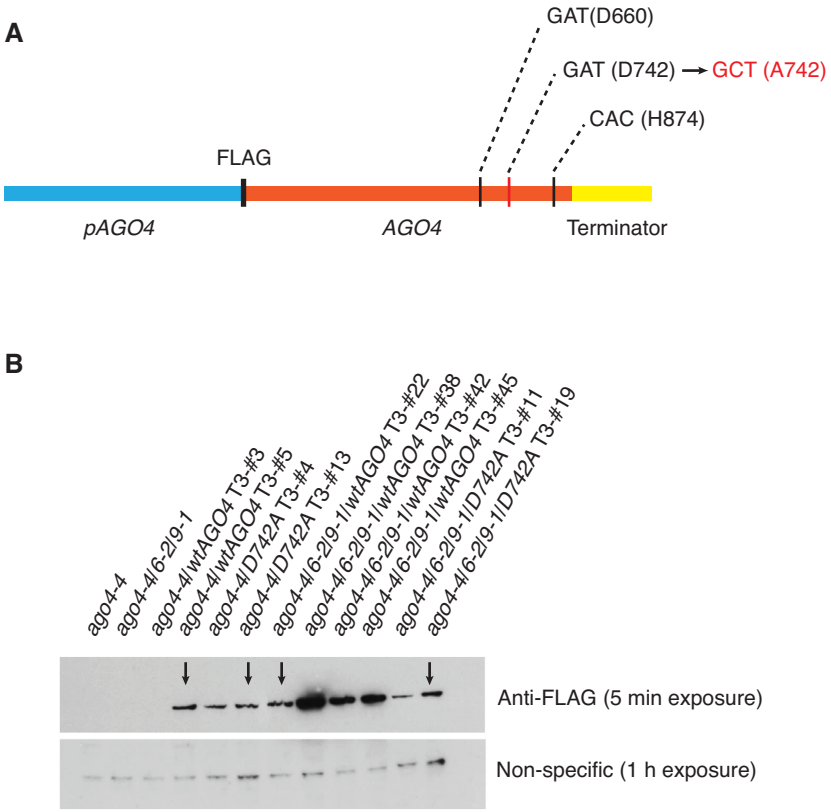

Figure EV1. Expression of wild-type AGO4 and slicing-defective AGO4 proteins in transgenic plants

A. Schematic of transgenes. Indicated codons correspond to the catalytic residues required for slicing. Codon color-coded by red represented the mutagenesis of codon 742.

B. Anti-FLAG immunoblot of FLAG-tagged AGO4 in T3 lines of the indicated transgenic plants. Transgenic lines that were chosen for sRNA-seq library preparation, based on approximately equal accumulation of AGO4 protein, are indicated by arrows.

