Research Article

Running Head: RDR6-RdDM Functions to Methylate Active TEs

Corresponding Author:
R. Keith Slotkin
500 Aronoff Laboratory
318 West 12th Avenue
Columbus, Ohio, 43210
614-292-1087
Slotkin.2@OSU.edu

Research Area: Genes, Development and Evolution
The Initiation of Epigenetic Silencing of Active Transposable Elements is Triggered by RDR6 and 21-22 Nucleotide Small Interfering RNAs

Saivageethi Nuthikattu¹*, Andrea D. McCue¹*, Kaushik Panda¹, Dalen Fultz¹, Christopher DeFraia¹, Erica N. Thomas¹, and R. Keith Slotkin¹,²

¹ Department of Molecular Genetics, The Ohio State University

² Corresponding Author:
500 Aronoff Laboratory
318 West 12th Avenue
Columbus, Ohio, 43210
614-292-1087
Slotkin.2@OSU.edu

*Authors contributed equally

Summary:
This work identifies a missing link between how an active transposable element is originally targeted for DNA methylation and epigenetic silencing.
Financial Sources

A.D.M. is a fellow of The Ohio State Center for RNA Biology.

D.F. is supported by NIH Training Grant T32 GM086252.

C.D. is supported by The Ohio State University’s Pelotonia Post-Doctoral Fellowship.

This work was supported by grant MCB-1020499 to R.K.S from the U.S. National Science Foundation.
Abstract

Transposable elements (TEs) are mobile fragments of DNA that are repressed in both plant and animal genomes through the epigenetic inheritance of repressed chromatin and expression states. The epigenetic silencing of TEs in plants is mediated by a process of RNA-directed DNA methylation (RdDM). Two pathways of RdDM have been identified: Pol IV-RdDM, which has been shown to be responsible for the de novo initiation, corrective reestablishment, and epigenetic maintenance of TE and/or transgene silencing, and RDR6-RdDM, which was recently identified as necessary for maintaining the repression for a few TEs. We have further characterized RDR6-RdDM using a genome-wide search to identify TEs that generate RDR6-dependent small interfering RNAs (siRNAs). We have determined that TEs only produce RDR6-dependent siRNAs when transcriptionally active, and we have experimentally identified two TE sub-families as direct targets of RDR6-RdDM. We used these TEs to test the function of RDR6-RdDM in assays for the de novo initiation, corrective reestablishment and maintenance of TE silencing. We found that RDR6-RdDM plays no role in maintaining TE silencing. Rather, we find that RDR6 and Pol IV are two independent entry points into RdDM and epigenetic silencing which perform distinct functions in the silencing of TEs: Pol IV-RdDM functions to maintain TE silencing and to initiate silencing in a Pol II expression-independent manner, while RDR6-RdDM functions to recognize active TE Pol II-derived transcripts to both trigger and correctively reestablish TE methylation and epigenetic silencing.
Introduction

Transposable elements (TEs) comprise large percentages of both animal and plant genomes. TEs are major targets of multiple endogenous gene silencing pathways that act to limit their expression and ability to generate new insertions and mutations (reviewed in Girard and Hannon, 2008). To study TE silencing, the process has been divided into three distinct mechanisms: the de novo initiation / triggering of silencing, the corrective reestablishment of silencing of TEs that were recently transcriptionally reactivated, and the epigenetic maintenance of TE silencing.

In the Arabidopsis genome, nearly all TEs are found in a transcriptionally silenced state (Lippman et al., 2004). This transcriptional gene silencing (TGS) is maintained by symmetrical DNA methylation, which is propagated through mitotic cell divisions (reviewed in Law and Jacobsen, 2010). In contrast to animals, plants do not erase the DNA methylation patterns of their gametes, and therefore CG and CHG (H=A,T or C) symmetrical DNA methylation patterns established in one generation are inherited and maintain TE silencing in the next generation through a process termed trans-generational epigenetic inheritance (Mathieu et al., 2007; Becker et al., 2011). In addition to the maintenance of symmetrical methylation, methylation of TEs is continually reinforced through a process of RNA-directed DNA methylation (RdDM)(reviewed in Law and Jacobsen, 2010; Haag and Pikaard, 2011). This pathway involves RNA Polymerase IV (Pol IV), a plant-specific DNA-dependent RNA polymerase that transcribes heterochromatic regions such as TEs into non-protein coding transcripts. These transcripts are converted into double-stranded RNA by the RNA-dependent RNA Polymerase RDR2 and cleaved into 24 nucleotide (nt) small interfering RNAs (siRNAs) by Dicer-like 3 (DCL3). The resulting 24 nt siRNAs associate with Argonaute 4 (AGO4) and AGO6 to target nascent TE transcripts in the nucleus produced by another plant-specific DNA-dependent RNA polymerase, RNA Polymerase V (Pol V). Targeting of Pol V transcripts by AGO protein complexes mediated through siRNA complementarity results in the recruitment of the DNA methyltransferase DRM2 to methylate cytosines at the Pol V-transcribed locus (reviewed in Law and Jacobsen, 2010). DNA targets of RdDM can be identified through CHH context DNA methylation, as this asymmetrical methylation pattern is not replicated. The Pol IV-RDR2-DCL3-AGO4/6-POL V-DRM2 pathway (from this point forward referred to as the Pol IV-RdDM pathway) acts as a loop reinforcing methylation states at regions of heterochromatin and silenced TEs.
Pol IV is thought to transcribe regions of the genome that are already DNA methylated (Zheng et al., 2009; Wierzbicki et al., 2012). Therefore, the initiation of TE silencing solely by the Pol IV-RdDM pathway presents a chicken-and-egg dilemma: if Pol IV transcription is guided to previously heterochromatic and methylated TEs, how can TE silencing be initiated by Pol IV transcription? A possible resolution to this conundrum has been potentially illuminated by several recent studies that have begun to uncover factors involved in a distinct small RNA-directed DNA methylation pathway that acts independently of the Pol IV-RdDM machinery (Eamens et al., 2008; Garcia et al., 2012; Pontier et al., 2012; Stroud et al., 2013). This pathway utilizes 21-22 nt siRNAs dependent on the RNA-dependent RNA Polymerase RDR6 and DCL2 to target methylation and maintain the silencing of a single AtCopia18A LTR retrotransposon fragment on chromosome 5, an AtREP1 Helitron family TE on chromosome 1, as well as several intergenic regions (Pontier et al., 2012). Importantly, this RDR6-dependent DNA methylation pathway (from this point forward referred to as the RDR6-RdDM pathway) was able to function in the absence of Pol IV, suggesting that it operates on siRNAs derived from RNA Polymerase II (Pol II) transcripts. When transcriptionally active, Pol II-derived TE mRNAs can be post-transcriptionally degraded into siRNAs that retarget complementary transcripts for further degradation in the cyclic RNA interference (RNAi) pathway (Sijen and Plasterk, 2003; Chung et al., 2008). In Arabidopsis, RDR6, DCL2, DCL4, and AGO1 degrade some TE mRNAs to produce siRNAs of 21-22 nt (McCue et al., 2012). Therefore, the RDR6-RdDM pathway could provide a link between post-transcriptional gene silencing (PTGS) / RNAi mediated by RDR6-dependent 21-22 nt siRNAs, and the DNA methylation responsible for the initiation of Pol IV-RdDM and TE trans-generational silencing.

The Pol IV-RdDM pathway has been previously shown to be necessary for the initiation of transgene silencing (Aufsatz et al., 2002; Chan et al., 2004; Greenberg et al., 2011), the corrective reestablishment of TE silencing (Teixeira et al., 2009; Ito et al., 2011), and the maintenance of some TE silencing (Herr et al., 2005; Huettel et al., 2006). In this report we did not focus on deciphering the precise RDR6-RdDM molecular mechanism, but rather we aimed to identify the major genome-wide targets of RDR6-RdDM, while also determining if RDR6-RdDM is involved in the initiation, corrective reestablishment and/or maintenance of TE silencing. If involved in these processes, we aimed to determine the relative contribution of the RDR6-RdDM and Pol IV-RdDM pathways. We have particularly focused on RDR6, a protein identified as a necessary component of the Pol IV-RdDM-independent DNA methylation of trans-acting
siRNA (TAS) loci as well as the single *Copia* and *Helitron* elements (Pontier et al., 2012; Wu et al., 2012). We have discovered that Pol IV-RdDM and RDR6-RdDM function differently, with RDR6-RdDM functioning only in the Pol II expression-dependent initiation and corrective reestablishment of TE silencing.

