A Dicer-Independent Route for Biogenesis of siRNAs that Direct DNA Methylation in *Arabidopsis*

**Graphical Abstract**

**Highlights**
- We identify laddered siRNAs independent of DCLs (sidRNAs)
- sidRNAs interact with AGO4 and direct DNA methylation
- sidRNA production requires distributive 3'-5' exonucleases
- We propose sidRNAs as triggers of de novo DNA methylation

**Authors**
Ruiqiang Ye, Zulong Chen, Bi Lian, ..., Xin-Jian He, Andrzej T. Wierzbicki, Yijun Qi

**Correspondence**
qiyijun@tsinghua.edu.cn

**In Brief**
In plants, DNA methylation can be directed by 24-nt siRNAs that are produced by DCL3. Ye et al. report a distinct class of DCL-independent siRNAs, sidRNAs. sidRNAs interact with AGO4 and require distributive 3'-5' exonucleases for their biogenesis. The authors propose that sidRNAs are the initial triggers of de novo DNA methylation.

**Accession Numbers**
GSE74398
A Dicer-Independent Route for Biogenesis of siRNAs that Direct DNA Methylation in Arabidopsis

Ruiqiang Ye, Zulong Chen, Bi Lian, M. Jordan Rowley, Ning Xia, Jijie Chai, Yan Li, Xin-Jian He, Andrzej T. Wierzbicki, and Yijun Qi

1Tsinghua-Peking Center for Life Sciences, Beijing 100084, China
2Center for Plant Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China
3School of Life Sciences, Peking University, Beijing 100871, China
4National Institute of Biological Sciences, Beijing 102206, China
5Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA
6Co-first author

*Correspondence: qiyijun@tsinghua.edu.cn
http://dx.doi.org/10.1016/j.molcel.2015.11.015

SUMMARY

DNA methylation directed by 24-nucleotide (nt) small interfering RNAs (siRNAs) plays critical roles in gene regulation and transposon silencing in Arabidopsis. 24-nt siRNAs are known to be processed from double-stranded RNAs by Dicer-like 3 (DCL3) and loaded into the effector Argonaute 4 (AGO4). Here we report a distinct class of siRNAs independent of DCLs (sidRNAs). sidRNAs are present as ladders of 23–60 nt in length, often having the same 5’ ends but differing in 3’ ends by 1-nt steps. We further show that sidRNAs are associated with AGO4 and capable of directing DNA methylation. Finally we show that sidRNA production depends on distributive 3’-5’ exonucleases. Our findings suggest an alternative route for siRNA biogenesis. Precursor transcripts are bound by AGO4 and subsequently subjected to 3’-5’ exonucleolytic trimming for maturation. We propose that sidRNAs generated through this route are the initial triggers of de novo DNA methylation.

INTRODUCTION

Cytosine DNA methylation is an epigenetic modification that regulates gene expression, represses transposable elements (TEs), and maintains genome integrity. In plants, DNA methylation is mainly found in TEs and occurs in three sequence contexts: CG, CHG, and CHH (where H is either A, T, or C) (Law and Jacobsen, 2010). Arabidopsis thaliana has evolved specific DNA methyltransferases to maintain methylation in each context. Methylation at CG and CHG sites is primarily by Methyltransferase 1 (MET1, an ortholog of DNMT1 in mammals) and Chromomethylase 3 (CMT3), respectively, whereas the maintenance of CHH methylation is catalyzed by Domain Rearranged Methyltransferase 2 (DRM2, an ortholog of DNMT3 in mammals) and CMT2 (Stroud et al., 2013; Zemach et al., 2013).

The establishment of DNA methylation in all sequence contexts is mediated by a specialized RNAi pathway referred to as RNA-directed DNA methylation (RdDM) (Law and Jacobsen, 2010; Matzke et al., 2009). RdDM entails the transcription of target loci by the plant-specific RNA polymerase IV (Pol IV) (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005). The resulting transcripts (P4RNAs) serve as templates for RNA-dependent RNA polymerase 2 (RDR2) to make double-stranded RNAs (dsRNAs) (Haag et al., 2012; Xie et al., 2004) that are processed by Dicer-like 3 (DCL3) into 24-nt small interfering RNAs (siRNAs) (Blevins et al., 2015; Qi et al., 2005; Xie et al., 2004; Zhai et al., 2015). These Pol IV-dependent siRNAs (P4siRNAs) are loaded into Argonaute 4 (AGO4) (Li et al., 2006; Qi et al., 2005; Xie et al., 2004; Zhai et al., 2015). Likely through base-pairing between siRNAs and nascent scaffold transcripts generated by another plant-specific RNA polymerase, Pol V (Wierzbicki et al., 2008), AGO4/siRNA complexes are recruited to target loci to direct DRM2-catalyzed DNA methylation (Wierzbicki et al., 2009; Zhong et al., 2014).

In addition to DCL3, Arabidopsis contains three other DCLs for generating distinct small RNA (sRNA) species. DCL1 acts in the maturation of 21-nt microRNAs (miRNAs) or siRNAs from hairpin-structured precursors (Kurihara and Watanabe, 2004; Rogers and Chen, 2013), DCL2 functions primarily in the production of 22-nt viral siRNAs (vsiRNAs) (Blevins et al., 2015; Qi et al., 2005; Xie et al., 2004; Bouché et al., 2006; Bouché et al., 2006; Xie et al., 2004), whereas DCL4 mainly acts in the biogenesis of 21-nt trans-acting siRNAs (ta-siRNAs) (Gasciolli et al., 2005; Xie et al., 2005; Yoshikawa et al., 2005). DCL2, DCL3, and DCL4 also function partially redundantly in P4siRNA production as well as the establishment and maintenance of DNA methylation (Henderson et al., 2006).

In this study, we identified a class of siRNAs that are independent of DCLs (named “sidRNA” for siRNAs independent of DCLs), sidRNAs mainly originate from transposons and intergenic sequences as well as transgenes. sidRNAs interact with and recruit AGO4 to target loci to direct DNA methylation.
Intriguingly, sidRNAs require 3'-5' exonucleases for their production. Our study reveals an unexpected siRNA biogenesis pathway that may act in the initiation of RdDM.

RESULTS

Genome-wide Identification of Small RNAs that Are Independent of DCL Proteins in Arabidopsis Seeds

The biogenesis of all sRNA species discovered so far in plants requires DCL proteins. In an attempt to examine whether there are DCL-independent siRNAs in Arabidopsis, we gel-purified RNAs between 18 and 30 nt from 3-week-old seedlings of the wild-type (Col-0), the triple mutant dcl2-1 dcl3-1 dcl4-2 (referred to as dcl2/3/4 hereafter) carrying null alleles of DCL2, DCL3, and DCL4 (Henderson et al., 2006), and the quadruple mutant dcl1-9 dcl2-1 dcl3-1 dcl4-2 (referred to as dcl1/2/3/4 hereafter) carrying an additional hypomorphic allele of DCL1 (Jacobsen et al., 1999; Figure S1A) and prepared cDNA libraries for deep sequencing on an Illumina high-throughput sequencing platform (Table S1).

