Mobile small RNAs regulate genome-wide DNA methylation

Mathew Graham Lewsey, Thomas J. Hardcastle, Charles W. Melnyk, Attila Molnar, Adrián Valli, Mark A. Urich, Joseph R. Nery, David C. Baulcombe, and Joseph R. Ecker

The Salk Institute for Biological Studies, La Jolla, CA 92037; Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037; Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037; Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom; and Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA 92037

Edited by Steven E. Jacobsen, University of California, Los Angeles, CA, and approved December 15, 2015 (received for review July 29, 2015)

RNA silencing at the transcriptional and posttranscriptional levels regulates endogenous gene expression, controls invading transposable elements (TEs), and protects the cell against viruses. Key components of the mechanism are small RNAs (sRNAs) of 21–24 nt that guide the silencing machinery to their nucleic acid targets in a nucleotide sequence-specific manner. Transcriptional gene silencing is associated with 24-nt sRNAs and RNA-directed DNA methylation (RdDM) at cytosine residues in three DNA sequence contexts (CG, CHG, and CHH). We previously demonstrated that 24-nt sRNAs are mobile from shoot to root in Arabidopsis thaliana and confirmed that they mediate DNA methylation at three sites in recipient cells. In this study, we extend this finding by demonstrating that RdDM of thousands of loci in root tissues is dependent upon mobile sRNAs from the shoot and that mobile sRNA-dependent DNA methylation occurs predominantly in non-CG contexts. Mobile sRNA-dependent non-CG methylation is largely dependent on the DOMAINS REARRANGED METHYLTRANSFERASES 1/2 (DRM1/DRM2) RdDM pathway but is independent of the CHROMOMETHYLASE (CMT)2/3 DNA methyltransferases. Specific superfamilies of TEs, including those typically found in gene-rich euchromatic regions, lose DNA methylation in a mutant lacking 22- to 24-nt sRNAs (dicer-like 2, 3, 4 triple mutant). Transcriptome analyses identified a small number of genes whose expression in roots is associated with mobile sRNAs and connected to DNA methylation directly or indirectly. Finally, we demonstrate that sRNAs from shoots of one accession move across a graft union and target DNA methylation de novo at normally unmethylated sites in the genomes of root cells from a different accession.

RNA-directed DNA methylation | plant grafting | transposable element | small RNA | transcriptional gene silencing

Mobile small RNAs regulate genome-wide DNA methylation

Small RNAs (sRNAs) of 24 nt are associated with transcriptional gene silencing by targeting DNA methylation to complementarity sequences. We demonstrated previously that sRNAs move from shoot to root, where they regulate DNA methylation of three endogenous transposable elements (TEs). However, the full extent of root DNA methylation dependent on mobile sRNAs was unknown. We demonstrate that DNA methylation at thousands of sites depends upon mobile sRNAs. These sites are associated with TE superfamilies found in gene-rich regions of the genome, which lose methylation selectively in an sRNA-deficient mutant. If the TEs were able to reactivate, they could cause genome instability and altered gene expression patterns, with negative effects on the plant. Consequently, mobile sRNAs may defend against these TEs.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the European Nucleotide Archive (accession no. E-MTAB-3473).

Significance

Small RNAs (sRNAs) of 24 nt are associated with transcriptional gene silencing by targeting DNA methylation to complementarity sequences. We demonstrated previously that sRNAs move from shoot to root, where they regulate DNA methylation of three endogenous transposable elements (TEs). However, the full extent of root DNA methylation dependent on mobile sRNAs was unknown. We demonstrate that DNA methylation at thousands of sites depends upon mobile sRNAs. These sites are associated with TE superfamilies found in gene-rich regions of the genome, which lose methylation selectively in an sRNA-deficient mutant. If the TEs were able to reactivate, they could cause genome instability and altered gene expression patterns, with negative effects on the plant. Consequently, mobile sRNAs may defend against these TEs.
recognize non-CG context methylated DNA and methylate lysine 9 of adjacent histone H3 (H3K9me2). CHROMOMETHYLASE (CMT) proteins CMT2 and CMT3 bind H3K9me2 and methylate adjacent non-CG sites (30) of the newly replicated DNA. Redundancy exists between target sites of CMT2 and CMT3, but their predominant functions are to maintain CHH and CHG context DNA methylation, respectively (31). The activity of CMT2 is substantially less efficient than that of MET1, so that CMT2 target sites exhibit variable levels of DNA methylation. CMT3 efficiency is intermediate between MET1 and CMT2. Both CMT2 and CMT3 typically target long TEs and gene-distal TEs (31).

There are also RNA-dependent mechanisms to maintain DNA methylation that involve plant-specific DNA-dependent RNA polymerases IV and V (POL IV and POL V). These polymerases are recruited to chromatin by methyl-DNA-binding proteins SHH1 and SUVH2 and 9 (32). POL IV produces precursor RNAs that are processed into 24-nt sRNAs, whereas POL V produces chromatin-bound scaffold transcripts at sites of DNA methylation (33). Together, they ensure maintenance of CG, CHG, and CHH context DNA methylation through an AGO-dependent mechanism similar to the mechanism that establishes methylation, in which an AGO–sRNA recruits DRM1 and DRM2 to maintain non-CG DNA methylation (8, 25). The target sites of the DRM1/DRM2 DNA methylation maintenance pathway are largely nonoverlapping with those of CMT2 and CMT3, and tend to be short, gene-proximal TEs and the edges of long TEs (31, 34).

RdDM can operate cell-to-cell and systemically due to translocation of 23- to 24-nt sRNAs from shoots to roots (35). In our previous studies, we confirmed that these mobile 23- to 24-nt sRNAs target RdDM and TGS at one transgene and RdDM at three endogenous TEs (35–37). Depletion of shoot sRNAs corresponded to reduction of 23- to 24-nt sRNAs in wild-type roots, indicating that shoot-derived sRNAs contribute to the total root sRNA population (36). In this study, we investigate the extent to which mobile sRNAs mediate genome-wide RdDM. Our approach, as before, is to analyze sRNA and DNA methylation in roots of grafted plants that are defective for the production of the 24-nt sRNA species associated with RdDM. The sRNAs in these grafted plants move predominantly from shoot to root following source–sink gradients (36) and, by grafting different genotypes as shoots, we identify changes in DNA methylation and gene expression in the roots that are dependent on mobile sRNAs.

