Maternal small RNAs mediate spatial-temporal regulation of gene expression, imprinting, and seed development in *Arabidopsis*

Ryan C. Kirkbride[a,b,c,1], Jie Lu[a,b,c,1], Changqing Zhang[a,b,c,1,2], Rebecca A. Mosher[d], David C. Baulcombe[e,3], and Z. Jeffrey Chen[a,b,c,3]

[a]Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX 78712; [b]Department of Integrative Biology, The University of Texas at Austin, Austin, TX 78712; [c]Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Austin, TX 78712; [d]School of Plant Sciences, University of Arizona, Tucson, AZ 85721; and [e]Department of Plant Sciences, University of Cambridge, CB2 3EA Cambridge, United Kingdom

Arabidopsis seed development involves maternal small interfering RNAs (siRNAs) that induce RNA-directed DNA methylation (RdDM) through the *NRPD1*-mediated pathway. To investigate their biological functions, we characterized siRNAs in the endosperm and seed coat that were separated by laser-capture microdissection (LCM) in reciprocal genetic crosses with an *nrpd1* mutant. We also monitored the spatial-temporal activity of the *NRPD1*-mediated pathway on seed development using the AGO4:GFP::AGO4 (promoter:GFP::protein) reporter and promoter:GUS sensors of siRNA-mediated silencing. From these approaches, we identified four distinct groups of siRNA loci dependent on or independent of the maternal *NRPD1* allele in the endosperm or seed coat. A group of maternally expressed *NRPD1*-siRNA loci targets endosperm-preferred genes, including those encoding AGAMOUS-LIKE (AGL) transcription factors. Using translational promoter:AGL::GUS constructs as sensors, we demonstrate that spatial and temporal expression patterns of these genes in the endosperm are regulated by the *NRPD1*-mediated pathway irrespective of complete silencing (*AGL91*) or incomplete silencing (*AGL40*) of these target genes. Moreover, altered expression of these siRNA-targeted genes affects seed size. We propose that the corresponding maternal siRNAs could account for parent-of-origin effects on the endosperm in interploidy and hybrid crosses. These analyses reconcile previous studies on siRNAs and imprinted gene expression during seed development.

DNA methylation | imprinting | epigenetics | polyploidy | heterosis

Seeds are important to plant evolution and global food supply. A typical seed consists of an embryo (the next generation of the plant), the endosperm (a triploid nutritive support tissue with a 2:1 maternal-to-paternal genome ratio), and the seed coat (consisting entirely of maternal tissue). Endosperm can compose approximately 80% of the seed mass in cereals like wheat or maize and provides the majority of calories consumed by humans (1). Biologically, the endosperm, like the placenta in mammals, provides nutrients and signals for embryonic development (2, 3).

Seed development in *Arabidopsis* is associated with small-interfering RNAs (siRNAs) (4, 5), whose biogenesis depends on *Nuclear RNA Polymerase D1* (*NRPD1*), encoding a large subunit of RNA polymerase IV (PolIV) (6, 7). A large number of siRNAs (*NRPD1*-siRNAs or PolIV-siRNAs) are known to be maternally expressed in seeds (5), and their abundance correlates positively with the maternal genome dosage and negatively with expression levels of siRNA target genes in interploidy crosses (4). A recent study also suggested an inverse relationship between siRNA and target mRNA levels in the endosperm of *Arabidopsis thaliana* and *Arabidopsis lyrata* seeds (8). The regulation of gene expression in seeds by these maternal siRNAs is likely to invoke RNA-directed DNA methylation (RdDM) (9) involving NRPD1, ARGONAUTE (AGO) family members, and other factors (10, 11). Furthermore, paternal *NRPD1*-siRNAs (also known as easiRNAs) are involved in the hybridization barriers in response to increased paternal ploidy in *Arabidopsis* (12).