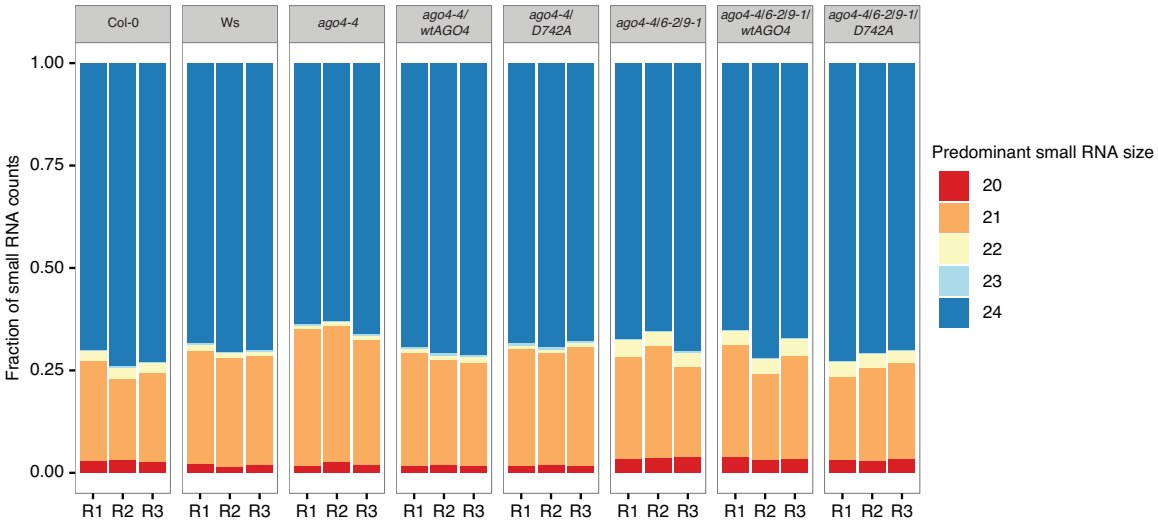

Figure EV2. Overall size profiles of small RNAs in tested genotypes

Fractions of small RNA clusters with different predominant sizes in indicated genotypes are shown. R1, R2 and R3 represent three biological replicates.

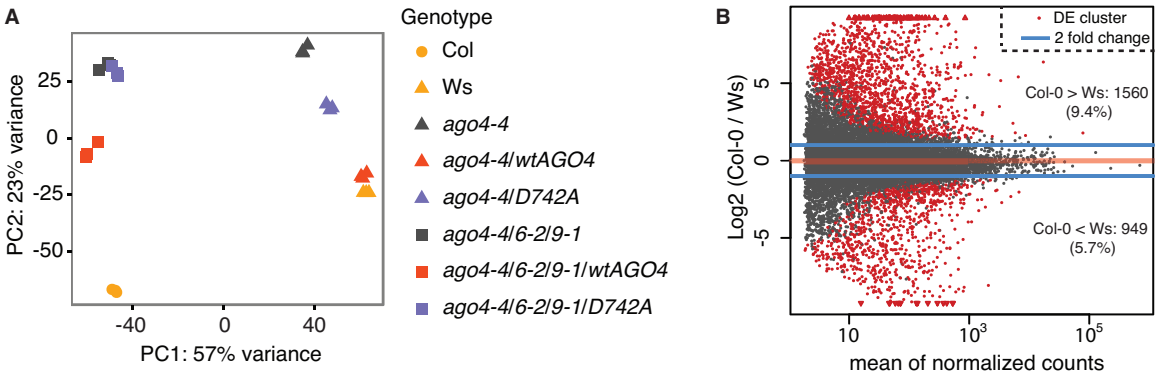

Figure EV3. Divergence of small RNA accumulation between Col-0 and Ws

A. Principal component analysis demonstrating overall relationships between sRNA-seq libraries. The first two principal components are shown for tested samples.

B. MA plot highlighting in red small RNA clusters with at least 2-fold differences (FDR=0.01) between Col-0 and Ws ecotypes.

Table EV1. Primers used in this study

For genotyping		
Primer name	5' to 3' sequence	Annotation
26.101	CGGAAGTAGTCAGCAACTGATACTTC	*AGO4* wild-type allele genotyping forward primer
26.102	CCAATGGGAATGAAAGTCCAA	*AGO4* wild-type allele genotyping reverse primer
20.80	TCTTAGAACGACAATGGTGG	*AGO6* wild-type allele genotyping forward primer
20.81	ACTCTAAGTGCATCCTGAGC	*AGO6* wild-type allele genotyping reverse primer
20.81	ACTCTAAGTGCATCCTGAGC	*ago6-2* mutant allele genotyping forward primer
20.49	GCGTGGACCGCTTGCTGCAAC	*ago6-2* mutant allele genotyping reverse primer
22.50	TTTTTCCTTTTTGCTTGTGGAT	*AGO9* wild-type allele genotyping forward primer
22.51	AACTTGCGTTACTTTTGGCATT	*AGO9* wild-type allele genotyping reverse primer
20.50	TTTTTCCTTTTTGCTTGTGGAT	*ago9-1* mutant allele genotyping forward primer
20.48	TGGTTCACGTAGTGGGCCATC	*ago9-1* mutant allele genotyping reverse primer
For subcloning		
Primer name	5' to 3' sequence	Annotation
pAGO4F	ACCGGGCCCGTTACAAAGCATCCGATCC	*AGO4* promoter forward primer
pAGO4R	CATACTAGTCTCCTGCTCAAAGAAACCAAAC	*AGO4* promoter reverse primer
AGO4F	GAGACTAGTATGGACTACAAGGATGACGATGA CAAGGATTCAACAAATGGTAACGGA	*AGO4* cDNA forward primer
AGO4R	TTATGCGGCCGCTTAACAGAAGAACATGGA	*AGO4* cDNA reverse primer
AGO4-ter-F	TTAAGCGGCCGCATGAGCAGCCCTACTTGGCTC A	*AGO4* terminator forward primer
AGO4-ter-R	AACTCCTAGGCATCAGCTCCAACCAATTTCA	*AGO4* terminator reverse primer
D742AF	CATAATTTTCAGGGCTGGTGTGAGTGAATC	D742A Mutagenesis forward primer
D742AR	GATTCACTCACACCAGCCCTGAAAATTATG	D742A Mutagenesis reverse primer