We show that mobile sRNAs influence genomic DNA methylation at thousands of loci, and that the affected loci are predominantly associated with transposons of specific classes. A very small number of protein-coding genes were influenced by this mobile RdDM. The mobile sRNA-dependent DNA methylation is associated with the DRM1/DRM2 RdDM pathway but not the CMT2/3 DNA methyltransferase pathway. Furthermore, we demonstrate that mobile sRNAs unique to one accession established DNA methylation de novo in unmethylated regions of the genome of a second accession.

Results
Identification of DNA Methylation Loci Targeted by Mobile sRNAs.

The primary aim of our study was to determine how many genomic loci in Arabidopsis may be targeted by RdDM due to the direct action of mobile sRNAs (Fig. 1) and to compare these loci with other sRNA-targeted and cytosine-methylated regions throughout the genome. We reasoned that mobile sRNAs that target RdDM may be associated with specific features of the genome and depend upon specific genes in the RdDM pathway. In addition, we aimed to determine how many of these directly affected loci might influence gene expression. It could be, for example, that DNA methylation, leading to TGS, of a transposon target of mobile sRNA affects expression of an adjacent gene.

In previous work, we compared sRNA populations in roots of WT Arabidopsis thaliana (accessions Col-0, termed Col, and C24) and a dicer-like 2 dicer-like 3 dicer-like 4 triple mutant (Col dcl234) background, termed dcl234 that had been graft to C24, Col, and dcl234 shoots to identify loci producing mobile sRNAs in the Arabidopsis genome. The dcl234 triple mutant is unable to produce 22- to 24-nt sRNAs associated with RNA silencing (23). We reasoned that any 22- to 24-nt sRNAs present in dcl234 roots grafted to WT shoots must have moved from the grafted WT shoot (36). Loci similarly represented by sRNA reads in C24/C24, C24/Col, Col/Col, and C24/dcl234 but not dcl234/dcl234 datasets (notation shoot/root, with underlining indicating the analyzed plant part) could be confidently assigned as producing mobile sRNA, because the WT shoot complements the inability of dcl234 roots to produce sRNA. In contrast, the sRNAs from loci that were absent in C24/dcl234 and dcl234/dcl234 grafts and present in grafts with Col or C24 roots were interpreted as being dependent on DCL2,3,4 but not mobile. Loci where sRNAs were present in all datasets indicate sRNAs produced by DCL1 or in a DCL-independent manner.

We incorporated genome-wide MethylC-seq (sequencing) data Q,16 in our analyses and applied the above reasoning to classify the genome-wide patterns of DNA methylation affected by graft-transmissible sRNAs. We used five genotype combinations in two-way grafts to define six locus classes (A–F) by their combinations of DNA methylation and sRNA representation across graft combinations (Fig. 1, where high and low sRNA abundance are indicated by dark and light blue, respectively, and high and low DNA methylation levels are indicated by dark and light red, respectively; also see SI Appendix, Figs. S1 and S2). Three of these classes were of primary interest:

A) “Direct” loci correspond to overlapping regions of cytosine methylation and sRNAs present in C24/C24, C24/Col, Col/Col, and C24/dcl234 but not dcl234/dcl234. This model
is consistent with direct targeting of DNA methylation by mobile sRNAs.

B) “Indirect” loci show the same pattern of DNA methylation between grafts as direct loci but these were not associated with sRNAs. Here we infer that there are undetectable levels of mobile sRNAs or that the change in methylation is due to indirect effects of a mobile signal.

C) “De novo methylated” loci correspond to overlapping regions of sRNAs and cytosine methylation present in C24/Col, C24/C24, and C24/dcl234 but not in dcl234/dcl234 or Col/Col. These loci likely correspond to sRNAs produced uniquely by C24 that target DNA methylation to previously unmethylated regions of the Col genome.

The class A (direct) and C (de novo) loci because these are most likely to be associated with the mobile sRNA. We compared their characteristics with those of class B (indirect) loci that are associated with a mobile signal but, by definition, do not correspond to detectable candidate sRNAs. Three classes of loci (D–F) not associated with regulation of DNA methylation by mobile sRNA were also identified, and are described in SI Appendix, Results.

Mobile sRNA and DNA Methylation. We developed a statistical framework that could be applied to sRNA and MethylC-seq data to accurately assess the models described in Fig. 1 (see SI Appendix, Materials and Methods, sections 1, 3, and 4 for a detailed explanation). Our previous sRNA abundance data (36) were reanalyzed using this framework to ensure consistency with analyses of the MethylC-seq datasets generated in the current study (Table 1). DNA methylation associated with mobile sRNA was examined separately by sequence context (CG, CHG, CHH), because the genetic requirements for each are distinct.

Our analyses revealed that non-CG DNA methylation is abundant at class A (direct) loci: We identified 13 CG, 398 CHG, and 401 CHH context class A loci with a false discovery rate (FDR) of 5% (Table 1 and Dataset S1). These included the three loci we identified previously as possessing mobile sRNA-associated DNA methylation (36). The association between regions of sRNA and DNA methylation that corresponded to the class A loci model was statistically significantly more frequent than expected by chance (Table 1, Z scores: CG, 154.69; CHG, 77.123; CHH, 84.250; P < 0.05). Furthermore, class A loci of all three DNA methylation contexts overlapped statistically significantly (SI Appendix, Fig. S3). Observation of colocalted DNA methylation in all three contexts is characteristic of RdDM. Fewer than 2.5% of the class A loci overlapped with regions identified as variable over 30 generations in Arabidopsis (18), and thus it is unlikely that the results are due to spontaneous epiallelic variation.