However, from previous studies, it is not clear whether the maternal expression of these *NRPD1*-siRNAs is from the maternal seed coat in the seed or from uniparentally expressed alleles in the endosperm (4, 5, 8). There was ambiguity in those and other studies because they typically used manually dissected seed tissues (13), in which there is the potential for cross-contamination of endosperm and seed coat tissues.

In this study, we investigated the expression of *NRPD1*-siRNA loci and RdDM pathway genes during seed development. Using laser-capture microdissection (LCM) to isolate endosperm and seed coat from reciprocal crosses between the wild-type (Wt; Colubmia-0 or Col-0) and *nrpd1* plants, we identified four distinct groups of *NRPD1*-siRNA loci, which we used to identify corresponding target genes subject to RdDM silencing. Using a sensor of *GUS* (encoding β-glucuronidase) fused with the genes targeted by *NRPD1*-siRNAs, including *AGL91* and *AGL40*, we generated several sets of stable transgenic plants and examined

Significance

Previous studies have revealed that RNA silencing and epigenetic regulation are important for seed development. Here we used laser-capture microdissection to analyze small RNA in the endosperm and seed coat and investigate their biological effects. We have identified four distinct groups of siRNA loci, including those expressed specifically from the maternal genome in the developing seed. One group of these siRNAs targets the endosperm-preferred genes, including those encoding AGAMOUS-LIKE (AGL) transcription factors. We show that the spatial-temporal expression of *AGL40* and *AGL91* is regulated by these maternal siRNAs, and disruption or overexpression of these genes in the endosperm alters seed size. Selection or manipulation of this siRNA-mediated gene expression may help improve seed production and crop yields.

PLANT BIOLOGY

spatial-temporal expression patterns in the seeds of Wt and the *nrpd1* mutant and their inheritance in the reciprocal crosses involving the mutant. We also investigated the effects of mutation or overexpression of *AGL91* or *AGL40* on seed development and size. Collectively, our data suggest that maternally expressed *NRPD1*-siRNAs can serve as suppressors of gene expression in developing seeds and also provide unique evidence that *NRPD1*-siRNAs mediate spatial-temporal expression of the endosperm-expressed genes, which regulate seed size.

Results

Biogenesis of siRNAs in the Seed Coat and Endosperm. To determine where in the seed *NRPD1*-siRNAs are produced, we used LCM (14) to separate the endosperm from the seed coat of seeds at 6 d after pollination (DAP). The plants were derived from reciprocal crosses using the *nrpd1* (*n*) mutant as the maternal (*n*XWt) or paternal (WtX*n*) parent (by convention, the maternal parent is listed first in a genetic cross) (*SI Appendix*, Fig. S1). All crosses used the Col-0 ecotype. Small RNA libraries were created from each subregion and sequenced. All 24-nt small RNA reads that were perfectly mapped onto the TAIR10 genome were assigned to 100-bp bins and normalized using scaling factors based on counts of the 21-nt known microRNA (miRNA) reads, as the miRNA abundance should be relatively unaffected in the crosses with the *nrpd1* mutant (*SI Appendix, SI Methods*). The data from two biological replicates were correlated ($r = 0.62$–0.87) (*SI Appendix*, Fig. S2), but the biological replicates were slightly separated in a principal component analysis (PCA) (*SI Appendix*, Fig. S3), which could result from the different tissue fixation methods used between the replicate sets. We circumvented this problem by selecting only the loci present in both biological replicates for further analysis. PCA revealed a clear distinction of siRNAs between the endosperm and seed coat samples in the WtX*n* cross (*SI Appendix*, Fig. S3). In the *n*XWt cross, the endosperm and seed coat samples were positioned more closely, resulting from a shared loss of large amounts of 24-nt *NRPD1*-siRNAs (*SI Appendix*, Fig. S4*A*).