Analyses of sequenced reads revealed that miRNAs were reduced by ~98% in abundance in dcl1/2/3/4 relative to Col-0 or dcl2/3/4 and that ta-siRNAs were reduced by ~99% in abundance in both dcl2/3/4 and dcl1/2/3/4 relative to Col-0 (Figure S1B), consistent with the known roles of DCL1 and DCL4 in processing miRNAs and ta-siRNAs, respectively (Kurihara and Watanabe, 2004; Xie et al., 2005). Intriguingly, considerable amounts of other siRNAs, including P4siRNAs, were still detected in dcl2/3/4 and dcl1/2/3/4 (Figure S1B). This was surprising because the biogenesis of P4siRNAs is known to be dependent mainly on DCL3 and partially on DCL2 and DCL4 (Henderson et al., 2006). To examine the origins of their biogenesis, after the removal of miRNAs and ta-siRNAs, siRNAs obtained from dcl1/2/3/4 were mapped to the Arabidopsis nuclear genome, and 14,360 siRNA-producing loci were identified using the following criteria: length \( \geq 100 \) nt; number of unique sRNAs \( \geq 5 \), each separated from the nearest neighbors by a maximum of 50 nt; and expression level \( \geq 15 \) reads per kilobase per million (RPKM) (Figure 1A; Table S2). These loci were mainly located in annotated TEs and repeats (74.4%) as well as intergenic regions (14.5%) (Figure 1A). Distinct from the siRNAs in Col-0 that are predominantly 24 nt in length, siRNAs in dcl1/2/3/4 and dcl2/3/4 exhibited a relatively even size distribution ranging from 18–30 nt, with a mild peak at 21 nt (Figure 1B).

This suggests that the detected siRNAs were probably not produced by DCLs because DCLs are known to generate sRNAs of diverse sizes (Qi et al., 2005; Xie et al., 2004). Moreover, in contrast to the dramatic reduction of miRNA accumulation in dcl1/2/3/4 relative to dcl2/3/4, siRNAs generated from these loci accumulated at comparable levels in dcl2/3/4 and dcl1/2/3/4 (Figure 1C), suggesting that they were not products of the residual DCL1 activity in dcl1/2/3/4. We therefore named them sidRNAs.

Further analyses indicated that total siRNAs produced from the sidRNA-generating loci in Col-0 were about five times more than those in dcl2/3/4 and dcl1/2/3/4 (Figure 1D). Interestingly, siRNAs of 25 nt or longer that very likely represent the production of sidRNAs were also detected in Col-0, although at much lower levels than those in dcl2/3/4 and dcl1/2/3/4 (Figure 1D). Most intriguingly, considerable amounts of sidRNAs produced from each locus appeared as ladders in size, often having the same 5' ends but differing in 3' ends by single-nucleotide steps (Figures 1E and 1F). Northern blot analysis confirmed that sidRNAs generated from several representative loci (AtREP2, SIMPLEHAT2, AtSN1, and At4TE27090) over-accumulated in dcl2/3/4 and dcl1/2/3/4 and displayed a wide size distribution between 20 and 60 nt, with most being in the range of 30–40 nt (Figure 1G). RNA species of 30–60 nt were also detected in dcl2/3/4 and dcl1/2/3/4 by deep sequencing analysis (Figure S1C).

Effects of Pol IV/RDR2 Mutations on sidRNA Production

A comparison of profiles of siRNAs in dcl1/2/3/4 and P4siRNAs in Col-0 revealed that over 85% of the siRNA loci overlapped with the loci that produced P4siRNAs in Col-0 (Figure 2A). The generation of P4siRNAs relies on coupled actions of Pol IV and RDR2 (Haag et al., 2012). To examine whether sidRNA production is also dependent on Pol IV and RDR2, we introduced a null mutation of NRPD1 (encoding the largest subunit of Pol IV) or RDR2 into the dcl2/3/4 mutant. Considering that dcl2/3/4 produces nearly equal amounts of siRNAs as dcl1/2/3/4 and that it is more convenient to combine mutations in dcl2/3/4 (Figures 1C, 1D, and 1G), we used dcl2/3/4 instead of dcl1/2/3/4 in the following studies. The deep sequencing results showed that sidRNA production at 1,229 loci was independent of NRPD1 and RDR2, whereas that at 12,992 loci was dependent on both NRPD1 and RDR2 (Figure 2B; Table S3), and dysfunction of either NRPD1 or RDR2 resulted in a severe reduction of sidRNA production at these loci (Figure 2C; Table S3). Intriguingly, the vast majority of the NRPD1/RDR2-dependent sidRNA loci were located in repeat regions, and only 7% of them resided in genic regions, whereas 38% of NRPD1/RDR2-independent sidRNA loci were located in genic regions (Figure 2D). Northern blot results confirmed that mutation in Pol IV or RDR2 had a dramatic effect on the generation of sidRNAs at the representative loci AtREP2 and SIMPLEHAT2 (Figure 2E). Expression of wild-type NRPD1, but not of an active site mutant (ASM) of NRPD1 (Haag et al., 2009), was able to restore the sidRNA levels in dcl2/3/4 (Figure 2E), further indicating that functional NRPD1 is essential for sidRNA generation at these loci. We noticed, however, that low levels of sidRNAs could still be detected at many of the NRPD1/RDR2-dependent sidRNA loci in nrpd1 dcl2/3/4 or rdr2 dcl2/3/4 and that they appear to have a strong strand bias (Figure 2F).

Detection of sidRNAs in Other Systems

The existence of siRNAs in Arabidopsis seedlings prompted us to determine whether sidRNAs are also prevalent in other systems. First, Arabidopsis endosperm is a specialized tissue that supplies nutrients for the developing embryo. There exists a large population of P4siRNAs in Arabidopsis endosperm (Mosher et al., 2009). These P4siRNAs are predominantly, if not exclusively, derived from the maternal allele, and the expression of a subset of these siRNAs is restricted to flowers and young siliques (Mosher, 2010; Mosher et al., 2009). Through deep sequencing of endosperm and embryo tissues of Col-0,
Second, the production of siRNAs can be induced by actively transcribed transgenes (Beclin et al., 2002; Mourrain et al., 2000). To explore whether sidRNA production can be detected in a transgene system, we transformed Arabidopsis with a 35S:GUS construct and monitored the production of sidRNAs derived from the transgenic region over six generations by deep sequencing (Table S1). sidRNAs expressed in the same ladder-like pattern were detected in each generation of transgenic lines (Figures S3A–S3C), suggesting that, like endogenous loci, transgenic loci can also produce sidRNAs. Intriguingly, Pol IV was dispensable for the generation of sidRNAs at the transgenic loci (Figures S3A and S3B), hinting that another RNA polymerase, presumably Pol II, is responsible for transcribing these sidRNAs.
Figure 2. Effects of Pol IV/RDR2 Mutations on sidRNA Production

(A) Venn diagram showing the overlap between sidRNA loci in dcl1/2/3/4 and P4siRNA loci in Col-0.
(B) Pie chart showing the numbers of sidRNA loci dependent on NRPD1 and/or RDR2.
(C) Heatmap of the abundance of sRNAs generated from sidRNA loci in the indicated plants.
(D) Pie chart summarizing the numbers of NRPD1/RDR2-dependent and -independent sidRNA loci in the indicated categories.
(E) Detection of sRNA production at representative sidRNA loci in the indicated plants by northern blot. SS rRNAs stained with ethidium bromide (EtBr) were used as loading controls.
(F) Box plots showing the strand bias of sidRNAs in the dcl2/3/4 and nrpd1 dcl2/3/4 mutants. The strand bias value was calculated as the absolute (abs) percentage differences between sRNAs reads produced from the positive (pos) and negative (neg) strands of sidRNA loci.