We investigated the genes potentially regulating DNA methylation at class A loci by analysis of MethylC-seq data from 86 Arabidopsis gene-silencing mutant lines (31). We examined the effects of mutation of a subset of genes with known roles in RdDM (dcl234/4, mel1, suv4/5, ddm1, suv4/5/6, vim1/2/3, cmt2, cmt3, and WT control) on DNA methylation at class A loci. We focused on non-CG methylation patterns because very few CG context class A loci were identified. Mutation of dml1/2 caused the strongest reduction in non-CG context methylation in class A loci (Fig. 2 and SI Appendix, Fig. S4). The other mutations examined showed weaker effects, of which cmt2 and cmt3 caused the least changes in DNA methylation: cmt3 reduced CHG context DNA methylation to levels intermediate between WT and dml1/dml2 but had no effect in the CHH context, whereas cmt2 had no effect in either context. DNA methylation was reduced to levels intermediate between WT and dml1/dml2 in both sequence contexts by all others of these mutations. It should be noted that the MethylC-seq data from gene-silencing mutant lines were generated using leaves of 3-wk-old plants, whereas our MethylC-seq data were generated using roots of 3-wk-old plants. We considered it appropriate to compare the two datasets because there is little evidence for tissue-specific variation in DNA methylation in Arabidopsis. We examined the similarity of DNA methylation in these two tissues by making comparisons between the likelihoods of methylation at loci identified as methylated in root tissue of the Col/Col graft from our dataset and in the 3-wk-old Col-0 (WT) leaf tissue of Stroud and colleagues (31). DNA methylation was broadly consistent between the two tissues (SI Appendix, Fig. S5 and Dataset S2). Additionally, a prior study demonstrated that DNA methylomes from two tissues in each of 11 Arabidopsis accessions cluster by accession, not tissue type (38).

DNA Methylation Under Indirect Regulation by the Mobile Signal. Class B (indirect) loci show the same pattern of DNA methylation between grafts as do class A (direct) loci. However, they

Table 1. Numbers of class A, B, and C loci in each DNA methylation context

<table>
<thead>
<tr>
<th>Class</th>
<th>mC context</th>
<th>mC loci (median length, nt)</th>
<th>sRNA loci (median length, nt)</th>
<th>mC overlaps to sRNA</th>
<th>sRNA overlaps to mC</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>CG</td>
<td>34 (33)</td>
<td>925 (147)</td>
<td>13*</td>
<td>8</td>
<td>154.69</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CHG</td>
<td>2,636 (94)</td>
<td></td>
<td>398*</td>
<td>257</td>
<td>77.173</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CHH</td>
<td>2,980 (114)</td>
<td></td>
<td>401*</td>
<td>278</td>
<td>84.250</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>CG</td>
<td>34 (33)</td>
<td>26,981 (1,759)</td>
<td>13*</td>
<td>12</td>
<td>−11.68</td>
<td>−5.5 × 10−32</td>
</tr>
<tr>
<td></td>
<td>CHG</td>
<td>2,636 (94)</td>
<td></td>
<td>868*</td>
<td>651</td>
<td>−92.603</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CHH</td>
<td>2,980 (114)</td>
<td></td>
<td>1,029*</td>
<td>787</td>
<td>−105.48</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>CG</td>
<td>5 (1)</td>
<td>63 (153)</td>
<td>0*</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>CHG</td>
<td>21 (48)</td>
<td></td>
<td>2*</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>CHH</td>
<td>125 (72)</td>
<td></td>
<td>4*</td>
<td>3</td>
<td>258.68</td>
<td>0</td>
</tr>
</tbody>
</table>

The columns “mC loci” and “sRNA loci” give the total numbers of mC and sRNA loci identified corresponding to each locus class A–C, as described in Fig. 1. The mC loci were counted individually by mC sequence context (CG, CHG, CHH). sRNA loci are not context-dependent, so only one set exists for each class A–C. The sRNA loci overlapping with the mC loci from the three sequence contexts within each class A–C. Comparisons were made in two directions: the number of mC loci overlapping sRNA loci (column “mC overlaps to sRNA”) and the number of sRNA loci overlapping mC loci (column “sRNA overlaps to mC”). Reversing the direction of comparison yielded different numbers of overlaps between sRNA and mC loci. This is possible because the sRNA and mC loci are of different sizes. Hence, one sRNA locus may overlap multiple mC loci.

Overlaps carried forward for downstream analyses of that class. A positive Z score indicates an enriched association between the sRNA and methylation loci, whereas a negative Z score indicates depleted association, relative to the overlap expected by chance. The final column gives the P values associated with the overlaps. Note that certain locus classes require low levels of mC or sRNAs in specific graft combinations (Fig. 1). For example, the number of overlaps column for class B gives the number of genomic loci where mC levels were high in roots of all graft combinations, except dcl234/dcl234, and where sRNA abundances were low in all graft combinations. Consequently, for this model, the sRNA loci column gives the number of loci where sRNA abundance was low in all graft combinations.
Fig. 2. DNA methylation of mobile sRNA-associated loci (class A, direct; B, indirect) is reduced most strongly by dml12 mutation. The plots show the mean proportion of DNA methylation across class A and B loci in mutants of RNA-directed DNA methylation pathway components and wild-type (WT) plants. The legend (right) indicates which color represents each mutant. Data are shown separately for non-CG (CHG and CHH) sequence contexts, which had the strongest association with mobile sRNAs. The proportion of methylation was calculated per base (between 0 and 1, unmethylated to fully methylated) for all loci within a class using published MethylC-seq data from RdDM mutants (31). Loci were then normalized to the same size, and the mean proportion of DNA methylation was calculated across them. The profiles of the mean proportion of methylation across the size-normalized loci are plotted between the dashed vertical lines (indicated by the solid black bars labeled “loci”). Mean proportions of methylation in flanking DNA, 4 kb upstream and downstream of the loci, are indicated to the left and right of the dashed vertical lines, respectively. Total numbers of loci assessed are given in parentheses.

are distinguished by the lack of associated sRNA. We found that class B and class A loci had extremely similar characteristics. Hundreds of class B loci were detected almost exclusively in non-CG contexts (13 CG, 868 CHG, 1,029 CHH; Table 1 and Dataset S1). These loci were unambiguously identified and located in the genome. We found that fewer than 0.5% of the class B loci overlapped with regions previously identified as spontaneously variable between parent and offspring plants, suggesting our results are not a consequence of this phenomenon (18).