Using a minimum expression-level filter (*SI Appendix*, Fig. S5), we identified 18,721 and 29,711 siRNA sites in the endosperm and seed coat, respectively (Dataset S1). Most sites in the seed coat were absent in the *n*XWt cross (Fig. 1*A*), suggesting that *NRPD1* is required for biogenesis of these siRNAs in the maternal tissue. The majority of siRNA sites in the endosperm, except for pericentromeric loci, were also lost in the *n*XWt cross despite the presence of a functional *NRPD1* paternal allele (5). Notably, 14,389 (~77%) of 18,721 siRNA sites in the endosperm overlapped with those (29,711) in the seed coat, and nearly all loci were lost in the *n*X*n* cross.

We ruled out the possibility that these data were distorted by contamination of the seed coat materials in the endosperm using quantitative RT-PCR (qRT-PCR) assays for expression of two endosperm-specific markers and two seed coat-specific markers (14) (*SI Appendix*, Fig. S6). The absent or low level of contamination confirms that many of the *NRPD1*-dependent siRNAs in the seed are expressed specifically from the maternal genome (4, 5). Some of these loci are expressed in the seed coat, and the siRNAs may have moved into the endosperm or may have been produced independently in the seed coat and the endosperm.

Four Major Groups of siRNA Loci in the Seed Coat and Endosperm. To better understand the biological function of siRNAs in the endosperm and seed coat, we further classified all siRNA loci relative to their genetic features and expression patterns. We used normalized 100-bp bin read counts to differentiate siRNA loci expressed in the endosperm (E, expressed; e, poorly or not expressed) or in the seed coat (C, expressed; c, poorly or not expressed) and based on their parent-of-origin expression in the WtX*n* cross (using the Wt *NRPD1* as the maternal parent) or *n*XWt cross (using the *nrpd1* mutant as the maternal parent) (Fig. 1*B*). Our nomenclature system has four letters, in which the first two represent endosperm (E or e) and seed coat (C or c)
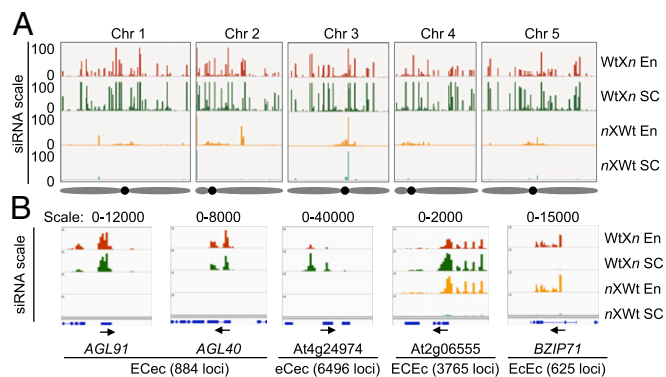

Fig. 1. Distribution of 24-nt siRNAs and classification of four distinct groups of siRNA loci in the endosperm and seed coat. (*A*) Overview of 24-nt siRNA distributions across five *A. thaliana* chromosomes (Chr 1–5) in endosperm (En) and seed coat (SC) samples from WtX*nrpd1* (WtX*n*; endosperm, red; seed coat, green) and *nrpd1*XWt (*n*xWt; endosperm, orange; seed coat, teal) seeds. The scale on the *y*-axis is based on genome browser windows displaying the 90th percentile scores for the normalized 100-bp bin read counts (*SI Appendix, SI Methods*). Diagrams of chromosomes with approximate centromeric locations (black dots) are shown (*Bottom*). (*B*) Examples of four distinct 24-nt siRNA classes (from left to right) present in both the endosperm (E) and seed coat (C) and maternal-*NRPD1* dependent (ECec; *AGL91* and *AGL40*), largely in the seed coat, and maternal-*NRPD1* dependent (eCec; *At4g24974*), in both endosperm and seed coat. Only the seed coat fraction is dependent on maternal *NRPD1* (ECEc; *At2g06555*), and only in the endosperm and independent of maternal *NRPD1* (EcEc; *BZIP71*). See the text for a detailed description. The range of siRNA is shown above each diagram, corresponding to the normalized 100-bp bin read counts, and the total number of siRNA loci in each group is shown in parentheses.

expression depending on the maternal *NRPD1*, and the second two indicate dependency on the paternal *NRPD1* (although siRNAs in the seed coat are exclusively maternally derived). For example, an ECec locus has dependency on maternal, but not paternal, *NRPD1* in the endosperm and seed coat.