See also Table S3.
Third, active transposons often trigger the production of
sRNAs that direct the establishment of DNA methylation and
inactivation of their transposition (Ito, 2012; Kim and Zilberman,
2014). To determine whether sidRNAs can be derived from
active transposable elements, we took advantage of an epige-
etic recombinant inbred line (epl15) in which a retrotransposon
called EVD became active and invasive from the eighth genera-
tion (F8) of inbreeding because of the loss of DNA methylation at
its 5’ long terminal repeat (LTR) region (Mari-Ordoñez et al.,
2013). DCL-dependent siRNAs that are 21–22 nt or 24 nt long
were found to be produced to initiate DNA methylation and sup-
press EVD mobilization from F8 to F14 (Mari-Ordoñez et al.,
2013). Analyses of the published sRNA datasets (Mari-Ordoñez et al.,
2013) revealed that high levels of laddered sidRNAs were also detected in the F11 and F14 generations (Figures S3D–S3F). The generation of sidRNAs derived from the 3’ gag region appeared to precede the production of sidRNAs tran-
scribed from the 5’ LTR (Figure S3E), coinciding with the 3’-5’
spreading of DNA methylation.

sidRNAs Are Associated with AGO4
In light of the prevalence of sidRNA expression, we asked
whether sidRNAs could be functional via associating with AGO
proteins. Because the loci generating sidRNA and 24-nt sRNA
largely overlap, and the majority of sidRNAs depend on Pol IV
and RDR2 for their biogenesis, we hypothesized that a large pro-
portion of sidRNA may be channeled to binding with AGO4. To
test this, we immunoprecipitated AGO4 from transgenic lines
expressing N-terminally GFP-tagged AGO4 (GFP-AGO4) in the
Col-0 and dcl2/3/4 background and purified sRNAs associated
with GFP-AGO4. Deep sequencing of purified sRNAs revealed
that AGO4 predominantly bound 24-nt siRNAs in Col-0 but bound
sidRNAs that were mainly 21–26 nt in length in dcl2/3/4 (Fig-
ure 3A). A considerable fraction of AGO4-associated sidRNAs
generated from each locus also appeared as ladders (Figures
3B and 3C), although to a lesser extent compared with the
bulk of sidRNAs (Figures 1E and 1F). Confirming the deep
sequencing results, northern blot analysis revealed that AGO4
bound sidRNAs at AtREP2 and SIMPLEHAT2 loci, although the
sizes of AGO4-associated sidRNAs were smaller compared with
those detected in total extract in dcl2/3/4 (Figure 3D). We
speculate that the longer forms of RNA species are precursors
of mature sidRNAs associated with AGO4. It is noteworthy that
a small fraction of presumed sidRNA precursors was also de-
tected in AGO4 immunoprecipitates (Figure 3D), suggesting
that these transcripts are loaded onto AGO4 before being pro-
cessed into mature sidRNAs. In summary, our findings suggest
that AGO4 is an effector protein of sidRNAs.

sidRNAs Are Capable of Directing DNA Methylation
The association of sidRNAs with AGO4 prompted us to examine
whether sidRNAs are capable of guiding AGO4 to target loci. To
this end, we performed chromatin immunoprecipitation (ChIP)-
qPCR to assess the occupancy of AGO4 at several representa-
tive loci in Col-0 as well as in mutants, including dcl2/3/4, nrpd1
dcl2/3/4, and ago4. As expected, AGO4 was enriched at these loci in Col-0 (Figure 4A). However, locus-specific AGO4 occupancy in dcl2/3/4 was reduced moderately but not
eliminated (Figure 4A), suggesting that factors other than 24-
nt sRNAs, presumably sidRNAs, are involved. Indeed, AGO4
occupancy was decreased further by mutation of NRPD1, which
is important for sidRNA biogenesis at these loci, and by
mutation of AGO4, which binds sidRNAs (Figure 4A). Therefore,
our data suggest that sidRNAs are capable of targeting AGO4 to
chromatin. We next examined whether sidRNAs play a role in directing DNA
methylation. By performing locus-specific bisulfite sequencing,
we found that nrpd1 dcl2/3/4 and rdr2 dcl2/3/4, mutants defi-
cient in both 24-nt siRNA and sidRNA biogenesis, exhibited
greater defects in DNA methylation than dcl1/2/3/4 and dcl2/3/4
(Figure 4B), suggesting that sidRNAs play a role in directing
DNA methylation. Although the extent of reduction in DNA
methylation caused by loss of sidRNAs appeared to vary in a lo-
cus-specific manner for all loci examined, DNA methylation in the
CHG and CHH contexts was affected more severely than that in
the CG context (Figure 4B). To investigate the effect of sidRNA
deficiency on DNA methylation genome-wide, we carried out whole-genome bisulfite sequencing and found that the overall CHG and CHH DNA methylation levels in \( \text{nrd}1 \text{ dcl}2/3/4 \) and \( \text{rdr}2 \text{ dcl}2/3/4 \) are significantly lower than those in \( \text{dcl}1/2/3/4 \) and \( \text{dcl}2/3/4 \) (Figures 4C and 4D). The level of CG methylation was reduced slightly by \( \text{NRPD}1 \) or \( \text{RDR}2 \) mutation (Figures 4C and 4D). Here our findings are in agreement with previous methylome surveys showing that \( \text{dcl}2/3/4 \) is a mutant causing reduced CHH methylation, whereas \( \text{nrd}1 \) and \( \text{rdr}2 \) are mutants leading to eliminated CHH methylation (Stroud et al., 2013). Taken together, our data suggest that sidRNAs primarily contribute to CHG and CHH methylation.

**Figure 4.** sidRNAs Are Capable of Directing DNA Methylation

(A) Detection of AGO4 occupancy at the indicated loci in Col-0 and the indicated mutants by ChIP analysis. AGO4 ChIP signals in the mutants are presented in relation to those in Col-0 (arbitrarily set to 1.0). Error bars indicate standard deviations of three biological replicates. Asterisks indicate a significant difference between the indicated groups (*p < 0.05, t test).

(B) Analysis of DNA methylation at the indicated loci in Col-0 and the indicated mutants by bisulfite sequencing. Presented is the overall percent methylation of cytosine sites in different sequence contexts. More than 20 clones were sequenced for each sample.

(C) Box plots of CG, CHG, and CHH methylation levels at sidRNA loci in Col-0 and the indicated mutants. Asterisks indicate a significant difference between the indicated groups (*p < 10\(^{-15}\), Mann-Whitney U test).

(D) Heatmap of CG, CHG, and CHH methylation levels at sidRNA loci in Col-0 and the indicated mutants. See also Tables S5 and S6.
Probing the Mechanism Underlying sidRNA Biogenesis

The ladder-like expression pattern of sidRNAs is reminiscent of the products of distributive 3’-5’ exonucleases. Distributive exonuclease are known to cleave, release, and engage another substrate, creating laddered products (Deutscher and Li, 2001).

We therefore hypothesized that the biogenesis of sidRNAs might involve binding of the 5’ end of sidRNA precursor transcripts to AGO4 and subsequent trimming from the 3’ end by a 3’-5’ distributive exonuclease (Figure 5A).

In accordance with this hypothesis, we found that the sidRNA level in dcl2/3/4 was reduced dramatically by mutation of ago4, suggesting the involvement of AGO4 in sidRNA production (Figure 5B). Moreover, transgenic expression of wild-type AGO4, but not a mutant AGO4 (AGO4\textsuperscript{YFAA}) defective in sRNA binding (Ye et al., 2012), was able to restore the generation of sidRNAs in ago4 dcl2/3/4 (Figure 5B), suggesting that the sRNA binding activity of AGO4 is important for sidRNA generation. Locus-specific examination of sidRNA production by northern blot also revealed that sidRNA generation required AGO4 and its sRNA binding activity (Figure 5C). It is known that some components acting at the effector stage of the RdDM pathway are also required for siRNA production, likely through a feedback regulation. We cannot exclude the possibility that the role of AGO4 in sidRNA production could also be indirect.

To gain insights into the mechanism of sidRNA biogenesis, we investigated the features of sidRNAs. We first examined whether sidRNAs have 5’ triphosphate or monophosphate ends. Total RNAs or AGO4-bound RNAs were prepared from Col-0 and dcl2/3/4 and treated with a 5’ to 3’ exonuclease Terminator that specifically degrades RNAs with 5’ monophosphate ends. ssRNAs generated from the AtREP2 locus were detected by northern blot. A 21-nt spike-in RNA with 5’ monophosphate was added in the reactions and used as a positive control.