The distribution of the class B loci across DNA methylation contexts mirrors that of the class A loci (13 CG, 398 CHG, and 401 CHH), although substantially more class B loci were identified. We focused on non-CG class B loci because so few CG loci were identified. The association between regions lacking sRNAs and possessing DNA methylation, corresponding to the class B loci model, was lower than would be expected by chance (Table 1; CHG, \( P = 0, Z = -92.603 \); CHH, \( P = 0, Z = -105.48 \)). This pattern suggests that methylation loci regulated by a mobile signal are negatively associated with regions lacking in small RNAs.

The class A and B loci were also influenced similarly by mutations in RdDM pathway genes (31) (Fig. 2 and SI Appendix, Fig. S4). Mutation of dml12 caused the strongest reduction in non-CG context DNA methylation (Fig. 2 and SI Appendix, Fig. S4). All other mutations examined reduced non-CG context methylation at class B loci except cmt2. Notably, cmt3 had the next-weakest effect, reducing CHG context DNA methylation but not affecting CHH context methylation.

In summary, there was very little to distinguish class A and class B loci, suggesting they are ultimately under the regulation of the same mobile signal. In principle, this signal could be an as-yet unidentified secondary regulatory factor dependent on the DCL234-dependent sRNA at the class A loci. Alternatively, it could be caused by the very low levels of DCL234-dependent sRNA that fail to pass the significance thresholds in our model (SI Appendix, Figs. S6–S8), or that are not well-sequenced.

We used an independent statistical approach to estimate the total number of class A and B loci in the dataset, as a complement to the results above (see SI Appendix, Materials and Methods, section 5 for a detailed explanation). This approach evaluates all loci that may exist without specifying genomic coordinates. We estimate that there exist 72 CG (59–85 with 95% confidence), 1,557 CHG (1,506–1,608 with 95% confidence), and 2,238 CHH (2,170–2,303 with 95% confidence) loci of class A, and 526 CG (485–568 with 95% confidence), 5,699 CHG (5,597–5,802 with 95% confidence), and 10,943 CHH (10,785–11,099 with 95% confidence) loci of class B by this approach. From these data, we conclude that thousands of DNA methylation loci found predominantly in the non-CG context may be regulated by the mobile signal.

DNA Methylation Regulated by the Mobile Signal Associates with Specific Features of the Genome. Non-CG context class A loci were significantly associated with TEs and promoters containing TEs but not with promoters that do not contain TEs (Fig. 3A). In contrast, they were significantly depleted in genes, coding regions, and 5′ UTRs. CG context class A loci were significantly associated with TEs only. Further analysis showed that class A loci associated significantly with many superfamilies of TEs (Fig. 3B). RAth elements and LINE and SINE superfamily TEs were targeted more strongly than other TEs, most clearly so by the non-CG context class A loci. TEs are broadly grouped according to their replication strategies as type 1 (retroelements) or type 2 (DNA elements), and the targeted superfamilies are all short, gene-proximal type 1 TEs.

Class A loci were targeted most highly by 23- to 24-nt mobile sRNAs, eliminated in dcl234/dcl234 grafts (Fig. 3C). These results indicate that DNA methylation directly targeted by the shoot-root mobile sRNA is associated with specific TE superfamilies.

Non-CG context class B loci were also associated significantly with TEs and promoters containing TEs (Fig. 4A and SI Appendix, Fig. S2), as were non-CG class A loci. They were associated with many different TE superfamilies, most strongly with certain type 1 retroelement superfamilies (RAths, LINE, SINE). They also showed clear associations with specific retroelement superfamilies (type 2 TEs) including DNA Mariner, Pogo, and HAT–

The genome features associated with class A and B loci might either be specific to these locus classes and to the mobile signal or they could be driven by dependence of the loci upon DCL234-deriv derived sRNAs. To address this point, we compared the features associated with class A and B loci with all features that lose DNA methylation in dcl234/dcl234 grafts. The features that lost DNA methylation in dcl234/dcl234 grafts were highly similar to those associated with class A and B loci (SI Appendix, Fig. S9). The data indicate that a specific subset of TEs lose DNA methylation in the dcl234 mutant. Furthermore, they suggest that dependency upon DCL234 drives the features associated with class A and B loci rather than these features being specific to the mobile signal.

The Influence of the Mobile Signal on Gene Expression. We examined whether DNA methylation regulated by the mobile signal may influence gene expression. Root transcriptomes of the graft combinations described by Fig. 1 and SI Appendix, Fig. S1 (excluding C24/C24) were profiled by RNA-seq. A total of 23 transcripts were identified as significantly differentially regulated in a manner that indicated association with the mobile signal (Dataset S3). These included transcripts coding for proteins involved in cell-wall formation, membrane transport, and regulation of osmotic potential. Three of the genes encoding these were associated with DNA methylation regulated either directly or indirectly by the mobile signal: AT2G01880, PURPLE ACID PHOSPHATASE 7; AT2G36490, REPRESSOR OF SILENCING 1 (ROSI); and AT3G43270, pectin methyltransferase. We measured...
Fig. 3. Mobile sRNA directly targets DNA methylation at promoters containing transposable elements. (A) CHG and CHH (non-CG) context class A loci are enriched in TEs and in promoters containing TEs but are depleted in genes and coding sequences (CDSs). Very few CG context loci were identified. (B) Non-CG context class A loci preferentially target classes of type 1 retroelement, indicated by significant enrichment. They also, less strongly, target some type 2 DNA element classes. Y-axis units in A and B normalize the feature/locus overlap by both sum of genome feature size and sum of methylation locus size, which permits comparison between columns. These units are the number of annotated features per total megabase (MB) of annotated feature per total MB of methylation in the locus class. Numbers of loci assessed are given in parentheses. Significance levels: #, $P < 10^{-5}$; +, $10^{-4} < P < 10^{-5}$; *, $10^{-3} < P < 10^{-4}$; !, $10^{-2} < P < 10^{-3}$; blue, underrepresented; red, overrepresented (relative to background). (C) Class A loci are targeted by graft-transmissible 23- to 24-nt sRNAs (the known mobile class). Loci were normalized to the same size on the graph, indicated by the dark bar on the position axis, flanked by sRNA coverage 4 kb upstream and downstream of the locus. Numbers of loci assessed are given in parentheses.
the number of differentially expressed genes between C24/Col and dcl234/col234 roots as a control. Twenty-seven transcripts were identified as significantly differentially regulated in roots of dcl234 compared with WT roots (Dataset S4). Differential expression of ROS1 (AT2G36490.1) was confirmed in roots of dcl234/col234 versus C24/col234 and C24/Col by quantitative real-time PCR (Q-RT-PCR) (SI Appendix, Fig. S10). These findings demonstrate that loss of DCL2, DCL3, and DCL4 causes differential expression of a small number of genes, consistent with data in previous reports (39).