Our further analysis was then based on empirically selected siRNA loci with an siRNA density that exceeded a minimum expression level in both biological replicates (*SI Appendix*, Fig. S5). After examining all possible combinations of expression or nonexpression of siRNA within 100-bp windows, we found four dominant siRNA expression patterns, each with a large number of corresponding loci: ECec, eCec, ECEc, and EcEc (*SI Appendix*, Fig. S7). Each siRNA locus usually did not overlap with another locus in a different group (*SI Appendix*, Fig. S8*A*). Other expression patterns with few or no siRNA loci were excluded from further analysis (*SI Appendix*, Fig. S8*B*).

The first group of siRNA loci (884), designated ECec, was dependent on maternal *NRPD1*, and the siRNAs were present in both the endosperm and the seed coat (Fig. 1*B* and Dataset S2). The maternal dependency in the ECec loci did not result from the *NRPD1* dosage in the endosperm (two copies in WtX*n* and one copy in *n*XWt), as the median ratio of all loci in the endosperm between the WtX*n* and *n*XWt crosses was 111-fold (*SI Appendix*, Fig. S9), which greatly exceeded the expected ratio of 2:1 (Fig. 1*B*). Similarly, the seed coat siRNA loci (eCec) had a large expression bias, while nonmaternally biased loci (EcEc and ECEc) had a ratio close to 1:1 (*SI Appendix*, Fig. S9).

These ECec *NRPD1*-siRNA loci were more likely to be associated with genes than other classes (*SI Appendix*, Fig. S10*A* and Dataset S6), most of which consist of transposable elements (TEs) or TE fragments in the gene vicinity or in upstream or downstream regions (4), including those encoding AGAMOUS-LIKE (AGL) transcription factors, such as AGL91 and AGL40 (Fig. 2*A*). They also match the genes expressed in the endosperm subregions (*SI Appendix*, Fig. S8*C*) (14). Notably, TE fragments were present in the coding and flanking sequences of these loci (Fig. 3*A* and *SI Appendix*, Fig. S11*A*), and they were more likely

to overlap with hypermethylated regions of the differentially methylated regions (DMRs) (15) in the endosperm than other groups (*SI Appendix*, Fig. S10*B*) but less likely to overlap with embryo DMRs (*SI Appendix*, Fig. S10*C*).

The second group, designated eCec, was the most abundant (6,496) among the four groups and was also dependent on maternal *NRPD1*, being expressed exclusively in the seed coat, a maternally derived component (Dataset S3). This group includes a locus adjacent to At4g24974, encoding a self-incompatibility protein S1 family (Fig. 1*B*). Other eCec siRNA loci were associated with the genes involved in sugar, sucrose, carbohydrate symporter and transporter activities, and extracellular regions (Dataset S3) that could contribute to the seed coat development.

The third group, designated ECEc, was present in both endosperm and seed coat (Dataset S4). The seed coat fractions of these siRNA loci (3,765) were present in the WtX*n* cross and lost in the *n*XWt cross (Fig. 1*B*), indicating the dependency of these loci on *NRPD1*. In the endosperm, these siRNA loci were present in both the WtX*n* and *n*XWt seeds, indicating their dependence on either maternal or paternal *NRPD1*.