Results

**Genome-wide identification of RDR6-dependent TE siRNAs**

A recent report suggested that 21 nt siRNAs and RDR6-RdDM function to maintain the transcriptional silencing of two individual TEs (Pontier et al., 2012). To identify which TEs are targets of RDR6-RdDM on a genome-wide scale, we performed deep sequencing of small RNAs from both wt Col and *rdr6* mutant inflorescences. We obtained 3,129,843 genome-matched 18-28 nt reads from wt Col and 1,479,287 from *rdr6*. In addition, we sequenced small RNAs from genomes with active TEs. Mutation in the *swi/snf* family chromatin remodeling protein DDM1 results in global loss of heterochromatin and genome-wide transcriptional reactivation of TEs (Gendrel et al., 2002; Lippman et al., 2004). We sequenced small RNAs from a homozygous *ddm1* single mutant (3,528,426 reads) and a *ddm1 rdr6* double mutant (2,677,800). We could therefore determine which TEs have RDR6-dependent siRNAs in both a TE-silenced (wt Col) and TE-activated (*ddm1*) background.

To characterize the TE content of our sRNA libraries, we began by separated the small RNA libraries into sizes of 21, 22 and 24 nts because these are the sizes with known biogenesis and function (reviewed in Chen, 2010). The overall distribution of perfectly genome-matching sRNAs of these small RNA size classes is similar in wt Col and *rdr6* (Figure 1A). As a control we quantified the number and size of tasiRNAs in each library (Figure 1B), and as predicted found that libraries without a functional RDR6 protein do not produce tasiRNAs (Allen et al., 2005), validating the library quality. We next classified the TE-derived siRNAs and again found very similar size distributions between wt Col and *rdr6* (Figure 1C), in agreement with previous studies which found that RDR6 is generally not involved in TE regulation in a TE-silenced background (Kasschau et al., 2007). Due to the short length of TE siRNAs and their repetitive nature, assigning the single TE locus that produced each TE siRNA is not possible. Instead, we categorized the TE class, family, and sub-family from which each TE siRNA was derived (see Materials and Methods for details). We compared wt Col and *rdr6* on the TE sub-family level to identify particular TE sub-families that have RDR6-dependent 21 or 22 nt siRNAs. Of the 318 TE
sub-families analyzed, we detected only 2 (0.63%) that display a 3-fold or greater difference between wt Col and rdr6 (red dots, Figure 1D). These two TE sub-families are RomaniAT5, an LTR retrotransposon, and AtRep10C, a short non-autonomous derivative of the Helitron1 family. We analyzed the siRNA size distribution of the RomaniAT5 and AtRep10C sub-families and found that RomaniAT5 has RDR6-dependent 22 nt siRNAs, while AtRep10C has RDR6-dependent 21 nt siRNAs (Figure 1E). Using genomic polymorphisms we determined that one individual TE copy (At5TE55255) from the 123 copy AtRep10C sub-family generates 95% of all AtRep10C sub-family RDR6-dependent 21 nt siRNAs. In our comparison between wt Col and rdr6 TE siRNAs we did not detect a change in siRNA levels of the TEs AtRep1 or AtCopia18A (green dots, Figure 1D), which represent the sub-families of the two individual elements identified by Pontier et al as regulated by RDR6-RdDM (Pontier et al., 2012). AtRep10C is a non-autonomous member of the larger autonomous AtRep1 element family, and together this data demonstrates that RDR6-RdDM regulates only a possible few individual elements of this larger Helitron family, perhaps due to position in the genome rather than TE identity. We conclude that, in a TE-silenced epigenome such as wt Col, very few TEs produce RDR6-dependent siRNAs.

Although RDR6-dependent TE siRNA production is very low when TEs are epigenetically silenced, we find that RDR6 plays a much larger role in TE siRNA production when TEs are transcriptionally active in the ddm1 mutant background. In the total small RNA pool, ddm1 mutants display increased amounts of 21 nt small RNAs compared to wt Col (Figure 1A). These 21 nt small RNAs are RDR6-dependent, as their levels revert back to wt Col levels in ddm1 rdr6 double mutants (Figure 1A). The increase in ddm1 21 nt small RNAs detected in the total small RNA sample is produced from TE siRNAs (Figure 1C). We found that in ddm1 mutants 90% of the abundant TE 21-22 nt siRNAs are dependent on RDR6, while in wt Col only 3% of the low level TE 21-22 nt siRNAs are RDR6 dependent (Figure 1C). The dependence of these TE 21-22 nt siRNAs on an RNA dependent RNA polymerase defines them as secondary siRNAs, amplified from primary TE transcripts. Upon characterization of the TE families and sub-families, we identified 15 of 318 TE sub-families (4.7%) that are responsible for all of the RDR6-dependent TE siRNAs when transposable elements are transcriptionally active in the ddm1 mutant epigenome. If only these 15 sub-families are analyzed, a 77-fold increase in 21 nt siRNAs and a 10-fold increase in 22 nt siRNAs compared to wt Col are observed, both of which are RDR6-dependent (Figure 1F). When these 15 sub-families are subtracted from the analysis,
no increase in 21 and/or 22 nt siRNAs in ddm1 mutants is observed (Figure 1G). These 15 TE sub-families are listed in Figure 1H with their relative contribution to the pool of 21-22 nt siRNAs in ddm1. The majority (96%) of RDR6-dependent TE siRNAs are from Athila family retrotransposons (Athila0, 1, 2, 3, 4, 5, 6A and 6B)(Figure 1H). Several AtENSPM family CACTA-like DNA transposons (AtENSPM2, 5 and 6), two AtCOPIA LTR retrotransposon sub-families (AtCOPIA93 and 52), and the individual sub-families of the AtGP1 LTR retrotransposon and Vandal3 Mutator super-family DNA transposon also produce RDR6-dependent siRNAs in ddm1 mutants (Figure 1H). The AtCOPIA93 sub-family is also called Évadé, and has been shown to produce 21-22 nt siRNAs only when transcriptionally active (Mirouze et al., 2009). These identified TE sub-families represent high copy (Athila2 has 413 copies/fragments present in the reference Arabidopsis genome), medium copy (AtENSPM2 has 114 copies/fragments), and low copy (AtCopia52 has 7 copies/fragments) TE sub-families. Although only 15 sub-family elements display RDR6-dependent siRNA accumulation, these 15 TE sub-families occupy 24.8% of the total TE space and 4.85% of the entire Arabidopsis genome. Therefore, we conclude that RDR6 plays a larger role in the production of TE siRNAs in ddm1 mutants, presumably when the TEs are transcriptionally active, and we have used this genome-wide small RNA sequencing data to generate candidate TEs regulated by RDR6-RdDM activity.

RDR6-RdDM and Pol IV-RdDM can target the same transcriptionally active TE region for CHH hypermethylation

We next aimed to determine if TE transcriptional activation leads to expression-dependent DNA methylation specifically through the production of 21-22 nt siRNAs and RDR6-RdDM. We specifically concentrated on whether TEs with RDR6-dependent siRNAs are targets of RDR6-RdDM by comparing TE methylation in ddm1 single mutants (with 21-22 nt siRNAs) to ddm1 rdr6 double mutants, which lose 21-22 nt siRNAs (Figure 1F). A problem that potentially complicates our analysis stems from the observation that 24 nt siRNAs from the Pol IV-RdDM pathway also decrease in ddm1 rdr6 double mutants (Figure 1F). Therefore, we searched our small RNA sequencing datasets for a TE sub-family that has RDR6-dependent 21-22 nt siRNAs in ddm1 in single mutants, and 24 nt siRNAs that are not RDR6-dependent. By focusing on this TE sub-family, we are able to separate methylation induced by 21-22 nt siRNAs and RDR6-RdDM from methylation induced by 24 nt siRNAs and Pol IV-RdDM. We identified the TE sub-family AtENSPM6, which has increased RDR6-dependent 21 and 22 nt siRNAs in ddm1
compared to wt Col (Figure 2A), and is one of the most minor contributors to the production of RDR6-dependent siRNAs in ddm1, accounting for only 0.19% of the total increase (Figure 1H). We further analyzed AtENSPM6 to identify a specific region of the element that displays increased levels of RDR6-dependent 21-22 nt siRNAs in ddm1 compared to wt Col, and consistent levels of 24 nt siRNAs in ddm1 and ddm1 rdr6. We identified the first and second exon of AtENSPM6 for further analysis (red box, Figure 2B). Examination of this region shows equal amounts of 24 nt siRNAs in ddm1 and ddm1 rdr6 (Figure 2C). Bisulfite sequencing of DNA methylation patterns for this region identified strong CHH hypermethylation in ddm1 single mutants: from 9% CHH methylation in wt Col to 63% in ddm1 (Figure 2D). CHH hypermethylation of TEs is known to occur when TEs are transcriptionally reactivated, particularly at Gypsy family LTR retrotransposons such as Athila (Stroud et al., 2013), either in TE-reactivated mutants (Saze and Kakutani, 2007) or when they are transcriptionally activated in the pollen vegetative nucleus of wt plants (Schoft et al., 2009; Slotkin et al., 2009; Calarco et al., 2012; Ibarra et al., 2012). In ddm1 rdr6 double mutants, CHH hypermethylation of AtENSPM6 is significantly reduced to 32% (p<0.01)(Figure 2D), demonstrating that the specific loss of the 21 and/or 22 nt siRNAs are responsible for a portion of the CHH hypermethylation. This analysis of AtENSPM6 demonstrates that activation of TEs can lead to their methylation through the RDR6-RdDM pathway, and that locating sites of RDR6-dependent 21-22 nt siRNA production can be used to successfully identify TE or other candidate genomic locations of RDR6-RdDM activity. In addition to the change in CHH context methylation levels, the CG and CHG methylation levels of AtENSPM6 in ddm1 single mutants (79% CG and 73% CHG) also decrease in ddm1 rdr6 (52.5% CG and 47.4% CHG)(Figure 2D). CG and CHG methylation are likely higher than CHH methylation in each genotype tested due to siRNA-directed targeting of new methylation to cytosines in any sequence context, followed by S-phase replication of DNA methylation patterns only in the symmetrical CG and CHG contexts.