We investigated whether transcripts dependent upon DCL234 (Dataset S4) or associated with the mobile signal (Dataset S3) might also depend upon DRM1/2. This was done by identifying transcripts differentially regulated between WT and the drm1/2 mutant using published RNA-seq data. These data were generated from leaves of 3-wk-old plants in parallel with the MethylC-seq data discussed in Fig. 3 and SI Appendix, Fig. S4 (31). Eighty-six transcripts were significantly differentially regulated between WT and drm1/2 plants (Dataset S5). The transcript of ROS1 (AT2G36490) was shared between these and transcripts associated with the mobile signal (Dataset S3B). The transcript AT1G53480.1 was dependent upon DRMI/2 (Dataset S5) and DCL234 (Dataset S4). We attribute the low degree of overlap between DRM1/2-dependent and mobile signal-associated transcripts to tissue specificity of gene expression. Our sRNA-seq and MethylC-seq experiments were conducted on roots of 3-wk-old plants. As discussed above, comparison between DNA methylation from different tissues is appropriate because there is little evidence for tissue-specific variation in DNA methylation. However, transcript abundance differs more between tissues. This was demonstrated by comparison of RNA-seq data from two tissues of multiple Arabidopsis accessions, where transcript abundance correlated more highly between like tissues of different accessions than different tissues of one accession (38).

Epialleles Are Transmitted Across Accessions by Mobile sRNAs. There is natural variation in RdDM and its associated sRNAs in Arabidopsis, and the term “epiallele” describes genomic locations where methylation differs between accessions (38, 40, 41). We investigated whether sRNAs produced uniquely by C24, and not by Col, could direct DNA methylation de novo at previously

Figure 4. Loci where DNA methylation is indirectly targeted by mobile sRNAs (class B) exhibit similar characteristics to loci that are directly targeted (class A). (A) Non-CG (CHG and CHH) context B loci are significantly enriched in TEs and in promoters containing TEs but are depleted in promoters that do not contain TEs, genes, and CDSs. Very few CG context loci were identified. (B) Non-CG context B loci target most classes of TEs but show most significant association with classes of type 1 retroelement. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes. They also, less strongly, target some type 2 DNA element classes.
unmethylated sites in the Col genome (Fig. 1, class C loci). Two
class C loci were unambiguously identified in the CHG DNA
methylation context, and four in the CHH context (Dataset S1).
Three distinct sites of de novo RdDM existed after taking into
account overlap of these loci between DNA methylation contexts
(Fig. 5). None of these loci had been previously identified as
spontaneously variable between parent and offspring plants (18).
These results indicate that exogenous sRNAs supplied by shoots
can target de novo DNA methylation at unmethylated sites in the
genome of root cells, thereby transmitting epiallelic states from
one Arabidopsis accession to another.

Discussion
In this study, we have characterized the genome-wide distribu-
tion of 23- to 24-nt shoot/root mobile sRNAs and identified
regions of DNA methylation that they target directly (class A
loci). We also identified regions of DNA methylation dependent
indirectly upon mobile sRNAs (class B loci). We found that loss
of mobile sRNAs generated in the shoots disrupts DNA meth-
ylation at thousands of sites in the roots. The class B (indirect)
loci were most numerous, but the two classes were otherwise
indistinguishable. Both were almost entirely in the non-CG
(CHG/CHH) DNA methylation context. The class A and B loci
were also significantly associated with the same TE superfami-
lies. We found that these TE superfamilies also lost DNA
methylation in the dcl234 mutant, which is deficient in 22- to 24-
nt sRNAs, indicating that the feature association of class A and
B loci was driven by the dependency of mobile sRNAs on
DCL234. Our results suggest that the function of mobile sRNAs
is to reinforce silencing of these TEs. Furthermore, we have
provided mechanistic insights into the specific RdDM pathways
regulating mobile sRNA-dependent methylation. Data can be
visualized on our interactive genome browser (neomorph.salk.
edu/mobile_methylome.php).

Our data confirm that mobile sRNA-dependent DNA meth-
ylation requires DRM1 and DRM2, key components of the 24-nt
RdDM pathway. There was only limited dependence upon
CMT3 and CMT2, both of which belong to a distinct DNA
methylation maintenance pathway. Furthermore, class A and B
loci were associated with the shortest superfamilies of TEs (Figs. 2–4 and 4 and SI Appendix, Table S1), including type 1 retroelements,
such as RATH elements (E1, E2, E3), SINEs, and LINEs, as well as
certain type 2 DNA elements (Mariner, Pogo, RC Helitron). The
DNA methylation pathways containing DRM1/2 and CMT2/3 are
distinct but overlapping, and both deposit non-CG methylation (8,
34). The DRMI/2 pathway methylates small TEs and TE edges,
whereas the CMT2/3 pathway is responsible for methylation of
long TEs (31, 34). Our data are in concordance with this pattern,
and establish a relationship of mobile sRNA-dependent methy-
lization with a specific RNA silencing mechanism.