The fourth group, designated EcEc (625), was present only in the endosperm but in both the WtX*n* or *n*XWt seed (Fig. 1*B* and Dataset S5). These siRNA loci (Dataset S5), as with the similar loci in ECEc, are dependent on either maternal or paternal *NRPD1* alleles acting redundantly or by an *NRPD1*-independent pathway (11). However, in seeds from parents homozygous for *nrpd1* (*n*X*n*), the majority of EcEc loci (569; 91%) including At3g32280, were absent, indicating dependency on *NRPD1*. These siRNAs could act as sensors for the paternal genome dosage (12). A small subset (56; 9%) of *NRPD1*-independent EcEc loci correspond to genes including At4g13570 (*SI Appendix*, Fig. S12*A*) and two DNA methyltransferase family genes (*SI Appendix*, Fig. S12 *B* and *C*). The *NRPD1*-independent EcEc loci also included five miRNA loci. *NRPD1*-independent loci were absent from the ECec group and rare in the ECEc (0.1%) and eCec (0.5%) groups (*SI Appendix*, Fig. S12*D*). Functions of these *NRPD1*-independent loci remain to be tested.
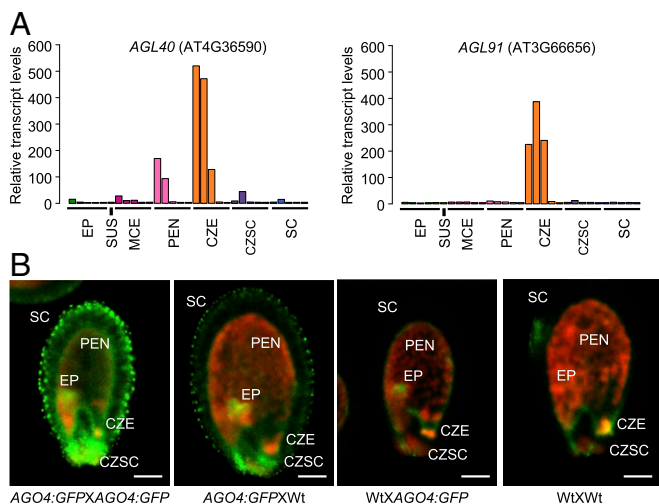

Fig. 2. Spatial-temporal expression of two *NRPD1*-siRNA target genes and *AGO4:GFP*. (*A*) Expression patterns of two *NRPD1*-siRNA target genes in the ECec group based on the LCM data reported by Belmonte et al. (14), showing *AGL91* expression in the CZE and *AGL40* expression in the CZE and peripheral endosperm. CZSC, chalazal seed coat; EP, embryo proper; MCE, micropylar endosperm; PEN, peripheral endosperm; SC, distal seed coat; SUS, suspensor. With each region (except the suspensor), five developmental stages (from left to right) are preglobular, globular, heart, linear cotyledon, and mature green. (*B*) Confocal images of seeds (at 6 DAP) exhibiting AGO4: GFP expression patterns in EP, SC, PEN, CZE, and CSC. GFP is shown in green; chlorophyll autofluorescence, in red. (Scale bars: 0.1 mm.)
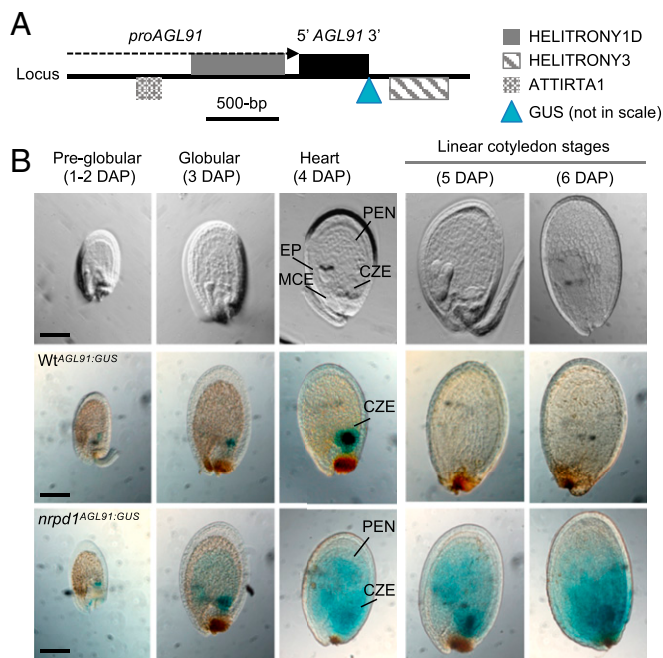