We then examined the sensitivity of sidRNAs to RNase T1. sRNAs in total extracts and AGO4 immunoprecipitates prepared from Col-0 and dcl2/3/4 were treated with RNase T1, and AtREP2 ssRNAs were probed by northern blot. Spike-in RNAs, including a 50-bp double-stranded RNA with a 2-nt 3’ overhang (ds) and a 50-nt ssRNA (ss), were added in the reactions and used as controls for RNase T1 digestion.
Figure 6. Distributive 3’-5’ Exonucleases Are Required for siRNA Accumulation and DNA Methylation

(A) Identification of distributive 3’-5’ exonucleases required for DNA methylation. Diagrams show the structure of Atrimmer1/RRP6L1 and Atrimmer2 (top). The positions of residues forming the conserved DEDD-Y active site are indicated. Bottom: DNA methylation levels at the AtSN1 and IGN5 loci in the indicated plants as measured by Chop-PCR. The methylation-sensitive restriction enzyme HaeIII was used to specifically cleave unmethylated DNA in the CHH context. A fragment of Actin that lacks HaeIII sites was amplified as a control. See Figure S4 for Chop-PCR results from all tested 3’-5’ exonuclease mutants.

(legend continued on next page)
Distributive 3'-5' exonucleases are required for sidRNA accumulation and DNA methylation

As proposed above, the production of sidRNAs needs 3'-5' distributive exonucleases for trimming. We therefore set out to identify the exonucleases that are required for sidRNA biogenesis. Because sidRNAs play a role in CHG and CHH DNA methylation, we attempted to search for candidate exonucleases involved in sidRNA production by performing methylation-specific Chop-PCR, a rapid way to examine CHH methylation levels at representative loci (Zhang et al., 2014b). Of 13 mutants carrying null mutations of 3'-5' distributive exonuclease genes, two mutants, with mutations mapped to At1g54440 and At1g56310, respectively, had reduced CHH methylation levels at IGN5 (Figure 6A; Figure S4). At1g54440 encodes ATRRPR6L1, a nuclear subunit that associates with the exosome core complex. ATRRPR6L1 has been reported to function in the processing of long non-coding RNAs (lncRNAs) involved in epigenetic silencing (Shin and Chekmanova, 2014). It also participates in the retention of Pol V-transcribed scaffold lncRNAs through a pathway independent of exosome-mediated RNA degradation (Zhang et al., 2014a). At1g56310 encodes an uncharacterized exonuclease. Considering their possible involvement in trimming sidRNAs, we named ATRRPR6L1 and the uncharacterized ribonuclease Atrimmer1 and Atrimmer2, respectively. Both Atrimmer1 and Atrimmer2 contain a conserved 3'-5' exonuclease domain (Figure 6A).

To better investigate the functions of Atrimmer1 and Atrimmer2 in sidRNA biogenesis, we carefully examined the change of DNA methylation levels at sidRNA-generating loci in several mutants. The CHH methylation levels at IGN5 and AtSN1 in nrpd1 and nrpe1 (NRPE1 encodes the largest subunit of Pol V) are lower than those in dcl2/3/4 (Figure 6A), suggesting that IGN5 and AtSN1 are sidRNA-generating loci. Surprisingly, the extents of reduction in IGN5 CHH DNA methylation levels in atrimmer1 and atrimmer2 were greater than that in dcl2/3/4 and comparable with those in nrpd1 and nrpe1 (Figure 6A), indicating that Atrimmer1 and Atrimmer2 dysfunction likely abolished DCL-dependent DNA methylation as well as sidRNA-mediated DNA methylation. One explanation for our results could be that sidRNA biogenesis is required for DCL-dependent biogenesis of 24-nt siRNAs.

To further explore whether Atrimmer1 and Atrimmer2 are required for sidRNA biogenesis genome-wide, we performed sRNA deep sequencing analysis with atrimmer1 and atrimmer2 mutants to identify Atrimmer1- and/or Atrimmer2-dependent sidRNA loci. We found that Atrimmer1 and Atrimmer2 regulated the production of sidRNAs at 715 and 107 loci, respectively, with 41 loci being regulated by both (Figure 6B). In atrimmer1 and atrimmer2, the levels of sidRNAs at the regulated loci were reduced dramatically (Figure 6C). To determine whether Atrimmers regulate sidRNA production and eventually regulate DNA methylation, we chose several loci with sidRNA production impaired by Atrimmer dysfunction (Figures 6D and 6E) and performed locus-specific bisulfite sequencing. Importantly, we found that dysfunction of Atrimmer1 or Atrimmer2 caused a further reduction of DNA methylation levels at some loci compared with dysfunction of DCLs. CHG and CHH DNA hypomethylation was always detected at loci where sidRNAs were produced at low levels in atrimmer1 and atrimmer2 (Figure 6F). Our data suggest that Atrimmer1 and Atrimmer2 promote locus-specific DNA methylation through regulating sidRNA production. Taken together, we identified Atrimmer1 and Atrimmer2 as potential trimmers of sidRNAs at a subset of loci.

**DISCUSSION**

Our study demonstrates the existence of a distinct class of sRNAs (sidRNAs) that are generated via an alternative route independent of DCLs in Arabidopsis. The biogenesis of sidRNAs likely entails 5'-end binding of sidRNA precursor transcripts to AGO4 followed by trimming by 3'-5' exonucleases (Atrimmers). Intriguingly, DNA methylation mediated by sidRNAs appears to be required for the canonical RdDM. We propose that sidRNAs are the initial triggers of RdDM.

---

(B) Venn diagram showing the numbers of sidRNA loci affected by atrimmer1/rrp6l1, atrimmer2, or both.
(C) Box plots of the levels of sidRNAs generated from Atrimmer1-affected (left) and Atrimmer2-affected (right) sidRNA loci in the indicated plants.
(D) Relative levels (shown in RPM) of representative Atrimmer1/2-controlled sidRNA loci in the indicated plants as measured by deep sequencing.
(E) Detection of sRNA production at representative sidRNA loci in the indicated plants by northern blot. SS rRNAs stained with ethidium bromide were used as loading controls.
(F) Analysis of DNA methylation at the indicated loci in Col-0 and the indicated mutants by bisulfite sequencing. Presented is the overall percent methylation of cytosine sites in different sequence contexts. More than 20 clones were sequenced for each sample. See also Figure S4 and Table S7.
sidRNA Biogenesis

In this study, we identified 14,360 loci that produce sidRNAs in Arabidopsis seedlings (Figure 1A). The majority of the identified sidRNA loci overlap with P4siRNA loci (Figure 2A), and dysfunction of NRPD1 or RDR2 largely reduced sidRNA production (Figures 2C and 2E), suggesting that Pol IV/RDR2 transcripts are precursors of sidRNAs at these loci. This is also supported by the fact that DNA regions producing Pol IV/RDR2-dependent transcripts in dcl2/3/4 strongly overlap with P4siRNA loci (Li et al., 2015).