In addition to being targeted for RDR6-RdDM methylation, this same region of AtENSPM6 is also targeted by 24 nt siRNAs and CHH methylation via Pol IV-RdDM. Mutations in the largest subunit of RNA Pol IV (nrpD1) in double mutant combination with ddm1 also show reduced methylation compared to ddm1 single mutants (Figure 2D). Mutations in either the Pol IV-RdDM or RDR6-RdDM pathways each produce roughly half of the level of AtENSPM6 CHH methylation (both 32%) compared to when the TE is active in ddm1 mutant plants (63%)(Figure 2D). This demonstrates that RDR6-RdDM and Pol IV-RdDM function on the same TE to
establish CHH hypermethylation upon TE transcriptional activation. Mutations in the largest subunit of RNA Pol V (nrpE1) in combination with ddm1 show a more severe reduction compared to ddm1 rdr6 and ddm1 nrpD1 (Figure 2D), corroborating previous findings that suggest that Pol V is a shared component of both the RDR6-RdDM and Pol IV-RdDM pathways (Wu et al., 2012).

Examination of the AtENSPM6 steady-state polyadenylated expression levels demonstrates that the methylation of this coding region of the AtENSPM6 element examined in Figure 2B-D does not correlate with or control the element’s transcriptional activity. AtENSPM6 activation in a ddm1 single mutant displays the same expression levels as ddm1nrpD1 and ddm1nrpE1 double mutants (Figure 2E), which have significantly less AtENSPM6 methylation (Figure 2D). ddm1 rdr6 double mutants have increased expression levels compared to the ddm1 single or other ddm1 double mutants (Figure 2E), however this change is not statistically significant. Rather than acting through methylation and on the transcriptional level, in this case RDR6 may be acting post-transcriptionally to degrade the activated AtENSPM6 mRNA into 21-22 nt siRNAs. Therefore, although targeted for methylation by both Pol IV-RdDM and RDR6-RdDM, once activated in ddm1, RDR6-RdDM and Pol IV-RdDM play no role in the repression of AtENSPM6 expression. This is similar to the TAS loci, where RDR6-RdDM-directed methylation plays no role in altering gene expression levels (Wu et al., 2012), as this methylation is likely too distant from the TAS or AtENSPM6 promoter to exert an influence on transcriptional rates.

Methylation induced by RDR6-RdDM and Pol IV-RdDM control the expression of the Athila6 TE

To determine if RDR6-RdDM has any role in regulating the expression of the TEs that produce RDR6-dependent 21-22 nt siRNAs, we next searched to identify a TE sub-family with RDR6-dependent 21-22 nt siRNAs near its TE 5' terminus, the site of TE promoter elements and transcript initiation. Steady-state expression levels of AtENSPM6 are not influenced by the methylation status of its coding region (Figure 2D-E), and the 5’ 500 bp that includes the 5’ terminal inverted repeat and transcriptional regulatory elements (Banks et al., 1988) produces very few siRNAs in wt Col, and none in ddm1 mutants (Figure 2B). Therefore, we focused on the TE family that is the major producer of RDR6-dependent siRNAs in ddm1 mutants: the Athila LTR retrotransposon family. We focused on the Athila6 sub-family, as it is the top contributor to the 21-22 nt siRNA pool in ddm1 mutants (Figure 1H). When transcriptionally activated, Athila6
21-22 nt siRNAs increase 72.5-fold compared to wt Col and comprise 15% of the total small RNA library in ddm1 mutants (Figure 3A).

To characterize the methylation status of the Athila6 transcriptional regulatory region, we first mapped the transcriptional start site (TSS) of the gag/pol protein-coding transcript of the Athila6 sub-family (Supplemental Figure 1) and subsequently used the surrounding 350 base pairs (bp) of this 5’ LTR region for analysis of DNA methylation. When transcriptionally active in ddm1 mutants, this TSS region undergoes a 24-fold increase in 21 nt siRNA, a 4.9-fold increase in 22 nt siRNAs, as well as a 6.2-fold increase in 24 nt siRNAs, all of which are dependent on RDR6 (Figure 3B). We performed bisulfite sequencing of the Athila6 LTR TSS and found in wt Col high levels of CG methylation (93.9%), intermediate levels of CHG methylation (54.7%) and low levels of CHH methylation (15.9%) (Figure 3C). When transcriptionally activated in ddm1 mutants, this methylation pattern is drastically altered: the CG methylation is reduced and the CHH methylation increases 4.3-fold compared to wt Col. Like AtENSPM6, we found that this CHH hypermethylation is dependent on both the RDR6-RdDM pathway and the Pol IV-RdDM pathway (Figure 3C). In ddm1 rdr6 double mutants the CHH methylation is significantly reduced to wt Col levels. As a control, we assayed sibling plants from a segregating family, both of which were ddm1 homozygous mutants but differed in the presence or absence of a functional RDR6 gene, and found a statistically significant (p<0.001) dependence on RDR6 for establishing CHH hypermethylation (Figure 3C). ddm1 double mutants with Pol IV-RdDM components, such as rdr2 and pol IV/nrpD1, also displayed significant reduction in CHH methylation (13.8% and 21.0%, respectively). Importantly, ddm1 nrpD1 double mutants lose all Athila6 TSS 24 nt siRNAs (Figure 3D), confirming that there is an alternate pathway besides production of 24 nt siRNAs for the establishment of the 21% CHH methylation in this double mutant. The least amount of methylation in any sequence context was detected in the ddm1 ago6 (6.2%) and ddm1 nrpE1 (2.7%) double mutants (Figure 3C). The CHH methylation detected in the ddm1 nrpE1 double mutant is the minimum baseline that can be detected in these samples, as this is very close to our experimentally determined conversion efficiency of 98.3% (Figure 3C). In addition to Pol V, AGO6 has been identified as downstream and shared component of both the RDR6-RdDM and Pol IV-RdDM pathways (Wu et al., 2012). CHH hypermethylation of the Athila LTR also occurs when Athila is transcriptionally activated in the wt pollen vegetative nucleus, and in these experiments CHH methylation was dependent only on Pol V and not Pol IV (Schoft et al., 2009). Due to the high levels of Athila 21-22 nt siRNAs detected in pollen (Slotkin et al.,
2009), we suggest that Pol IV-RdDM plays no role in methylating *Athila* and possibly other TEs in the pollen grain, but rather these TEs are methylated by RDR6-RdDM.

Our results from both *Athila6* and *AtENSPM6* suggested that RDR6-RdDM and Pol IV-RdDM function independently to establish CHH hypermethylation on transcriptionally active TEs. To determine if these are the only two pathways responsible for establishing TE methylation, we constructed a *ddm1 nrpD1 rdr6* triple mutant, inactivating critical components of both the RDR6-RdDM and Pol IV-RdDM pathway at the same time. The *ddm1 nrpD1 rdr6* triple mutant has very low levels of *Athila6* LTR TSS methylation in all sequence contexts (10.6% CG, 2.2% CHG, 1.6% CHH)(Figure 3C), lower than both the *ddm1 rdr6* and *ddm1 nrpD1* double mutants that inactivate RDR6-RdDM and Pol IV-RdDM individually. Therefore, we conclude that RDR6 and Pol IV are the only two entry points into RdDM responsible for the expression-dependent CHH hypermethylation of *Athila6*, and these pathways work independently and additively on their target(s) to establish TE methylation levels.