Fig. 3. Spatial-temporal regulation of *AGL91* is dependent on maternal *NRPD1*. (*A*) The genomic region of *AGL91* (AT3G66656) is surrounded by transposons in the promoter (*proAGL91*) and 3′ regions. The locations of *GUS* insertion and endogenous p4-siRNAs are shown. (*B*) Cleared (*Upper*) and GUS-stained seed images in the Wt^*AGL91:GUS*^ (*Middle*) and *nrpd1*^*AGL91:GUS*^ (*Lower*) transgenics. Wt^*AGL91:GUS*^ and *nrpd1*^*AGL91:GUS*^ were transgenic plants that expressed *AGL91:GUS* in the Wt (Col-0) and the *nrpd1* mutant, respectively. Locations of the embryo proper (EP), CZE, micropylar endosperm (MCE), and peripheral endosperm (PEN) are marked at the heart stage. (Scale bars: 0.1 mm.)

Spatial-Temporal Regulation of *NRPD1*-Mediated Pathway Genes During Seed Development. The presence of siRNAs in the seed coat and endosperm is consistent with the expression patterns of RdDM pathway genes, including *NRPD1*, *NRPE1*, *RDR2*, *DCL3*, *AGO4*, and *AGO9*, in those tissues (14). Expression levels of the RdDM pathway genes are very high in the embryo proper early in development (up to the heart stage) (*SI Appendix*, Fig. S13, box with dashed lines), and late in development (linear cotyledon and mature green stages), they increase to even higher levels in the endosperm (box with solid lines), while their expression levels in the seed coat are generally low. In contrast, non-RdDM pathway genes, such as *MET1* and *DDM1*, do not exhibit this expression pattern.

With variable spatial-temporal expression levels of the RdDM pathway genes in the seed, we sought to confirm the pathway activity using an existing reporter line (16), in which *pAGO4: GFP::AGO4* (promoter:GFP::protein) was expressed in the *ago4* mutant (Fig. 2*B* and *SI Appendix*, *SI Methods*). We observed AGO4:GFP expression in the seed coat, embryo, and peripheral endosperm at 6 DAP, when crosses were performed using transgenic plants for both parents (Fig. 2*B* and Movie S1, and with replicates in *SI Appendix*, Fig. S4*B*). This expression pattern was similar in the cross when the *AGO4:GFP* line was used as the maternal parent (*AGO4:GFP*XWt) (Movie S2); however, when the *AGO4:GFP* line was used as the paternal parent in the cross (WtX*AGO4:GFP*) (Movie S3), GFP expression was still present in the embryo but was reduced in the endosperm and absent in the seed coat relative to WtXWt control (Movie S4). These data indicate that *AGO4* and other RdDM pathway genes, including possibly *NRPD1* (5), are maternally expressed in the endosperm and seed coat but biparentally expressed in the embryo.

Spatial-Temporal Expression of siRNA Target Genes Depends on Maternal *NRPD1* or *RDR2*. In the group of *NRPD1*-siRNAs present in both the endosperm and seed coat (ECec), the siRNA-associated AGL

transcription factor members (Datasets S2–S5) were statistically overrepresented among all genes (hypergeometric distribution; $P < 0.01$). These include proteins able to regulate endosperm cellularization and seed development (17, 18), and we hypothesized that these TE-associated siRNA loci (ECec group) in the endosperm can affect spatial-temporal expression of these endosperm-expressed genes and, consequently, seed development.