RDR2 dysfunction, which was thought to cause the accumulation of single-stranded nascent Pol IV transcripts, surprisingly resulted in a dramatic loss of sidRNA signals (Figures 2C and 2E). A previous study has also shown that RDR2 has the same effect on the production of Pol IV-dependent transcripts, as does Pol IV (Li et al., 2015). One possible explanation is that RDR2 may play a role in stabilizing Pol IV transcripts and that, when RDR2 binds to Pol IV transcripts, the stabilization facilitates the conversion of Pol IV transcripts into dsRNAs. However, in light of the weak Pol IV transcriptional activity but robust RDR2 activity (Haag et al., 2012), we favor another possibility: that RDR2 converts single-stranded Pol IV transcripts into dsRNAs that are then unwound and separated into ssRNAs. These ssRNAs then serve as RDR2 substrates and undergo exponential amplification by RDR2. In this case, a large amount of single-stranded sidRNA precursors are generated, and RDR2 dysfunction leads to loss of most, if not all, RNA signals. This process resembles the amplification of single-stranded primordial RNA (priRNA) precursors by an RNA-dependent RNA polymerase complex (RDRC) in fission yeast (Halic and Moazed, 2010) and replication of ssRNA viruses from limited amounts of viral genomic RNAs by viral RNA-dependent RNA polymerases (RdRPs) (Anquist, 2002).

Nascent transcripts of Pol IV/RDR2 are expected to have a 5′ triphosphate. However, recent studies have shown that Pol IV/RDR2 transcripts bear 5′ monophosphate (Li et al., 2015; Zhai et al., 2015). We found that sidRNAs and their precursors also bear 5′ monophosphate (Figure 5D), suggesting that Pol IV/RDR2 transcripts are likely processed by a mechanism yet to be identified before they are loaded onto AGO4. The absence of double-stranded sidRNA precursors in the AGO4 immunocomplex (Figure 5E) suggests that sidRNA precursors are loaded onto AGO4 in single-stranded forms or in double-stranded forms but that one of the strands is quickly dissociated from the complex. The anchoring of single-stranded sidRNA precursors then facilitates the trimming process as discussed below.

The majority of AGO4-associated sidRNAs are shorter than sidRNAs in total cellular extract (Figure 3), suggesting that sidRNAs are processed further upon anchoring to AGO4. The laddered pattern of AGO4-associated sidRNAs is in favor of an evolutionarily conserved trimming mechanism involving the actions of 3′-5′ distributive exonucleases. Exonucleases have been found to participate in RNAi, transposon silencing, and heterochromatin formation in multiple species. For instance, in fission yeast, RNA degradation products associate with AGO1 to recruit Triman for their 3′ end processing, resulting in priRNAs that initiate heterochromatin formation (Halic and Moazed, 2010; Marasovic et al., 2013). In Neurospora, the exonuclease QIP mediates 3′-5′ trimming and maturation of miRNA-like sRNAs (Xue et al., 2012). The exonuclease MUT-7 has been reported to interact with a protein complex that generates sRNAs and is therefore required for RNAi and transposon silencing in worms (Ketting et al., 1999; Tops et al., 2005). In the silkworm, primary PIWI-interacting RNA (piRNA) precursors are trimmed to the mature piRNA size by the activity of an unknown 3′-5′ exonuclease (Kawaoka et al., 2011). In this study, we showed that two 3′-5′ exonucleases, Atrimmer1 and Atrimmer2, are required for sidRNA generation in Arabidopsis (Figure 6). It appeared that Atrimmer1 and Atrimmer2 merely regulate sidRNA generation at ~5% of the sidRNA loci we identified. Because our screening is based on the methylation status of two loci, exonucleases responsible for sidRNA generation at other loci were very likely neglected. Also, functional redundancies probably hindered the isolation of more Atrimmer mutants. It is very likely that there are other exonucleases that control sidRNA generation at different loci. It will be interesting to identify these exonucleases in the future.

Loss of Atrimmers unexpectedly resulted in diminished sidRNA production instead of accumulation of longer sidRNA precursors (Figure 6E), suggesting that, in addition to performing their trimming functions, Atrimmers may also stabilize sidRNA precursors. Without Atrimmers, sidRNA precursors would dissociate from AGO4 and undergo degradation. Although Atrimmer1/RRP6L1 is a nuclear auxiliary subunit of core complex, the trimming function of Atrimmer1 seems to be independent of an exosome function because depletion of two exosome subunits (RRP4 and RRP41) caused neither defective sRNA accumulation nor defects in DNA methylation (Shin et al., 2013; Zhang et al., 2014a).

In addition to Pol IV/RDR2-dependent sidRNA, sidRNA production at 1,229 loci appears to be independent of NRPD1 and RDR2 (Figures 2B and 2C). We also found that low levels of sidRNAs were still produced from many of the Pol IV/RDR2-dependent loci in the absence of Pol IV or RDR2 and that they are mainly derived from one of the DNA strands (Figure 2F). In light of the fact that Pol II-dependent RNAs with a strong strand bias have been detected previously at loci that generate P4siRNAs (Li et al., 2015), we propose that Pol II contributes to the production of sidRNA precursors. Supporting this notion, we found that sidRNAs were produced from a transgenic locus under the control of a 3S promoter that is transcribed by Pol II (Figures S3A-S3C).

A Proposal for sidRNAs as Initial Triggers of DNA Methylation

The canonical RdDM pathway involves 24-nt siRNAs that are produced by the coordinated activities of Pol IV, RDR2, and DCL3. These 24-nt siRNAs are bound by AGO4 and used as a guide through base-pairing with Pol V-generated scaffold transcripts to recruit DRM2 that catalyzes de novo DNA methylation (Law and Jacobsen, 2010; Matzke et al., 2009). Targeting of Pol IV and Pol V is therefore critical for the specificity of RdDM. It has been shown that the recruitment of Pol IV is assisted by two Pol-IV-interacting proteins, Classy 1 (CLSY1) (Law et al., 2011) and Sawadee Homeodomain Homolog 1 (SHH1)/DNA-Binding Transcription Factor 1 (DTF1) (Law et al., 2013; Zhang et al., 2014a).
stranded RNA precursors. Double-stranded precursors are produced by Pol II and RDR2, which produces either double-stranded or single-stranded siRNAs processed by DCL2 and DCL4 from Pol II/RDR6-generated dsRNAs. These siRNAs initiate low levels of de novo DNA methylation in an AGO6-dependent manner and suggest the initial methylation activates the canonical RdDM pathway (Kim and Zilberman, 2014; Matzke and Mosher, 2014; McCue et al., 2015; Panda and Slotkin, 2013). Another study utilizing a unique system for initiating DNA methylation suggests that dsRNAs are produced by Pol II/RDR6, when present at high levels, can be processed by DCL3 into 24-nt siRNAs to induce de novo methylation (Mari-Ordoñez et al., 2013).

Our finding of siRNAs leads us to propose an alternative model for the initiation of RdDM (Figure 7). In this model, Pol II transcripts (possibly amplified by RDR6) associate with AGO4 and undergo 3’-5’ trimming by Atrimmers. The resulting siRNAs then initiate DNA methylation at a locus without any epigenetic modification by recruiting DRM2. Low levels of DNA methylation established by siRNAs may mark the locus for subsequent recruitment of Pol IV and RDR2. The vast majority of the Pol IV/RDR2 transcripts are processed by DCL3 to generate 24-nt siRNAs that are subsequently loaded onto AGO4, whereas single-stranded precursors bind directly to AGO4 and undergo 3’-5’ trimming. Despite being generated through different routes, both 24-nt siRNAs and siRNAs then recruit DRM2 by base-pairing with scaffold transcripts generated by Pol V to reinforce DNA methylation. The role of siRNAs in maintaining DNA methylation is supported by the results showing that siRNAs mediate the enrichment of AGO4 on chromatin and CHG/CHH DNA methylation (Figure 4).