In contrast to *AtENSPM6*, the methylation status of the *Athila6* TSS has a direct effect on the steady-state transcript accumulation of the *gag/pol* protein-coding transcript. Activation of expression is detected in *ddm1* single mutants, and increases (p<0.001) when hypermethylation is reduced in *ddm1 rdr2, ddm1 nrpD1* and *ddm1 nrpE1* double mutants (Figure 3E). This suggests that the transcript increase (p<0.001) in *ddm1 rdr6* compared to *ddm1* single mutants (Figure 3C-D) is due to the loss of hypermethylation in the double mutant, rather than or in addition to RDR6 acting post-transcriptionally. *Athila6* expression was particularly high in *ddm1 nrpE1* double mutants, 10-fold higher compared to *ddm1* single mutants. This expression level likely represents close to the maximum transcript accumulation potential of *Athila6* when virtually uninhibited by repressive DNA methylation.

**RDR6-RdDM functions to initiate expression-dependent TE methylation**

The process of TE silencing has been divided into three distinct mechanisms: the *de novo* initiation of silencing, the corrective reestablishment of silencing of recently reactivated TEs, and the epigenetic maintenance TE of silencing. After identifying RDR6-RdDM target TEs, we set out to determine if the RDR6-RdDM pathway functions to initiate, reestablish, or maintain TE silencing. If it does play a role in these separate processes, we aimed to identify the relative contributions of Pol IV-RdDM and RDR6-RdDM to these functions. We have focused on the *Athila6* retrotransposon, as hypermethylation of the LTR TSS functions to regulate TE
expression (Figure 3C-E). Even though other studies have suggested that additional 21-22 nt siRNA generating proteins, such as RDR1, play a role in DNA methylation (Pontier et al., 2012; Stroud et al., 2013), we find no alteration of Athila6 siRNA levels between ddm1 single mutants and ddm1 rdr1 double mutants (Figure 3D). Therefore, we have focused our studies on RDR6, which is responsible for production of all Athila6 LTR TSS 21-22 nt siRNAs, as well as POL IV, which is responsible for the production of 24 nt siRNAs (Figure 3D).

We wanted to assess if RDR6-RdDM plays a role in the initiation of TE silencing, and if so, its relative contribution compared to Pol IV-RdDM. To test the initiation of TE silencing, we transformed plants with a transgene containing a 946 bp fragment of Athila6. We placed an Athila6 fragment under the control of an estradiol-inducible promoter to test if Pol II expression is necessary for the function of either RdDM pathway. We transformed wt Col, rdr6 and pol IV/nrpD1 mutants, which do not have detectable levels of Athila6 expression (Figure 4A), with our inducible-Athila6 (ind-Athila) transgene. After a mock-induction of T1 generation plants, very low levels of ind-Athila expression were detected, which were just above the limit of detection (Figure 4A). With estradiol-induction of expression, wt Col plants with ind-Athila undergo a 13.7-fold increase in expression, rdr6 a 18.8-fold increase, and pol IV/nrpD1 a 129-fold increase compared to the mock-induced plants (Figure 4A). We compared the expression of the induced genotypes and found that wt Col has the lowest expression level, rdr6 is intermediate (although not statistically significantly higher than wt Col), and pol IV/nrpD1 has the highest expression of the ind-Athila transgene (p<0.05)(Figure 4A). As a control we estradiol-treated non-transgenic wt Col plants and determined that estradiol treatment does not activate expression of endogenous Athila6 elements (Figure 4A).

We next determined that similar to Pol II-derived reactivated endogenous Athila6 transcripts (in ddm1 mutants for example), estradiol-induced transcripts of ind-Athila6 are processed into 21-22 nt siRNAs (Figure 4B). This processing is dependent on RDR6, as rdr6 mutant plants with estradiol-induced Athila6 transgene expression fail to produce 21-22 nt siRNAs (Figure 4B). We are unable to determine if our ind-Athila transgene also produces Athila6 24 nt siRNAs, as the endogenous 24 nt siRNAs present in wt Col plants mask this production (Figure 3A&D, Figure 4B)(McCue et al., 2012). There are nine endogenous Athila6 TEs that match the ind-Athila transgene with ≥95% nucleotide identity over the length of the TE portion of the transgene. The 24 nt siRNAs these endogenous Athila6 elements produce are dependent on Pol IV and RDR2, as pol IV/nrpD1 and rdr2 mutants fail to accumulate Athila6 24
nt siRNAs (McCue et al., 2012). Using our ind-\textit{Athila} transgene system, we are able to control TE expression and 21-22 nt siRNA production to determine the role of RDR6-RdDM on the initiation of TE methylation.

To determine if RDR6-RdDM plays a role in the initiation of TE silencing, we performed bisulfite sequencing of the promoter and \textit{Athila6} portion of the ind-\textit{Athila} transgene in T1 generation plants using estradiol-induced inflorescence tissue as well as mock-induced sibling plants. To differentiate between TE and transgene silencing, we bisulfite sequenced a wt Col line with the same transgene backbone expressing the \textit{RAN1} gene (At5g20010) and no repetitive DNA. We found that even without induction of expression this transgene does accumulate some methylation (7.5% CG, 13.0% CHG, 6.3% CHH). However, the symmetrical CG and CHG methylation level was 1/3 or less of the methylation of mock-induced ind-\textit{Athila} in wt Col, demonstrating that ind-\textit{Athila} is specifically recognized as a repetitive TE fragment and methylated beyond the levels of a standard genic transgene even without high levels of expression (Figure 4C). Without induction of expression, ind-\textit{Athila} methylation levels in wt Col are higher in the symmetrical CG and CHG contexts (22.3% CG, 33.8% CHG) compared to the asymmetrical CHH context (5.2% CHH), suggesting that methylation was initiated on the transgene earlier in development compared to the floral bud stage we assayed, and we are detecting only the replicated symmetrical methylation from this previous event. In \textit{rdr6} mutants, the methylation level of mock-induced ind-\textit{Athila} (31.2% CG, 34.9% CHG, 7.7% CHH) was similar compared to wt Col, demonstrating that without TE expression RDR6 plays no role in the initiation of silencing and \textit{de novo} methylation (Figure 4C). Alternatively, pol IV/nrpD1 mutants display less CG and CHG methylation even without induction of expression (9.4% CG, 13.2% CHG), demonstrating that Pol IV-RdDM functions in a Pol II expression-independent manner (or at least with very little expression) to initiate TE methylation.

Upon estradiol-induction of expression RDR6-RdDM functions to establish CHH methylation. When ind-\textit{Athila} is estradiol-treated in wt Col plants, the CHH methylation triples from 5.2% in mock-induced wt Col to 15.5% in induced wt Col (Figure 4C). This demonstrates that externally activating TE expression and increasing transcription from the locus does not directly remove DNA methylation. The expression-dependent doubling of CHH methylation levels found in wt Col is not detected in estradiol-induced \textit{rdr6} mutants (3.7% CHH), demonstrating that RDR6-RdDM is required for Pol II expression-dependent CHH methylation (Figure 4C). In addition, there is a reduction in the total methylation levels (CG and CHG
contexts) in estradiol-induced \textit{rdr6} mutants compared to estradiol-induced wt Col plants (Figure 4C). Methylation in all sequence contexts is lowest in \textit{pol IV/nrpD1} estradiol-induced plants (5.7\% CG, 9.9\% CHG, 2.8\% CHH), demonstrating that Pol IV-RdDM plays a larger cumulative role in the \textit{de novo} initiation of TE methylation compared to RDR6-RdDM.

**The RDR6-RdDM pathway functions to correctly reestablish methylation and silencing**

We next aimed to determine if the RDR6-RdDM pathway functions to correctly reestablish silencing of recently reactivated TEs. The Pol IV-RdDM pathway has been previously identified as necessary to correctly re-silence TEs that were transcriptionally active in a \textit{ddm1} mutant (Teixeira et al., 2009). To test the corrective reestablishment of silencing of recently reactivated TEs we conducted a series of genetic crosses between mutant \textit{ddm1} and WT DDM1 plants (Table 1). For the majority of TEs, the transcriptionally reactivated state in \textit{ddm1} mutants is inherited and TEs retain transcriptional activity during the subsequent formation of \textit{ddm1/+} heterozygous progeny (Kakutani et al., 1999). In contrast, we find that \textit{Athila6} re-silencing of \textit{gag/pol} expression occurs very efficiently and quickly in an F1 \textit{ddm1/+} heterozygote (Figure 5A), which is likely due to the abundant siRNAs these TEs produce when active (Figure 3A). This re-silencing occurs as early as in the +/\textit{ddm1} F1 seedling (Supplemental Figure 2). We used \textit{ddm1} double mutants to perform a TE re-silencing assay with parents that were both homozygous for a mutation in a second gene that produces a protein involved in either the RDR6-RdDM or Pol IV-RdDM pathway (see Table 1). We found that \textit{Athila6} expression is not fully re-silenced in mutant backgrounds of either the Pol IV-RdDM pathway or the RDR6-RdDM pathway (Figure 5A). The Pol IV-RdDM pathway mutants (\textit{rdr2}, \textit{dcl3}, \textit{pol IV/nrpD1}) play a larger role in this re-silencing, as their \textit{Athila6} expression levels are on average higher than the RDR6-RdDM pathway mutants. However, the RDR6-RdDM pathway mutants (\textit{rdr6} and \textit{dcl2}) do show statistically significant (p<0.001) expression that is greater than the control cross of wt Col x \textit{ddm1}, demonstrating that both the Pol IV-RdDM and RDR6-RdDM pathways play a role in the re-silencing of \textit{Athila6} expression.