The direct and indirect (class A and B) loci are essentially
identical according to our data, except that no mobile sRNAs
were identified as being associated with indirect class B loci.
These findings lead to three hypotheses for further investigation.
The first is that the mobile sRNAs from the shoot regulate an
unidentified secondary signal in the roots. This secondary signal
would then directly regulate DNA methylation at the B loci.
The second is that mobile sRNAs can direct the RNA silencing
machinery to sites that they target imperfectly (i.e., with which they
have mismatches in sequence homology). We required perfect
mismatching between sRNAs and genomic sequence when identifying
sRNA target loci, but it is possible that a certain degree of
mismatch between mobile sRNAs and their targets is either tolerated
or required in certain circumstances. However, allowing mis-
matches during sRNA mapping (and therefore in targeting)
permits individual sRNAs to target multiple sites due to their
inherently short sequence, many of which are presumably spurious
(42). Therefore, detailed analysis of individual methylation regions
and thorough experimental validation are required to test this
hypothesis. Moreover, it appears unlikely that the conclusions of
our study would be altered by the outcome of such experiments,
because we find the characteristics of class A (direct) and class B
(indirect) loci to be essentially identical. The third hypothesis is
that these loci are targeted by mobile sRNAs of very low abundance,
which could not be detected in our sRNA-seq libraries.

Class A and B loci were substantially more numerous than the
other classes (Fig. 1, Table 1, and SI Appendix, Figs. S1 and Table
S2). The class D–F loci described loci not methylated by mobile

![Image](image_url)

**Fig. 5.** C24-derived sRNAs can target DNA methylation de novo at unmethylated regions of the Col-0 genome. Three such de novo loci are shown in AnnoJ
genome browser screenshots. The screenshots display sRNA reads and methylated cytosine residues (mCs) for graft combinations C24/C24, Col/Col, and C24/
Col. Results from two independent biological replicates, suffixed 1 and 2, are shown for each graft combination. Note the groups of sRNA reads present only
in C24/Col and C24/C24, which correlate with the presence of mC. Red and green sRNAs indicate that they map to the Watson (W) and Crick (C) strands,
respectively. Gold, blue, and pink mC positions indicate methylation in the CG, CHG, and CHH contexts, respectively.
sRNAs, not remethylated by mobile sRNAs, and whose methylation was independent of DCL234-derived sRNAs, respectively. Class F loci were the least numerous. This observation suggests there may be few mobile sRNA-targeted loci at which that sRNA does not regulate DNA methylation.

sRNAs of 24-nt length were proportionally the most common sRNA species detected in our sRNA-seq experiments and demonstrated the clearest shoot–root transmission (Fig. 2) (36). This study and another indicate that 21- to 22-nt sRNAs are also mobile and associated with DNA methylation (37). These data are consistent with previous proposals of a role for the 21- to 22-nt size class in RdDM (6, 7, 37, 43). Furthermore, the type 1 retroelements that we find to be the predominant targets of mobile sRNA-dependent DNA methylation produce RNA intermediate replication stages, making it possible that if transcriptionally active they are also targets of PTGS via mobile 21-nt sRNAs (44). However, fewer 21- to 22-nt sRNAs were mobile than 24-nt sRNAs, suggesting 24-nt sRNAs are the primary shoot–root signal. The design of our experiment did not permit us to completely eliminate the possibility that “mobile” 21-nt sRNAs are produced in the roots. This is because the dcl234 mutant contains functional DCL1, which generates 21-nt sRNAs, and the dcl234 mutant eliminates their production only partially (25, 45–47). Detailed investigation of their contribution would involve generation of an additional series of Arabidopsis mutant grafts and is complicated by the lethality of a dcl1 null plant (48).

Although mobile sRNAs have minimal effects on gene expression in Arabidopsis, we predict that their influence would be much greater in TE-rich genomes. In such genomes, including those of many common crops, mobile sRNAs may be an important mechanism of genome defense. The TE superfamilies targeted by mobile sRNAs are enriched in euchromatic regions, typically within c. 500 nt of the nearest gene (49, 50). TEs inserted near genes can have dramatic effects on gene expression (15, 50). Expression of a small number of transcripts was differentially regulated in our study, consistent with previous observations (39). However, the A. thaliana genome contains substantially fewer TEs than related outcrossing species, such as Arabidopsis lyrata (51, 52). Furthermore, TEs in the A. thaliana genome show a lower rate of active transposition (51). This suggests that this genome may lack the highly repetitive elements characteristic of other genomes with less influence on gene expression than TEs in outcrossing species, which would result in fewer mobile sRNA-dependent regulated genes. Nonetheless, those genes that did exhibit mobile sRNA-dependent regulation had diverse functions, indicating they may have significant influence in the correct conditions. Moreover, mobile sRNAs are able to move into both meristematic and meiotically active tissues, where they can alter DNA methylation in these tissues it is essential to protect genome stability by repression of TEs, so that gametes and developing organs are not harmed (54, 55).

Grafting is routinely used in agriculture to combine rootstocks and shoots (scions) with desirable characteristics, such as for grapevine, apples, and tomatoes (36, 57). We have demonstrated that mobile sRNAs regulate patterns of DNA methylation genome-wide, and that expression of specific genes exhibits mobile regulation. Base-resolution methymethylomes are being actively generated in multiple species, including crops and nonmodel plants (38, 58–60). Clear diversity in the epigenomes of closely related subspecies and accessions has been observed from these data (38, 58). With our demonstration that site-specific transmission of epiallelic states from one accession to another can be achieved by grafting and by de novo methylation of unmethylated DNA, it is likely that at least some effects of grafting are due to the movement of RNA. Our findings also indicate that DNA methymethylomes may provide a potential new resource to growers who use grafting. They could consider potential modification of gene expression patterns in sink tissues via sRNA transmission from source tissues. The mobile sRNAs can alter DNA methylation in germ-line tissue (37), so that DNA methylation patterns altered by grafting may be heritable by progeny plants (37, 61).