In our model, DCL3-dependent biogenesis of canonical 24-nt siRNAs and Atrimmer-dependent biogenesis of siRNAs are not mutually exclusive. Instead, they act in parallel in wild-type plants to produce siRNAs from TEs at which RdDM has been established. At loci with established RdDM, the vast majority of Pol IV/RDR2 transcripts is transcribed from the over-accumulation of laddered sRNAs in these mutants (Figures 1E–1G). At active loci, Pol IV/RDR2 transcripts are processed by DCL3 into canonical 24-nt siRNAs, and a small fraction of these siRNAs are processed into siRNAs by the trimming mechanism. In the dcl2/3/4 mutant, where DCL3 is absent, a larger fraction of the Pol IV/RDR2 transcripts is processed by the trimming mechanism, as evident from the over-accumulation of laddered sRNAs in these mutants (Figures 1E–1G). At active loci, Pol II/RDR6 transcripts are processed by the trimming mechanism to produce siRNAs, which trigger RdDM. In summary, we propose that siRNAs produced by the trimming mechanism make a minor contribution to the maintenance of DNA methylation at established RdDM loci but play key roles in the initiation of RdDM.

**EXPERIMENTAL PROCEDURES**

**Plant Materials and Oligonucleotides**

All Arabidopsis lines used in this study were in the Columbia (Col-0) background. Detailed information about mutants, transgenic lines, and oligonucleotides used in this study can be found in the Supplemental Experimental Procedures.

**Deep Sequencing and Data Analysis**

siRNAs of 18–30 nt or 30–60 nt were gel-purified and subjected to library construction as described previously (Ai et al., 2008). Bisulfite sequencing libraries
were constructed using the NEXTflex bisulfite sequencing (bisulfite-seq) kit (Bioo Scientific) according to the manufacturer’s instructions. Libraries were sequenced on the Illumina HiSeq 2000 platform by Bionova Biotech. Detailed protocols for library construction and methods for sequencing data analyses can be found in the Supplemental Experimental Procedures.

sRNA Northern Blot
sRNAs were purified from total extracts or AGO4 immunoprecipitates and subjected to northern blot analysis as described previously (Qi et al., 2006).

Immunoprecipitation
Immunoprecipitation of GFP-AGO4 complexes from 3-week-old seedlings was performed as described previously (Ye et al., 2012). The quality of purification was examined by SDS-PAGE followed by silver staining.

AGO4 ChiP
The ChiP assay was performed with anti-AGO4 antibody (xAGO4) as described previously (Rowley et al., 2011).

Bisulfite Sequencing and Chop-PCR
Genomic DNA was extracted from 3-week-old seedlings using the DNeasy plant mini kit (Qiagen). 500 ng of DNA was used for bisulfite sequencing analysis as described previously (Wu et al., 2010). 500 ng of DNA was used for Chop-PCR as described previously (Zhang et al., 2014b) after digestion with HaeIII.

Analyses of Features of sidRNAs
The sensitivities of sidRNAs to RNase T1 and Terminator exonuclease were examined to determine whether sidRNAs are single-stranded and contain 5' monophosphate ends. Detailed protocols can be found in the Supplemental Experimental Procedures.

ACCESSION NUMBERS
The accession number for the sequencing data reported in this paper is GEO: GSE74398.

SUPPLEMENTAL INFORMATION
Supplemental Information includes Supplemental Experimental Procedures, four figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2015.11.015.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS
This work was supported by National Natural Science Foundation of China Grants 31225015, 31330042, and 31421001 (to Y.Q.). We would like to acknowledge the assistance of the Computing Platform of the National Protein Science Facility (Beijing).

Received: September 15, 2015
Revised: October 27, 2015
Accepted: November 4, 2015
Published: December 17, 2015

REFERENCES


Molecular Cell, Volume 61

Supplemental Information

A Dicer-Independent Route for Biogenesis of siRNAs that Direct DNA Methylation in Arabidopsis

Ruiqiang Ye, Zulong Chen, Bi Lian, M. Jordan Rowley, Ning Xia, Jijie Chai, Yan Li, Xin-Jian He, Andrzej T. Wierzbicki, and Yijun Qi
Figure S1, Identification of sidRNAs, Related to Figure 1.
Figure S2, Identification of Endosperm-Specific sidRNAs, Related to Figure 1.
Figure S3, Detection of sidRNAs from Transgenic Loci and Transposons Undergoing de novo Silencing, Related to Figure 1.
Figure S4, Identification of Distributive 3'-5' Exonucleases Required for DNA Methylation, Related to Figure 6.
Table S1, Summary of Small RNA Sequencing Datasets, Related to Figures 1, 2, 3, 5, and 6
Table S2, Identification of sidRNA Loci (sidRLs), Related to Figure 1
Table S3, Effects of Pol IVRDR2 Mutations on sidRNA Production, Related to Figure 2
Table S4, Identification of Endosperm-Specific sidRNA Loci (ES-sidRLs), Related to Figure S2
Table S5, Summary of Whole Genome Bisulfite Sequencing Datasets, Related to Figure 4
Table S6, Profiles of DNA Methylation at sidRNA Loci by Whole-Genome Bisulfite Sequencing, Related to Figure 4
Table S7, Identification of sidRNA Loci that Are Controlled by Atrimmer1 and/or Atrimmer2, Related to Figure 6
Figure S1, Identification of sidRNAs, Related to Figure 1.

(A) Diagrams showing the gene structures of DCLs. The positions of T-DNA insertions are indicated.

(B) Box plots of the levels of miRNAs, ta-siRNAs and P4siRNAs in the indicated plants.

(C) Box plots of the levels of 30-60-nt sRNAs generated from sidRNA loci in the indicated plants. Asterisks indicate a significant difference between Col-0 and the mutants ($P < 10^{-15}$, Mann–Whitney U test).
Figure S2, Identification of Endosperm-Specific sidRNAs, Related to Figure 1.

(A) Box plots of levels of sRNAs of the indicated sizes produced from endosperm-specific sidRNA loci in Col-0, dcl2/3/4 and dcl1/2/3/4. Asterisks indicate a significant difference between Col-0 and the mutants ($P < 10^{-15}$, Mann–Whitney U test).

(B) Density of sidRNAs mapped to the ES-sidRL47 (cluster55) locus in the indicated plants.

(C) A representative group of laddered sRNAs produced from ES-sidRL47 (cluster55) locus. Numbers represent the abundance (RPM) of sRNAs with different lengths ranging from 19 to 28 nt. Values in the parentheses indicate the percentages of sRNAs with different lengths.
Figure S3, Detection of sidRNAs from Transgenic Loci and Transposons Undergoing de novo Silencing, Related to Figure 1.

(A, B) Detection of sidRNAs derived from transgenic loci in T1-T6 generations of the indicated genotypes by deep sequencing analysis.

(C) A representative group of laddered sRNAs produced from transgenic loci in T1 generation in Col-0 background. Numbers represent the abundance (RPM) of sRNAs with different lengths ranging from 18 to 28 nt. Values in the parentheses indicate the percentages of sRNAs with different lengths.

(D, E) Detection of sidRNAs mapped to EVD locus in epi15 F8, F11, F14, and wild-type (WT) plants by deep sequencing analysis.

(F) A representative group of laddered sRNAs produced from EVD locus in epi15 F11 plants. Numbers represent the abundance (RPM) of sRNAs with different lengths ranging from 18 to 30 nt. Values in the parentheses indicate the percentages of sRNAs with different lengths.
Figure S4, Identification of Distributive 3’-5’ Exonucleases Required for DNA Methylation, Related to Figure 6.

DNA methylation levels at the AtSN1 and IGN5 loci in the indicated plants were measured by Chop-PCR. The methylation-sensitive restriction enzyme HaeIII was used to specifically cleave unmethylated DNA in CHH context. A fragment of Actin that lacks HaeIII sites was amplified as a control.