We next determined if the lack of full re-silencing in an \textit{rdr6} mutant background is due to incomplete reestablishment of DNA methylation. We performed bisulfite DNA sequencing on the progeny of the crosses in Table 1, and found that the expression levels in Figure 5A inversely correlate with the methylation levels of the \textit{Athila6} LTR TSS. The high level of CG and low level of CHH methylation found in wt Col is restored in the progeny of wt Col x \textit{ddm1} (86\% CG and
9% CHH)(Figure 5B). However, when this cross is performed in an *rdr6* background, CG and CHH methylation levels are only partially restored (57% CG and 27% CHH)(Figure 5B). This data demonstrates that RDR6 is necessary for the complete re-methylation and re-silencing of *Athila6* TEs. In the *pol V/nrpE1* mutant background, which encodes what is likely a shared component of the RDR6-RdDM and Pol IV-RdDM pathways, *ddm1* heterozygous progeny from Table 1 display very low CG and CHH methylation levels (33% CG and 8% CHH)(Figure 5B), resulting in the highest levels of *Athila6* expression (Figure 5A). The combined data from our crosses performed in Table 1 demonstrate that at least for *Athila6*, RDR6-RdDM and Pol IV-RdDM both function to correctively reestablish TE DNA methylation. However, RDR6-RdDM plays a minor role compared to Pol IV-RdDM in the reestablishment of TE transcriptional silencing.

**RDR6-RdDM does not globally act to maintain TE silencing**

We next aimed to determine if RDR6 and RDR6-RdDM play a role in maintaining the repression of TE expression in a wt epigenome with transcriptionally silenced TEs. There are few RDR6-dependent TE siRNAs when TEs are transcriptionally silenced (Figure 1D)(Kasschau et al., 2007), and only a few individual TE copies have been reported to transcriptionally activate in *rdr6* single mutants (Pontier et al., 2012). In addition, we find that in a wt Col epigenome restrictive to TE activity, RDR6 plays no role in the maintenance of TE methylation or expression of either *AtENSPM6* (Figure 2D-E) or *Athila6* (Figure 3C-E). Recent analysis of the global methylation change in *rdr6* seedlings has shown that genes, and not TEs, are the targets of RDR6-RdDM in a wt Col epigenome (Stroud et al., 2013). We used publicly available gene expression microarray data to assay global TE expression in three biological replicates of wt Col and *rdr6* single mutant plants (Allen et al., 2005). We characterized the microarray expression values of 1155 TEs (see Materials and Methods) and found 95% of TEs examined have no or less than 2-fold expression difference in *rdr6* compared to wt Col, and no TE showed a >2.5-fold increase in expression in *rdr6* mutants (Figure 6A), suggesting that RDR6-RdDM plays no global role in the maintenance of transcriptionally silencing TE expression. Pontier et al identified two individual TEs (*AtRep1* and *AtCopia18A*) for which RDR6-RdDM functions to maintain cytosine methylation patterns and silenced expression (Pontier et al., 2012). We did not detect any RDR6-dependent siRNAs for these TE sub-families (Figure 1D) and these TEs are not present on the ATH1 microarray. However, we did identify two TE sub-families (*AtRep10C* and
that have RDR6-dependent 21-22 nt siRNAs in a wt Col epigenome (Figure 1D-E). We focused on the individual \textit{AtRep10C} element At5TE55255 and the \textit{RomaniAT5} sub-family to attempt to detect any role of RDR6-RdDM in the maintenance of TE silencing. We found that both the \textit{AtRep10C} element At5TE55255 and the \textit{RomaniAT5} sub-family display no loss of DNA methylation or statistically significant increase in steady-state expression levels in \textit{rdr6} mutants (Figure 6B-E). In contrast to symmetrical DNA methylation and Pol IV-RdDM, our data demonstrates that RDR6-RdDM has no or very little function in maintaining TE silencing.
Discussion

In this report we used the AtENSPM6 TE to demonstrate that methylation can be established by RDR6-dependent 21-22 nt siRNAs through a separate pathway from Pol IV-RdDM. We found that RDR6 and Pol IV represent distinct entry points for different sized siRNAs into a RdDM mechanism that shares the components AGO6 and Pol V, and these two entry points work independently and additively on TE targets to establish TE methylation. We used small RNA deep sequencing to determine that TEs only produce RDR6-dependent 21-22 nt siRNAs when they are transcriptionally active, and we have identified 15 TE sub-families that constitute this pool of siRNAs and are therefore potential targets of RDR6-RdDM. The TE copy number does not correlate with siRNA production, and hence copy number does not account for the ability of a TE to produce RDR6-dependent siRNAs when transcriptionally active. Therefore, why only these 15 TE sub-families generate RDR6-dependent siRNAs is currently unknown.

The Athila super-family of retrotransposons contributes the vast majority of RDR6-dependent 21-22 nt siRNAs and is the major target of RDR6-RdDM when transcriptionally active. Athila may generate large quantities of these siRNAs due to the tight distribution of Athila to the centromere core, their high copy number, highly nested configuration, and/or their ancient integration into the plant genome, each of which differs from the average Arabidopsis TE family including other LTR retrotransposons (reviewed in Slotkin, 2010).

We demonstrated that both RDR6-RdDM and Pol IV-RdDM regulate Athila6 expression levels. However, RDR6-RdDM may not regulate the expression of all 15 potential RDR6-RdDM target TE sub-families, as some methylation does not result in a change in expression, similar to the methylation established at TAS loci (Wu et al., 2010). For example, the AtENSPM6 TE is a direct target of RDR6-RdDM, but this methylation does not affect the TE’s transcript levels.

These differences may be due to the region of each TE that is targeted by RdDM. Methylation of promoters and regulatory regions by RdDM will likely affect the TE’s expression state, whereas the methylation of Pol II-transcribed protein-coding regions may not. For example, the Athila6 promoter / TSS is targeted by RdDM, resulting in the regulation of expression, while methylation of the AtENSPM6 protein-coding region does not result in regulation of expression. Retrotransposons may be more efficient targets for transcriptional repression by RDR6-RdDM compared to DNA transposons or genes, because their promoters and regulatory information must be converted into RNA for element duplication. This may account for why LTR retrotransposons, such as Athila, are targets of RDR6-RdDM expression regulation, because...
they contain a downstream transcribed LTR that is processed into 21-22 nt siRNAs which match the 5’ LTR containing the promoter and TSS.

In contrast to previous findings, we find that RDR6 and RDR6-RdDM play or no role (or a very minor role) in the maintenance of TE silencing. Transcriptionally silenced TEs in a wt Col epigenome produce very few RDR6-dependent siRNAs, while even the two TE sub-families we identified with RDR6-dependent siRNAs (AtREP10C and RomaniAT5, Figure 1) do not show increased expression in rdr6 mutants. Furthermore, both AtENSPM6 and Athila6 methylation and expression is unaltered in rdr6 single mutants compared to wt Col. The individual TEs identified by Pontier et al that are maintained in a silenced state by RDR6-RdDM may be slightly Pol II transcriptionally active, but quickly turned over into siRNAs. We theorize that these and/or other individual TEs in the Arabidopsis genome may retain their activity in the repressive wt Col epigenome due to the influence of their local euchromatic context and adjacent genes. However, the vast majority of Arabidopsis TEs are Pol II transcriptionally inactive. In fact, both the AtENSPM6 and Athila6 TEs are so efficiently and deeply silenced in wt Col that their transcriptional silencing is not dependent on the Pol IV-RdDM pathway to continuously re-target methylation and silencing. Although these two TEs continue to be re-targeted by the Pol IV-RdDM pathway to produce 24 nt siRNAs, their expression is only dependent on the maintenance of symmetrical DNA methylation (Figures 2E, 3E)(Kato et al., 2003; Lister et al., 2008), while other TEs are not as deeply silenced and therefore are still dependent on constant re-targeting of DNA methylation by Pol IV-RdDM (Herr et al., 2005).