Table EV2. Data sources and accession numbers of *Arabidopsis thaliana* sRNA-seq libraries

Accession number	Genotype	Ecotype	3'-Adapter (first 8 nt)	Source
GSM2086247	Col-0	Col-0	TGGAATTC	This study
GSM2086248	Col-0	Col-0	TGGAATTC	This study
GSM2086249	Col-0	Col-0	TGGAATTC	This study
GSM2327936	Ws	Ws	TGGAATTC	This study
GSM2327937	Ws	Ws	TGGAATTC	This study
GSM2327938	Ws	Ws	TGGAATTC	This study
GSM2327939	*ago4-4*	Ws	TGGAATTC	This study
GSM2327940	*ago4-4*	Ws	TGGAATTC	This study
GSM2327941	*ago4-4*	Ws	TGGAATTC	This study
GSM2327945	*ago4-4/wtAGO4*	Ws	TGGAATTC	This study
GSM2327946	*ago4-4/wtAGO4*	Ws	TGGAATTC	This study
GSM2327947	*ago4-4/wtAGO4*	Ws	TGGAATTC	This study
GSM2327948	*ago4-4/D742A*	Ws	TGGAATTC	This study
GSM2327949	*ago4-4/D742A*	Ws	TGGAATTC	This study
GSM2327950	*ago4-4/D742A*	Ws	TGGAATTC	This study
GSM2327942	*ago4-4/6-2/9-1*	Col-0/Ws	TGGAATTC	This study
GSM2327943	*ago4-4/6-2/9-1*	Col-0/Ws	TGGAATTC	This study
GSM2327944	*ago4-4/6-2/9-1*	Col-0/Ws	TGGAATTC	This study
GSM2327951	*ago4-4/6-2/9-1/wtAGO4*	Col-0/Ws	TGGAATTC	This study
GSM2327952	*ago4-4/6-2/9-1/wtAGO4*	Col-0/Ws	TGGAATTC	This study
GSM2327953	*ago4-4/6-2/9-1/wtAGO4*	Col-0/Ws	TGGAATTC	This study
GSM2327954	*ago4-4/6-2/9-1/D742A*	Col-0/Ws	TGGAATTC	This study
GSM2327955	*ago4-4/6-2/9-1/D742A*	Col-0/Ws	TGGAATTC	This study
GSM2327956	*ago4-4/6-2/9-1/D742A*	Col-0/Ws	TGGAATTC	This study
GSM1103235	Col-0	Col-0	TGGAATTC	[26]
GSM1103237	*nrpd1-4*	Col-0	TGGAATTC	[26]
GSM1103238	*nrpe1-4*	Col-0	TGGAATTC	[26]
GSM1103240	*drm2-2*	Col-0	TGGAATTC	[26]
GSM1103239	*ssh1-1*	Col-0	TGGAATTC	[26]
GSM893112	Col-0	Col-0	CACTCGGG	[27]
GSM893113	Col-0	Col-0	CACTCGGG	[27]
GSM893114	Col-0	Col-0	CACTCGGG	[27]
GSM893115	*nrpe1-1*	Col-0	CACTCGGG	[27]
GSM893116	*nrpe1-1*	Col-0	CACTCGGG	[27]
GSM893117	*nrpe1-1*	Col-0	CACTCGGG	[27]