In summary, we have shown that transmission of sRNAs from shoots to roots of Arabidopsis regulates DNA methylation at thousands of sites genome-wide. Mobile sRNA-dependent methylation is predominantly in the non-CG context (CHH and CCHH), and is associated with short type 1 retroelements found in gene-rich regions of the genome. We confirm that deposition of mobile sRNA-dependent methylation is dependent upon the DRM1 and DRM2 RdDM pathway and largely independent of the CMT2 and CMT3 methylation pathway. Our conclusions underpin future research into why plants possess a system for communicating methylation status from shoot to root tissues.

Materials and Methods

See also SI Appendix, Materials and Methods.

Plant Materials and Grafting. Plants of A. thaliana accessions Col-0 and C24 were used in our experiments, as well as a previously described dcl2-1/rdDM-1 (36) and a rdDM2 triple mutant (Col-0 background) (23). The C24 plants had a GFP transgene silenced by a partial GFP inverted repeat, termed GxGxIR, as previously described (36). Arabidopsis plants were grown under 10-h supplemental fluorescent lighting at 20 °C on vertical plates of 0.8% agar, 1/2 Murashige and Skoog media (pH 5.7). Micrografting was conducted 7 d after germination as previously described (57, 62). Plant tissue was harvested 5 wk after germination. Tissues were taken care-shoot (0.1 cm) and 0.3–0.5 cm surrounding the graft junction. Two independent bioreplicates were conducted for all RNA-seq samples, and two to four for each MethylC-seq sample (SI Appendix, Table S3).

MethylC-Seq and RNA-Seq Library Construction. DNA for the MethylC-seq libraries was prepared from three or four pooled plants per sample using the Gentra Puregene Tissue Kit (Qiagen) according to the manufacturer’s instructions. MethylC-seq libraries were prepared from total RNA extracted as previously described (36). RNA-seq libraries were generated with the ScriptSeq RNA-Seq Library Preparation Kit (Epiconcept) according to the manufacturer’s instructions.

Sequencing Operations. MethylC-seq and RNA-seq libraries were sequenced using Illumina HiSeq 2000 and 2500 platforms, respectively, according to the manufacturer’s instructions. MethylC-seq libraries were sequenced as described previously (63). RNA-seq libraries were prepared from total RNA extracted as previously described (36). RNA-seq libraries were generated with the ScriptSeq RNA-Seq Library Preparation Kit (Epiconcept) according to the manufacturer’s instructions.

Sequencing Analyses. See SI Appendix, Materials and Methods for a detailed description of analytical methods and software. In brief, sRNA data were aligned to the TAIR9 reference genome requiring perfect matching. When Q25 is being used MethylC-seq libraries were aligned to make use of the Bowtie 2 alignment (26) and Cufflinks (84), or Cuffdiff (85) and Aligner, based upon Bovtiew 2, to map reads to the genome (source code is Q26 available at https://github.com/htbg/RTYAMA). Mapping statistics of libraries are given in SI Appendix, Table S3. MethylC loci were then identified independently in each context (CG, CHG, CHH) using the segmentSeq R package (64), accounting for nonconversion rates in each sample (65). Subsequently, loci exhibiting various models of differential methylation were identified using the empirical Bayesian methods for analysis of methylation data (64) with the baySeq R package (66, 67). Average methylation profiles were calculated by dividing defined regions into an equal number of windows and calculating methylation across these windows. Gene annotations were retrieved from the Arabidopsis Information Portal (https://www.araport.org) (68) and transposon annotations from the Arabidopsis Information Resource (69). Estimations of the numbers of loci expected to fit a model and confidence intervals were calculated by simulating 10,000 sets of true and false positives based on the posterior likelihoods. Unambiguous lists of specific differentially regulated loci were identified by applying a false discovery rate of lower than 0.05 to the model, which in general represents a more conservative analysis. Associations between DMRs, gene features, and sRNAs were assessed using a block-bootstrap method (70). Intersections between loci and genome features were normalized for abundance of loci and features, the sizes of loci and features, and how loci and features might cluster within the genome when plotting.

Lewsey et al.

8 of 10 | www.pnas.org/cgi/doi/10.1073/pnas.1515072113
We did so by calculating the number of features overlapped by the DNA methylation5′RNA loci per megabase total feature length per megabase total locus length. This approach gave a visually accurate representation of significance of association. Independent measurements of statistical significance based upon the block-bootstrap analyses are presented beneath the bars also. Promoters were defined as 2,000 nt preceding a transcriptional start site. Data were deposited in the European Nucleotide Archive (accession no. E-ETAB-3473).

Q-RT-PCR. Transcript abundance was assessed in root of independent plants as described previously (37). Fold changes in abundance of test transcripts relative to the ACTIN 2 transcript (AT3G18780), a stable control (71), were calculated using delta delta CT methodology (72, 73). Primer sequences were 5′-GGGA-TATTGTGACAAGGCCCAG-3′ (ROSI forward), 5′-CAAACTGATCCTGGGAAAGAAA (ROSI reverse), 5′-GCCATACCAAGCTGTGCTCCTC-3′ (ACTIN 2 forward), and 5′-CCCTGTAGATTTGACAG-3′ (ACTIN 2 reverse).

ACKNOWLEDGMENTS. We thank Matthew D. Schultz and Yupeng He for assistance with analyses; Huaming Chen for assistance with web browser development; and Roberto Solano, Robert J. Schmitz, Taiji Kawakatsu, Chongyuan Luo, Ryan Lister, Ronan C. O’Malley, and Lindsay Robinson for helpful discussions. M.G.L. was funded by an EU Marie Curie FP7 International Outgoing Fellowship (252475). Work in J.R.E.’s laboratory is funded by the Gordon and Betty Moore Foundation (2034) and the National Science Foundation (MCB-1024999). J.R.E. is an investigator of the Howard Hughes Medical Institute. C.W.M. was funded by a Clare College (Cambridge, United Kingdom) Junior Research Fellowship. Work in the D.C.B. laboratory was supported by the Gatsby Charitable Foundation, EU FP7 Collaborative Project Grant AENEAS, and ERC Advanced Investigator Grant ERC-2013-AdG 304602. D.C.B. and M.C.W. are the Royal Society Edward Penley Abraham Research Professor.


AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

Q: 1_Please contact PNAS_Specialist.djs@sheridan.com if you have questions about the editorial changes, this list of queries, or the figures in your article. Please include your manuscript number in the subject line of all email correspondence; your manuscript number is 201515072.

Q: 2_Please (i) review the author affiliation and footnote symbols carefully, (ii) check the order of the author names, and (iii) check the spelling of all author names, initials, and affiliations. Please check with your coauthors about how they want their names and affiliations to appear. To confirm that the author and affiliation lines are correct, add the comment “OK” next to the author line. This is your final opportunity to correct any errors prior to publication. Misspelled names or missing initials will affect an author’s searchability. Once a manuscript publishes online, any corrections (if approved) will require publishing an erratum; there is a processing fee for approved erratum.

Q: 3_Please review and confirm your approval of the short title: Mobile sRNAs regulate genome-wide DNA methylation. If you wish to make further changes, please adhere to the 50-character limit. (NOTE: The short title is used only for the mobile app and the RSS feed.)

Q: 4_Please review the information in the author contribution footnote carefully. Please make sure that the information is correct and that the correct author initials are listed. Note that the order of author initials matches the order of the author line per journal style. You may add contributions to the list in the footnote; however, funding should not be an author’s only contribution to the work.

Q: 5_You have chosen not to pay an additional $1350 (or $1000 if your institution has a site license) for the PNAS open access option. Please confirm this is correct and note your approval in the margin.

Q: 6_Please verify that all supporting information (SI) citations are correct. Note, however, that the hyperlinks for SI citations will not work until the article is published online. In addition, SI that is not composed in the main SI PDF (appendices, datasets, movies, and “Other Supporting Information Files”) have not been changed from your originally submitted file and so are not included in this set of proofs. The proofs for any composed portion of your SI are included in this proof as subsequent pages following the last page of the main text. If you did not receive the proofs for your SI, please contact PNAS_Specialist.djs@sheridan.com.

Q: 7_Please check the order of your keywords and approve or reorder them as necessary. Note that PNAS allows up to five keywords; please do not add new keywords unless you wish to replace others.

Q: 8_Please indicate whether the data have been deposited in PDB or another publicly accessible database before your page proofs are returned. It is PNAS policy that the data be deposited BEFORE the paper can be published.

Q: 9_Please check that the data deposition footnote is complete and correct as set and provide a URL.

Q: 10_Please choose a single email address for the first corresponding author, per journal style.
Q: 11_Please note that sRNAs are defined in *Significance* and the Abstract as “small RNAs” rather than “small noncoding RNAs,” as in the main text. Please rectify the discrepancy.

Q: 12_Please advise regarding “CMT3 efficiency is intermediate between MET1 and CMT3” (i.e., “CMT2 efficiency,” or “between MET1 and CMT2”?)

Q: 13_PNAS requires gene and protein abbreviations to be defined. Please provide definitions for any gene and protein abbreviations that appear more than once in your paper (it is not necessary to do so for abbreviations that appear only once in the text). If this is not possible, please provide a modifier for each term that identifies what the term is (for example, “protein kinase _____”)

Q: 14_Please advise regarding “AGO–sRNA” (i.e., currently set as a complex between AGO and sRNA).

Q: 15_PNAS italicizes genes and alleles. All genes, alleles, and proteins will appear as indicated in your proofs. Please carefully check use of italics and capital letters throughout the proof and correct as necessary. If, by “___ gene,” you mean “the gene that encodes protein ___,” then italic type is not necessary. (NOTE: If all instances of a gene/allele should be changed, please make only one correction and indicate that it is a global change.)

Q: 16_Please note the spelling out of “seq” at first mention.

Q: 17_Quotation marks are used at the first appearance of a word or phrase; they do not appear with each subsequent use of that word or phrase.

Q: 18_Please complete “The class A (direct) and C (de novo) loci ___ because these are…” (i.e., “were studied further”?)

Q: 19_Please note that the Z value for CHG is 77.123 in the text yet 77.173 in Table 1 itself. Please advise.

Q: 20_Please clarify “by all others of these mutations.”

Q: 21_Please note the edit “RAth elements and LINE and SINE superfamily TEs.”

Q: 22_Please define RAth, LINE, and SINE at first mention.

Q: 23_Please advise regarding “They also showed clear associations with specific retroelement superfamilies (type 2 TEs)” (i.e., DNA elements?).

Q: 24_Please advise regarding “shortest superfamilies of TEs” (i.e., “superfamilies of the shortest TEs”?).

Q: 25_Please advise regarding “TAIR9 reference genome” [i.e., please provide a web address or reference (ref. 69 is first cited below)]

Q: 26_Please provide a web address or reference for Bowtie 2, as needed.

Q: 27_Please advise regarding “presented beneath the bars” (i.e., please specify the figures, or that this is set throughout the figures).
Q: 28. Please note the spelling out of TSS.
Q: 29. Please spell out ERC and EU.
Q: 30. Please note the resetting of ref. 64.
Q: 31. Please note the resetting of ref. 65.
Q: 32. Please advise regarding the significance level definition indicated by an asterisk in the legend for Fig. 3, as there is no asterisk in the figure itself.
Q: 33. Please advise regarding the significance level definition indicated by a plus sign in the legend for Fig. 4, as there is no plus sign in the figure itself.
Q: 34. Please provide a web address for “AnnoJ genome browser.”
Q: 35. Please note the abbreviation “mC” indicating both “methylated cytosine residue” in the legend for Fig. 5 and “DNA methylation” elsewhere in the paper.
Q: 36. Please define “NA” in the legend for Table 1.
Q: 37. Please advise regarding “the number of overlaps column” in the footnote for Table 1 (i.e., please specify which “overlaps” column, or clarify that both are meant).