In contrast to maintenance of TE silencing, RDR6-RdDM plays a significant role in the corrective reestablishment of TE silencing. When Athila6 is transcriptionally active, RDR6-RdDM functions to establish high levels of CHH methylation at the LTR TSS, apparently in an attempt to re-silence this TE. However, in the ddm1 mutant background this re-silencing of expression cannot be established (Figures 2-3). When a ddm1 mutant individual with transcriptionally active TEs is outcrossed, the resulting F1 hybrid now has a functional DDM1 protein, and methylation induced by both Pol IV-RdDM and RDR6-RdDM establishes re-silencing of the TE (Figure 5). Compared to RDR6-RdDM, Pol IV-RdDM plays a larger role in correctly reestablishing transcriptional silencing of Athila6, as well as in the initiation of silencing of the ind-Athila6 transgene (see below). This larger dependence on Pol IV-RdDM may be due to the previous finding that in pol IV and rdr2 mutants, the AGO6 protein does not accumulate to its normal wt Col levels (Havecker et al., 2010). AGO6 is a shared component of the RDR6-RdDM and Pol IV-
RdDM pathways (Wu et al., 2012), and when limited in pol IV and rdr2 mutants, an indirect dampening effect of RDR6-RdDM is expected. Therefore, the reduction of TE methylation and increase in TE expression in Pol IV-RdDM pathway mutants nrpD1 and rdr2 are overestimations of the impact of the Pol IV-RdDM pathway, accounting for the observed larger contribution of Pol IV-RdDM compared to RDR6-RdDM in the initiation and corrective reestablishment of TE silencing. However, why similar levels of partial methylation in +/ddm1 rdr2 compared to +/ddm1 rdr6 progeny from Figure 5 result in different expression levels is currently unknown. Another puzzling question remains: When transcriptionally activated in ddm1 mutants, why are Athila6 24 nt siRNAs dependent on RDR6, while all other TE 24 nt siRNAs (including from AtENSPM6) are RDR6-independent? We hypothesize that Athila6 methylation induced by RDR6-RdDM indirectly helps establish the Pol IV-RdDM template. We speculate that without the initiation of CHH methylation via RDR6-RdDM at the Athila6 TSS, the low level of methylation at this region results in a loss of Pol IV targeting to this region. Therefore, without RDR6, the drop in methylation results in less Pol IV transcription and fewer 24 nt siRNAs. Conversely, increased methylation by RDR6-RdDM may generate more Pol IV transcription from these loci, creating more 24 nt siRNAs.

In addition to the corrective reestablishment of TE silencing, RDR6-RdDM functions in the de novo initiation of TE silencing. By analyzing the initiation of silencing of ind-Athila6 (Figure 4), we can separate the distinct activities of Pol IV-RdDM from RDR6-RdDM. Pol IV-RdDM functions independent of ind-Athila6 expression, or at least with very low detectable expression levels. This silencing is likely due to the abundant 24 nt siRNAs that the endogenous Athila6 elements consistently produce, and silencing must be dependent on homology between the siRNAs and the target TE. In contrast, RDR6-RdDM functions only after the induction of expression from ind-Athila6. Similar to the corrective reestablishment of TE silencing, pol IV mutants have a more severe effect on the initiation of element silencing, but again this could be due to Pol IV’s influence on AGO6 protein accumulation (see above). Whether RDR6-RdDM function is dependent on homology to the already present and silenced TEs in the genome is currently unknown. Alternatively, RDR6-RdDM may function in a homology-independent manner by somehow recognizing Pol II-derived TE transcripts and initiating silencing at these loci.

Two general mechanisms to initiate TE silencing
The distinction in the epigenetic state of their targets (silenced TEs vs. transcriptionally active TEs) allows us to propose a model of separate Pol IV-RdDM and RDR6-RdDM functions. The distinction is best illustrated considering two examples of a transcriptionally active TE entering a genome. In the first instance the active TE enters the genome through cross hybridization within a single species or from two closely related species. If these two species share a recent common ancestor (and therefore can mate), their TEs will share homology to each other. If one of the two parent species contains a TE copy that has been previously silenced and whose silencing is maintained by the Pol IV-RdDM pathway, the complementarity between the silenced TE’s Pol IV-dependent 24 nt siRNAs and the other parent’s transcriptionally active TE(s) will result in homology-dependent trans-silencing of the active TEs. Where and when in the F1 hybrid plant this silencing occurs remains an open question, but if we use the model of the corrective reestablishment experiment that we performed (with one parent having transcriptionally active TEs while the other has the same TEs in a silenced state, Figure 5), the F1 silencing occurs very early during embryogenesis. In this way, the previously silenced TEs in a plant genome act as a battle history, cellular memory and inoculum for any incoming homologous active TEs (Jensen et al., 1999). We theorize that Pol IV-RdDM plays a larger role in our corrective reestablishment and initiation of TE silencing experiments due to at least one parent in each of these experiments having silenced Athila6 elements and producing Pol IV-dependent Athila6 24 nt siRNAs.

The second instance of TE silencing considers active TEs that are unique to the genome they enter. This may occur through horizontal transfer, which on an evolutionary timescale may be a somewhat common event for TEs in plant genomes (Diao et al., 2006; Roulin et al., 2008). Due to the unique TE sequence, the cell has no ability to silence the active TE based on homology. We speculate that RDR6-RdDM functions in the recognition of Pol II-derived transcripts from this new TE. How the RDR6-RdDM pathway recognizes TE transcripts is not understood, but evidence suggests that mobilization of active TEs often produces spontaneous rearrangements or nested elements that drive the production of double-stranded RNA, triggering RNAi and siRNA production (Slotkin et al., 2005). Once produced, these 21-22 nt siRNAs can initiate TE DNA methylation through RDR6-RdDM. Since many organisms do not have Pol IV and 24 nt siRNAs, the Pol II-derived transcript-identification mechanism may be the evolutionarily older mechanism responsible for surveillance of the transcriptome and targeting active TEs for initiation of silencing, which in some cases in animals must be reset each
generation (reviewed in Smallwood and Kelsey, 2011). Once established by RDR6-RdDM, low levels of TE methylation in plants will be reinforced by the methylation-dependent Pol IV transcription of this TE locus and re-targeting by Pol IV-RdDM. Over time this TE will become deeply silenced and only dependent on symmetrical DNA methylation to propagate its trans-generational epigenetic silencing. It will produce the 24 nt siRNAs required for initiating Pol IV-RdDM homology-dependent silencing of any incoming active TEs with sequence similarity.
Materials and Methods

Plant Material
The mutant alleles used in this study are described in Supplemental Table 1. Plants were grown under standard long day conditions at 23°C. Inflorescence tissue was used in each experiment unless otherwise noted.

Expression Analysis by qRT-PCR
Three biological replicates were performed for each genotype. Each replicate consisted of a non-overlapping pool of individuals. qRT-PCR was performed and analyzed as in McCue et al. (McCue et al., 2012), with the exception that cDNA was generated using an oligo-dT primer and Tetro Reverse Transcriptase (Bioline). Quantitative Real-Time PCR was performed using SensiMix (Bioline) on a Mastercycler ep realplex thermocycler (Eppendorf). The At1g08200 gene was used as a reference in all qRT-PCR reactions. qRT-PCR primers are listed in Supplemental Table 1.

Bisulfite Conversion and Sequencing
DNA was bisulfite converted using the EZ DNA Methylation-Gold kit (Zymo Research). Converted DNA was amplified using EpiTaq DNA polymerase (Takara) using the primers listed in Supplemental Table 1. PCR products were TOPO-TA cloned into pCR4 (Invitrogen) and sequenced. Analysis of individual DNA sequences was performed using the Kismeth analysis tool (Gruntman et al., 2008). For each bisulfite-converted DNA sample, an unmethylated exon of the gene At2g20610 was amplified, and at least three clones were sequenced to determine the C→T conversion efficiency, which is listed on each figure. Error bars represent Wilson score interval 95% confidence limits (Henderson et al., 2010). Differences between genotype methylation status were analyzed with a two-tailed Student’s t-test. Each cloned sequence was given a methylated percentage based on the number of unconverted cytosines that were present out of total cytosines in the sequence. Each methylation context (CG, GHG, CHH or total) was analyzed independently in this manner. Individual methylation percentages for each clone were grouped into populations by genotype, and these populations were compared with a student’s t-test. This statistical approach is more stringent compared to the Chi-squared tests previously used for statistical comparisons of bisulfite data (Henderson et al., 2010), as it takes into account clone-to-clone methylation variation. All of our bisulfite methylation comparisons
that were statistically significant (p<0.05) using the student’s t-test were also statistically significant at the p<0.0001 level using the Chi-squared approach.