Extended Experimental Procedure

Plant Materials

The dcl2-1 dcl3-1 dcl4-2 (dcl2/3/4), nrpd1-3 (nrpd1), rdr2-1 (rdr2) and ago4-1 (ago4) mutants in the Col-0 ecotype were described previously (Henderson et al., 2006; Onodera et al., 2005; Wierzbicki et al., 2009; Xie et al., 2004). The quadruple mutant dcl1-9 dcl2-1 dcl3-1 dcl4-2 (dcl1/2/3/4) was obtained by crossing dcl2-1 dcl3-1 dcl4-2 with dcl1-9 dcl2-5 dcl3-1 dcl4-2 (Blevins et al., 2011). The quadruple mutants nrpd1 dcl2/3/4, rdr2 dcl2/3/4, and ago4 dcl2/3/4 were obtained by crossing dcl2/3/4 with nrpd1, rdr2, and ago4, respectively. T-DNA insertion lines for genes encoding 3’-5’ exonucleases in the Col-0 ecotype were obtained from the Arabidopsis Biological Resource Center (ABRC). All mutants were genotyped by PCR. The transgenic lines NRPD1-FLAG and NRPD1(ASM)-FLAG were described (Berger et al., 2009) and crossed into nrpd1 dcl2/3/4. The transgenic lines GFP-AGO4 in Col-0 or dcl2/3/4 background were described (Ye et al., 2012).

Constructs pAGO4:GFP-AGO4 and pAGO4:GFP-AGO4YF/AA were described previously (Ye et al., 2012) and transformed into the ago4 dcl2/3/4 mutant using the floral-dip method (Clough and Bent, 1998). Positive transformants were identified through selection for antibiotic resistance and confirmed by Western blot. The GUS gene was PCR-amplified, cloned into pENTR/D-TOPO (Invitrogen) and introduced into pEarley-Gate202 (Earley et al., 2006) through LR recombination reaction. The resulting p35S:GUS construct was transformed into Col-0, nrpd1, dcl2/3/4 and nrpd1 dcl2/3/4. At least 50 lines were obtained for each genotype. Propagation of the transgenic lines was performed by single-seed descent procedures. In every generation, ten flowers from each plant of the same genotype were harvested and pooled for further analysis.

sRNA Sequencing and Analysis
sRNAs of 18-30 nt or 30-60 nt were gel-purified on a 15% denaturing PAGE gel and subjected to library construction as described (Mi et al., 2008). A detailed protocol is available upon request. The libraries were single-end sequenced on an Illumina HiSeq2000 platform.

After removing adapters and low-quality reads, sRNA-seq reads were mapped to the Arabidopsis genome (TAIR10 version) with Bowtie (Langmead et al., 2009) allowing no mismatches, and the mapped reads were retained for further analyses.

sRNA reads from dcl1/2/3/4 were used for identification of sidRNA loci. A sidRNA loci were annotated using the following criteria: 1) length ≥ 100 nt; 2) number of unique sRNAs ≥ 5, each separated from nearest neighbors by a maximum of 50 nt; 3) expression level ≥ 15 RPKM (reads per kilobase per million).

sRNA abundance was calculated as reads per million (RPM) or reads per 10 million (RP10M). To identify sidRNA loci differentially expressed in different tissues or genotypes, edgeR was applied to calculate the fold change and p-value. The resulting p-values were adjusted by Benjamini-Hochberg’s approach to control false discovery rate (FDR). Those showing ≥ 3 fold change at a FDR < 0.01 were considered as differentially expressed loci.

Whole-genome Bisulfite Sequencing and Analysis

Paired-end bisulfite sequencing libraries were constructed using the NEXTflex Bisulfite-Seq kit (BIOO SCIENTIFIC) according to the manufacturer’s instructions. About 2 µg of genomic DNA was sonicated to ~300 bp, end repaired and ligated to methylated adapters. Ligated products were subjected to bisulfite treatment using the EZ DNA Methylation-Gold kit (ZYMO Research) according to the manufacturer’s instructions, and then PCR amplified for 15 cycles using EX Taq DNA polymerase (Takara). The amplified libraries were purified twice using Ampure XP purification beads (Beckman) prior to quantification with the Agilent 2100 Bioanalyzer. The libraries were paired-end sequenced on the Illumina HiSeq 2000 generating 2x100-mer reads.

Bisulfite sequencing reads were mapped to Arabidopsis genome (TAIR10) by BRAT-BW (Harris et al., 2012), allowing up to three mismatches. Only uniquely mapped non-redundant reads were retained for further methylation analysis. Chloroplast bulk fractional methylation was used to monitor bisulfite conversion efficiency. Fractional DNA methylation levels were calculated by \#C/(\#C+\#T). CHH DMRs (differentially methylated regions) were defined as previously described (Stroud et al., 2013; Zhang et al., 2013). DNA methylation levels in every 200 bp sliding window with 50 bp step size were compared in wild-type and mutants. Only windows with more than four informative cytosine (i.e. coverage ≥ 4) were retained. The windows with absolute methylation difference > 0.1 and Benjamini-Hochberg-adjusted FDR < 0.001 (Fisher’s exact test) were considered as differentially methylated bins. Then, those differentially methylated bins separated by gaps within 100 bp were merged together as differentially methylated regions (DMRs). The final coordinates were further adjusted from the first methylated cytosine to the last methylated cytosine.

Analyses of Features of sidRNAs

To analyze whether sidRNAs have 5’ monophosphate ends, sRNAs in total extracts and AGO4 immunoprecipitates prepared from Col-0 and dcl2/3/4 were treated with Terminator exonuclease (Epicentre) following the manufacturer’s standard procedure. A synthetic 21-nt spike-in RNA with 5’ monophosphate was added in the reactions and used as a positive control. To examined whether the
sRNAs are single- or double-stranded, the sRNA samples were incubated with RNase T1 (Thermo SCIENTIFIC) at 37°C for 20 mins and AtREP2 sRNAs were probed by Northern blot. Spike-in RNAs including a 50-bp double-stranded RNA with 2-nucleotide 3’ overhang (ds) and a 50-nucleotide single-stranded (ss) RNA were added in the reactions and used as controls for RNase T1 digestion. Reactions were stopped by phenol/chloroform extraction, and the RNAs were collected by ethanol precipitation. The resultant RNAs were subjected to Northern blot analysis using a probe to detect AtREP2-derived sRNAs.

**Oligonucleotides**

Primers and probes used in this study are listed below.

### Probes for small RNA Northern blot

<table>
<thead>
<tr>
<th>Oligo</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>anti-miR171</td>
<td>GATATTGGCGCGGCTCAATCA</td>
</tr>
<tr>
<td>anti-ta-siR255</td>
<td>TACGCTATGTGAGACTTAA</td>
</tr>
<tr>
<td>anti-AtREP2</td>
<td>GCGGGACGGGTTTGGCAGGACGTTACTTAAT</td>
</tr>
<tr>
<td>anti-SIMPLEHAT2</td>
<td>TGGGTTACCCATTTTGGACACCCCTA</td>
</tr>
<tr>
<td>anti-AtSN1-1</td>
<td>CACCAACCGTGGTGGCCCGCAATGTTAAACTCTGAGATAGAGTGCTGGAGTACGACA</td>
</tr>
<tr>
<td>anti-AtSN1-2</td>
<td>CTCTCAGATAGAGGTGCTGGATTCGAGACACGTTGGGAGATCTCTTTTCCAAGAAGA</td>
</tr>
<tr>
<td>anti-AtSN1-3</td>
<td>CGTGGAGGAAGCTGCTTTCCTCAAGAAGATGAAATTCTGAGTGGTCCCCTGGG</td>
</tr>
<tr>
<td>anti-AtSN1-4</td>
<td>ATTTCTGTATGGGCTCCCGCTCTGGAGATGTAAGGCTTCTTCTGCGTACGAG</td>
</tr>
<tr>
<td>anti-A4TE27090</td>
<td>GTCTGAGAAAGATTTGTTGGCTTCCCGCAGACCTCAGATGATGTT</td>
</tr>
<tr>
<td>anti-IGN5-1</td>
<td>AGTTTTTGTGGGATTTGGGCGCAATAAACGCAAGTAGCTCCTTT</td>
</tr>
<tr>
<td>anti-IGN5-2</td>
<td>TCTGAGATTTGCCATACCCCTGCCATACCTCTTTTACTACATAG</td>
</tr>
<tr>
<td>anti-sidRL2514</td>
<td>GATACGGGCTTTGCCCTCCATGCACACGCGCCCAACAC</td>
</tr>
<tr>
<td>anti-sidRL5049</td>
<td>TGGGAACAGACGACCTCTTCAAGACCGGAGCAACCTGCTGAGGAGATG</td>
</tr>
</tbody>
</table>