**Northern Blotting**

Total RNA was isolated using Trizol reagent (Invitrogen), and small RNAs were concentrated with polyethylene glycol. Fourteen micrograms of small RNA-enriched RNA were loaded in each lane. Gel electrophoresis, blotting and cross-linking were performed as described in Pall *et al* (Pall et al., 2007). *Athila* probes were generated by randomly degrading a P32-labelled *in vitro* RNA transcript. PCR primers used to generate the probes are listed in Supplemental Table 1. The miR161 control probe was generated by 5’ end labeling the DNA oligonucleotide shown in Supplemental Table 1.

**Transgene Construction and Expression**

The *Athila6* inducible expression transgene (ind-*Athila*) was produced using the primers listed in Supplemental Table 1 from *ddm1* oligo-dT primed cDNA. This fragment was cloned into vector pMDC7. Sequencing the transgene identified the *Athila6A* element At5g32197 as the specific element that was incorporated into this transgene. The inducible RAN1 control transgene was constructed using the same pMDC7 vector and expresses a HA-tagged version of the RAN1 (At5g20010) gene. Induction was achieved using the 20 uM estradiol spray from Borghi (2010) or mock induction using the same spray without the inclusion of estradiol. T1 plants for the ind-*Athila* transgene were sprayed as they transitioned to flowering, and were sprayed every day for 5 or more days before inflorescence tissue collection. Biological replicates consisted of three non-overlapping pools of T1 sibling plants. Each biological replicate consisted of at least three individual T1 plants. Primers used for qRT-PCR and bisulfite sequencing are listed in Supplemental Table 1. Bisulfite amplification utilized one transgene-anchored primer to ensure amplification of ind-*Athila* only, and not other endogenous *Athila6* elements.

**Small RNA Library Production, Sequencing and Analysis**

Inflorescence small RNA was isolated with Trizol reagent (Invitrogen) and concentrated using the *miR*Vana microRNA isolation kit (Ambion). Libraries were produced using the TruSeq Small RNA Sample Preparation Kit (Illumina) as recommended by the manufacturer. Each library was barcoded and sequenced in the same lane of an Illumina Genome Analyzer IIx. The resulting
sequences were de-multiplexed, adapter trimmed, and filtered on length and quality, and tRNA/rRNA and low complexity reads were removed. Small RNAs were matched to the Arabidopsis genome, and sequences that did not perfectly align were discarded. Library size is normalized by calculating reads per million of 18-28 nt genome-matched small RNAs. Small RNAs were also matched to the TAIR10 (www.arabidopsis.org) and Repbase (www.girinst.org) annotation of the TE portion of the Arabidopsis genome using bowtie. To best handle multi-mapping sequences generated from repetitive regions of the genome, the bowtie modifiers --best -M1 --strata were employed (Treangen and Salzberg, 2012). If more than one genome perfect match for a TE siRNA exists, only one random match is assigned per small RNA read. Counts of multi-mapping siRNAs were not amplified using our approach, and we have not overestimated read number for high copy TEs. Small RNA tracks and display of the data in Figure 2B were performed using the Integrated Genome Browser (Nicol et al., 2009). The raw sequencing and genome-matched small RNAs analyzed are available from NCBI GEO repository number GSE41755.

Analysis of TE expression from microarrays

Microarray analysis of gene expression in wt Col and rdr6 single mutant inflorescences was performed by Allen et al (GSE2473)(Allen et al., 2005). 1155 TE probes have previously been identified on the ATH1 gene expression microarray (Slotkin et al., 2009). Fold change in TE expression was calculated by NCBI GEO2R.

Acknowledgements

The authors thank Sarah H. Reeder, Xiao Zhou, Gregory Booton and Yao Wan for their contributions, reagents and statistical advice. A.D.M. is a fellow of The Ohio State Center for RNA Biology. D.F. is supported by NIH Training Grant T32 GM086252. C.D. is supported by The Ohio State University’s Pelotonia Post-Doctoral Fellowship. This work was supported by grant MCB-1020499 to R.K.S from the U.S. National Science Foundation.

Literature Cited


argonautes functionally diverge based on their expression and interaction with target loci. Plant Cell 22: 321–334


McCue AD, Nuthikattu S, Reeder SH, Slotkin RK (2012) Gene Expression and Stress
Response Mediated by the Epigenetic Regulation of a Transposable Element Small RNA. PLoS Genet 8: e1002474


Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccara M, Ciaudo C, Cruaud C, Poulain J,


Figure Legends

Figure 1. Identification of TEs that produce RDR6-dependent siRNAs.
(A) The accumulation of 21, 22 and 24 nt small RNAs in the four genotypes sequenced. Library size is normalized by calculating reads per million of 18-28 nt genome-matched small RNAs. An increase in *ddm1* 21 nt siRNAs is apparent in the total small RNA pool.
(B) TasiRNAs were investigated as a library control. TasiRNAs of specifically the 21 nt size accumulate only in genotypes with a functional RDR6 protein.
(C) Accumulation of TE-derived siRNAs. Few 21 or 22 nt siRNAs accumulate from TEs in any genotype except for *ddm1*.
(D) Comparison of 21-22 nt TE siRNA accumulation between wt Col and *rdr6* single mutants. Each point represents a TE sub-family. Only two TE sub-families (*AtRep10C* and *RomaniAT5*) have a 3-fold or more difference between wt Col and *rdr6* (red dots). The *AtRep1* and *AtCopia18A* sub-families (green dots) show no dependence on RDR6 for accumulation of their siRNAs.
(E) The small RNA accumulation of the *AtRep10C* and *RomaniAT5* TE sub-families.
(F) 15 TE sub-families have 3-fold or higher levels of 21 or 22 nt siRNAs in *ddm1* compared to wt Col. These siRNAs are dependent on RDR6, as they are lost in *ddm1 rdr6* double mutants.
(G) TE siRNA distribution when the 15 sub-families identified in part F are removed from the analysis. All of the 318 TE sub-families (with the exception of the 15 subtracted from this data) display low levels of 21 and 22 nt siRNAs and high accumulation of 24 nt siRNAs in all four genotypes.
(H) The combined 21 and 22 nt siRNA accumulation of the identified 15 sub-families with increased RDR6-dependent siRNA production in *ddm1*. The relative contribution of the sub-families is shown.

Figure 2. CHH hypermethylation of the transcriptionally active *AtENSPM6* TE is dependent on both the RDR6-RdDM and Pol IV-RdDM pathways.
(A) Profile of siRNAs derived from *AtENSPM6* sub-family TEs.
(B) Small RNA profile of the first 5000 bp of the 8825 bp *AtENSPM6* consensus element. The number of reads per million of 21, 22 and 24 nt siRNAs are mapped for wt Col, *ddm1* and *ddm1*
rdr6. The scale of each track is 1-100 reads per million. The red-box represents the region further interrogated in parts C-E. TIR = terminal inverted repeat.

(C) Profile of siRNAs derived from the red-boxed region of AtENSPM6 in part B.

(D) Bisulfite sequencing of DNA methylation levels of the AtENSPM6 region from the red box in part B. Methylation in the CG (red), CHG (blue) and CHH (green) sequence contexts are shown (H=C,T or A). The number of clones sequenced and the C→T conversion efficiency for each bisulfite reaction (as judged by sequencing an a non-methylated genic exon) are shown for each sample. This region of AtENSPM6 undergoes hypermethylation in ddm1 mutants. This hypermethylation is reduced in ddm1 double mutants with the RDR6-RdDM pathway (ddm1 rdr6), the pol IV mutant of the Pol IV-RdDM pathway (ddm1 nrpD1) and the shared component Pol V (ddm1 nrpE1).

(E) qRT-PCR analysis of the steady-state polyadenylated transcript accumulation of AtENSPM6. ns = not a statistically significant difference.

Figure 3. The expression-dependent CHH hypermethylation of the Athila6 transcriptional start site is dependent on both RDR6-RdDM and Pol IV-RdDM.

(A) The accumulation of Athila6 sub-family-derived siRNAs.

(B) The accumulation of siRNAs from the Athila6 long terminal repeat (LTR) transcriptional start site (TSS) region assayed for DNA methylation by bisulfite sequencing in part C.

(C) Bisulfite sequencing of the Athila6 LTR gag/pol TSS demonstrates that components of both the RDR6-RdDM and Pol IV-RdDM pathways are required to attain the full CHH hypermethylation present in ddm1 single mutants. Two biological replicates of ddm1 rdr6 were performed.

(D) Small RNA Northern blot of the Athila6 LTR TSS region. All 24 nt siRNA production is dependent on RDR2 and Pol IV/NRPD1. All 21-22 nt siRNA production is dependent on RDR6. MiR161 is used as a loading control.

(E) qRT-PCR analysis of the steady-state polyadenylated gag/pol transcript levels of Athila6 demonstrate that the expression level increases in ddm1 double mutants that lose hypermethylation compared to ddm1 single mutant plants.

Figure 4. RDR6-RdDM dependent initiation of TE silencing is dependent on expression.
(A) qRT-PCR analysis of the expression of *Athila6* in plants without the inducible *Athila* (ind-*Athila*) transgene, with the mock-induced transgene, and with estradiol-induction of transgene expression.

(B) Small RNA Northern blot of *Athila6* siRNAs. The production of 21-22 nt siRNAs from the estradiol-induced ind-*Athila* transgene is dependent on RDR6. MiR161 is shown as a loading control.

(C) Bisulfite sequencing of the *Athila6* region of the ind-*Athila* transgene in mock- and estradiol-induced conditions. The non-TE containing ind-RAN1 transgene is used as a control to separate background transgene silencing from TE silencing.

**Figure 5. RDR6-RdDM functions in the corrective reestablishment of TE silencing.**

(A) qRT-PCR analysis of *Athila6* gag/pol transcript accumulation in the progeny of the crosses detailed in Table 1. Active *Athila6* TEs in a *ddm1* background are efficiently re-silenced upon crossing to wt plants. Mutations in the Pol IV-RdDM and RDR6-RdDM pathways result in only partial re-silencing of the *Athila6* gag/pol transcript.

(B) Bisulfite sequencing of the *Athila6* LTR around the gag/pol TSS demonstrates that RDR6, RDR2 and Pol V are required for efficient return of methylation levels to their wt states of high CG and low CHH. Plants examined are the progeny of the crosses detailed in Table 1.

**Figure 6. RDR6-RdDM does not function to maintain TE silencing.**

(A) Comparison of TE expression levels between wt Col and *rdr6* mutant inflorescences. Microarray gene expression values for 1,155 individual TEs are shown (see Materials and Methods). In this boxplot, the centerline represents the median fold change; the upper and lower box borders are the 75% and 25% percentile fold-change, while the whiskers represent the 90% and 10% percentile. Individual outlier TEs are shown as points, however, no single TE has a >2.5-fold increase (red line) in expression in *rdr6* single mutants compared to wt Col. The gene ARF4 (At5g60450) is shown as a positive control for a gene with increased expression in *rdr6* mutants, as this gene is targeted by RDR6-dependent tasiRNAs (Allen *et al.*, 2005).

(B) Bisulfite sequencing of the DNA methylation levels of the *AtRep10C* TE At5TE55255. No reduction of methylation is detected in *rdr6* mutants.

(C) qRT-PCR analysis of the steady-state polyadenylated expression from the *AtRep10C* TE At5TE55255. No increase in expression is detected in *rdr6* mutants.
(D) Bisulfite sequencing of the DNA methylation levels of the *RomaniAT5* TE sub-family.

(E) qRT-PCR analysis of the steady-state polyadenylated expression from the *RomaniAT5* TE sub-family. Expression reactivated in a *ddm1* mutant is shown for comparison. The expression difference between wt Col and *rdr6* is not statistically significant (ns)(p>0.05).
Supplemental Figure Legends

**Supplemental Figure 1. Start site of the Athila6 gag/pol transcript.**

(A) The transcriptional start site of the *Athila6 gag/pol* transcript within the 5' LTR as mapped by 5' RACE RT-PCR. The sequence and number corresponds to the consensus *Athila6A* LTR compiled by Repbase. Fifteen RACE clones were sequenced, and the start sites were mapped. Each red arrow points to the mapped transcript initiation site, with the number of times that start site was identified indicated above the arrow.

**Supplemental Figure 2. Athila6 re-establishment of silencing occurs early in plant development.**

Expression of the *Athila6 gag/pol* transcript as determined by qRT-PCR in (A) juvenile leaves or (B) whole seedlings. Plants heterozygous for *ddm1 (+/ddm1)* were generated from the parents indicated in Table 1.
Table 1. Crosses performed to test the role of RDR6-RdDM and PolIV-RdDM components in TE re-silencing

<table>
<thead>
<tr>
<th>Female Parent</th>
<th>Male Parent</th>
<th>Progeny Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt Col</td>
<td>ddm1 F3</td>
<td>+/-ddm1</td>
</tr>
<tr>
<td>ddm1 F3</td>
<td>wt Col</td>
<td>ddm1/+</td>
</tr>
<tr>
<td>rdr6</td>
<td>ddm1 rdr6</td>
<td>+/-ddm1 rdr6</td>
</tr>
<tr>
<td>dcl2</td>
<td>ddm1 dcl2</td>
<td>+/-ddm1 dcl2</td>
</tr>
<tr>
<td>rdr2</td>
<td>ddm1 rdr2</td>
<td>+/-ddm1 rdr2</td>
</tr>
<tr>
<td>dcl3</td>
<td>ddm1 dcl3</td>
<td>+/-ddm1 dcl3</td>
</tr>
<tr>
<td>nrpD1</td>
<td>ddm1 nrpD1</td>
<td>+/-ddm1 nrpD1</td>
</tr>
<tr>
<td>nrpE1</td>
<td>ddm1 nrpE1</td>
<td>+/-ddm1 nrpE1</td>
</tr>
</tbody>
</table>
Figure 1

A. All small RNAs

B. tasRNAs

C. TE small RNAs

D. rdr6 21+22 nt RPM vs. Col 21+22 nt RPM

E. RomaniAT5 vs. AtRep10C

F. 15 TE sub-families

G. TE small RNAs minus 15

H. 15 TE sub-families with increased 21&22 nt siRNAs
Figure 2

A

**AtENSPM6**

whole element

<table>
<thead>
<tr>
<th></th>
<th>24 nt</th>
<th>22 nt</th>
<th>21 nt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Col</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rdr6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ddm1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ddm1 rdr6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

Col 24 nt

Col 22 nt

Col 21 nt

ddm1 24 nt

ddm1 22 nt

ddm1 21 nt

ddm1 rdr6 24 nt

ddm1 rdr6 22 nt

ddm1 rdr6 21 nt

AtENSPM6 methylations (%): 97.4, 98.2, 97.3, 96.1, 100, 98.3

D

sequences: 9, 9, 11, 23, 7, 8

control gene conversion %: 97.4, 98.2, 97.3, 96.1, 100, 98.3

E

**AtENSPM6** bisulfite region
**Figure 3**

**A**

*Athila6 siRNAs*

![Bar graph showing reads per million for different genotypes: Col, rdr6, ddm1, ddm1 rdr6. Bars for 24nt, 22nt, and 21nt siRNAs are shown.]

**B**

*Athila6 5’ LTR TSS*

![Bar graph showing reads per million for different genotypes: Col, rdr6, ddm1, ddm1 rdr6. Bars for 24nt, 22nt, and 21nt siRNAs are shown.]

**C**

<table>
<thead>
<tr>
<th>Control gene conversion %</th>
<th>30</th>
<th>8</th>
<th>14</th>
<th>8</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>9</th>
<th>8</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>97.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**D**

*Athila6 LTR TSS methylation (%)*

![Chart showing LTR TSS methylation percentages for different genotypes: Col, rdr6, ddm1, ddm1 rdr6, etc. SD bars are shown.]

**E**

*Atlhila6 gag/pol relative expression*

![Bar graph showing relative expression of gag/pol for different genotypes: Col, rdr6, nrpD1, etc. SD bars are shown.]

*p<0.001*
Figure 4

A

![Graph showing relative expression of Athila6](image)

B

![Images of sequencing results](image)

C

<table>
<thead>
<tr>
<th>sequences:</th>
<th>12</th>
<th>13</th>
<th>9</th>
<th>16</th>
<th>12</th>
<th>22</th>
<th>16</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>control gene conversion %:</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99.2</td>
<td>100</td>
<td>100</td>
<td>98.3</td>
<td>99.2</td>
</tr>
</tbody>
</table>
Figure 5

A

controls

RDR6-
RdDM

Pol IV-
RdDM

Athila6 gag/pol relative expression

wt Col
+/-ddm1
+/-ddm1 rdr6
+/-ddm1 dcl2
+/-ddm1 dcl3
+/-ddm1 nrpD1
+/-ddm1 nrpE1

B

sequences:
control gene conversion %:
30 14 8 8 8 8 8
97.4 97.3 98.3 98.3 100 99.4

Athila6 LTR TSS methylation (%)

wt Col
ddm1
+/-ddm1 rdr6
+/-ddm1 dcl2
+/-ddm1 dcl3
+/-ddm1 nrpD1
+/-ddm1 nrpE1

CG
CHG
CHH

p<0.05

p<0.05
Figure 6

A

Absolute Fold Change (rdr6/ wt Col)

wt Col vs rdr6 inflorescence microarray expression

B

sequences: control gene conversion %:

AtRep10C

C

Relative AtRep10C expression

D

Relative RomaniAT5 expression

E

ns