### Primers for bisulfite sequencing

<table>
<thead>
<tr>
<th>Oligo</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtREP2-BiF</td>
<td>AYAYTTTGTGGATATTTTGAATAG</td>
</tr>
<tr>
<td>AtREP2-BiR</td>
<td>CCACCACCTACCTGGAATAAACATAT</td>
</tr>
<tr>
<td>SIMPLEHAT2-BiF</td>
<td>TAGGGGTCTAAAAATGGGTTAAAAATT</td>
</tr>
<tr>
<td>SIMPLEHAT2-BiR</td>
<td>ATAAAAAAAAATTTACGAATTTTACCTCTTC</td>
</tr>
<tr>
<td>IGN5-BiF</td>
<td>GTTYYYYGAAGAGATGAAYAAATGTTAAATGATAYATGGYGGT</td>
</tr>
<tr>
<td>IGN5-BiR</td>
<td>RRACAACTTTCCTGACAACCACCAACCTCCATCCCTCTTTA</td>
</tr>
<tr>
<td>AtSN1-BiF</td>
<td>GTGGTATAAGTTTAGTTTTAATTTTAYGGYATAGTTAATT</td>
</tr>
<tr>
<td>AtSN1-BiR</td>
<td>CAATATACRATCCAAAAACCGTATAAAATATATCTTAA</td>
</tr>
<tr>
<td>sidL5049-BiF</td>
<td>AGAGTAAGTTATGGGTTGGATG</td>
</tr>
<tr>
<td>sidL5049-BiR</td>
<td>CAATCTATATATGCTCAACRTCTT</td>
</tr>
<tr>
<td>sidL2514-BiF</td>
<td>TGAGTGAGTATGGTGGTGGT</td>
</tr>
<tr>
<td>sidL2514-BiR</td>
<td>CCAATTCCTAATCATRAAAAAATTTAAA</td>
</tr>
<tr>
<td>sidL14057-BiF</td>
<td>TTAATTTGTYAYATTTAGYGGT</td>
</tr>
<tr>
<td>sidL14057-BiR</td>
<td>AAATTACTCCRCACTTACTCCTAT</td>
</tr>
</tbody>
</table>

### Primers for Chop-PCR
IGN5-HaeIII-F TCCCGAGAAGAGTAGAAACAATGTACTAAAA
IGN5-HaeIII-R CTGAGGTATCCATAGCCCCCTGATCC
AtSN1-HaeIII-F ACCAACGTCTGTTG6CAGTGGTAAATC
AtSN1-HaeIII-R AAAATAAGTGGTGGTGTACAAAGC
ACT2-HaeIII-F CGAGCAAGATGGGAAACCTCAA
ACT2-HaeIII-R AAGAATGGAACCACCGATCCAGACA

**Primers for AGO4-ChIP**

ACTIN-ChIP-F GAGAGATTCAGATGCCCAGAAGTC
ACTIN-ChIP-R TGGATTCCAGCAGCTTCCA
ArREP2-ChIP-F AATGGCAGCATCAGGATTTT
ArREP2-ChIP-R TGGAGAGATTTTGGGAAGATTG
solo LTR-ChIP-F GTATAGAGATGAAATGATGGATAATGACA
solo LTR-ChIP-R TTATTTTGATCAGTGTTATAACCGGATA
IGN5-ChIP-F AAGCCCCACCACTACACTATAAACCTTAAT
IGN5-ChIP-R CCGAATAACAGCAAGTCTCCTTTAATA
SIMPLEHAT2-ChIP-F TGATGGGTCAAATGAGTTGATGA
SIMPLEHAT2-ChIP-R TTGGGTTAATTGGGTAAACCAT

**Primers for genotyping**

LBb1.3 ATTTTGGCGATTTTCGGAAC
LB3 TAGCATCTGAAATTTCTAATAACCATCTGACAC
GABI-T8474 ATAAATAACGTCCGGACATCTACATTT
dc11(del1-9)_geno_R1 GTATGAGATGGAAGATGATTG
LB-pGV_CAF1_2(del1-9) GCCACACACATCTCATTGATGCTTGG
dc12-1-genotyping-LP GACACACACATCTCATTGATGCTTGG
dc12-1-genotyping-RP TGGAGAGATTTTGGGAAGATTG
dc13-1-genotyping-LP AGATTAACTCTGCACTATGTTG
dc13-1-genotyping-RP TGGAAAATTTTGGCTACAACGG
dc14-2-genotyping-P1 TCTCCATATTGACCATCATACTCATT
dc14-2-genotyping-P2 GCTGCACAGCTGATGATTACAA
dc14-2-genotyping-P3 GCCGCTCGAGATCATCAGCAAAGGAAT
nrpd1a-3-genotyping-LP TTTTGATCCCTTGTGATCACCTG
nrpd1a-3-genotyping-RP CTGATGGTCCCTGTTGAAGATA
rdr2-1-genotyping-P1 ACACATATTGACAGATACG
rdr2-1-genotyping-P2 ATGGTCTAGACAGACGACGACGCCAGATCAAC
rdr2-1-genotyping-P3 TAGCATCTGAAATTTCTAATAACCATCTGACAC
ago4-1-genotyping-F TGACTGACAGCTGAAAATGGGATGTGGAT
ago4-1-genotyping-R GCCACTCTGACAGCTACCCACACTAAGTT
SALK_012547 LP CTCTGCCCAGAGGGTATATG
| SALK_012547 RP | TGATGCTGATTTTGCTGATTG |
| SALK_066543 LP | ATAAAGAGCCGCTGGAGAAG |
| SALK_066543 RP | CGTCATCGTCAATGCTTAAGC |
| CS381695 LP | TGTTGAAAATGATATGGGG |
| CS381695 RP | GAAAGGTAGTGAAGGCTT |
| SALK_144940 LP | GTTCAGATACAAAAAGCAG |
| SALK_144940 RP | CAGTCATCTTGAAAGCTAAT |
| SALK_072496 LP | GAAAGGTAGATGAAGGGTCC |
| SALK_072496 RP | TGTTTGGAAAATGATATGGGG |
| SALK_078063 LP | CGCAAAATCAATCAACAG |
| SALK_078063 RP | CACACGTGATCCAAACACAAG |
| SALK_113786 LP | ACTCACCACTTAATTGCACG |
| SALK_113786 RP | ACCATTCGAAGGTTCCTC |
| SALK_145649 LP | ACTTTCTGCTCTCTTTGCCC |
| SALK_145649 RP | CCAATGGCTTCTTCTTCTT |
| SALK_024707 LP | AACAATAGATCATTGCGACG |
| SALK_024707 RP | GATTTCCAGACGAGGAGAAG |
| SALK_068417 LP | GATTATGATTCTGATAGGCG |
| SALK_068417 RP | GATTTGCTGACGGAATATCAC |
| SALK_050279 LP | CGCAAAATCAATCAACAG |
| SALK_050279 RP | TCACAGTCGTGACTCGTGAAG |

**Supplemental Reference**