To test this possible effect in the endosperm, we selected two *NRPD1*-siRNA associated genes, *AGL91* and *AGL40*, for an in-depth study. This is partly because *AGLs* affect endosperm development (17) and that *AGL91* and *AGL40* loci produce the most abundant siRNAs among the list of ECec loci (Fig. 1*B* and Dataset S6), in contrast to *AGL62*, which is associated with relatively low siRNA levels. These siRNAs depend on the maternal *NRPD1* in seeds and could induce RdDM (4). Indeed, using published data (15), we found that both *AGL40* and *AGL91* were heavily methylated in the CHG and CHH contexts in the endosperm but not in the embryo (*SI Appendix*, Fig. S14), although there was CG methylation in both the endosperm and embryo. Moreover, based on published transcriptome data of LCM samples (14), *AGL91* was highly expressed in the chalazal endosperm (CZE) from preglobular to heart stages, while *AGL40* was expressed in both the CZE and peripheral endosperm, but at higher levels in the former (Fig. 2*A*).

We next generated multiple independently derived transgenic lines using a translational *promoter:AGL::GUS* (encoding β-glucuronidase) sensor driven by either the *AGL91* or the *AGL40* promoter (including siRNA-generating regions) (Fig. 3*A* and *SI Appendix*, Fig. S11*A*) and used histochemical staining to assess GUS expression in different regions of the seed. Using the reporter lines allowed us to test the function of *NRPD1* in different genetic crosses and in a spatial-temporal manner. In the Wt transgenic plants (Wt$^{AGL91:GUS}$), GUS was undetectable in the micropylar or peripheral endosperm but was present in the CZE at preglobular (1–2 DAP), globular (3 DAP), and heart (4 DAP) stages (Fig. 3*B*, *Middle*). However, CZE expression was lost by the linear cotyledon stage (5–6 DAP), when the RdDM pathway genes were up-regulated in the endosperm (*SI Appendix*, Fig. S13) (14). In the multiple lines of the *nrpd1* mutant transgenic plants (*nrpd1*$^{AGL91:GUS}$), *AGL91:GUS* was ectopically expressed in all endosperm domains and persisted after the linear cotyledon stage (Fig. 3*B*, *Bottom*).

Notably, when the *nrpd1* or *rdr2* mutant was used as the maternal parent in the genetic cross (*nrpd1*XWt$^{AGL91:GUS}$) (Fig. 4*A*, *Left*) or *rdr2*XWt$^{AGL91:GUS}$ (Fig. 4*A*, *Right*), *AGL91* was ectopically expressed. This finding rules out the possibility that overexpression in the *nrpd1* transformants relative to Wt was due to variation between transgenic lines. It indicates that maternal *NRPD1*-siRNAs mediate RdDM, which in turn silences *AGL91* expression in other endosperm domains at 1–2 DAP and later in the CZE after the heart stage. The loss of late-stage silencing in the *nrpd1*$^{AGL91:GUS}$XWt cross implies that the RdDM in the WtXWt seed is established after fertilization, when maternal *NRPD1* is required for RdDM (Fig. 4*A*, *Left*).

Likewise, *AGL40:GUS* was expressed in the CZE endosperm from the preglobular stage to the heart stage and was absent in the linear cotyledon stage (*SI Appendix*, Fig. S11*B*, *Middle*). This late-stage silencing of the *AGL40* allele was relieved whenever the *nrpd1* or *rdr2* mutation was used as the maternal parent (*nrpd1*XWt$^{AG40:GUS}$ or *rdr2*X Wt$^{AG40:GUS}$) (*SI Appendix*, Fig. S11*C*), indicating that maternal *NRPD1*-siRNAs mediate late-stage silencing of *AGL40* or *AGL91*, as suggested previously (4, 5). The persistent expression of *AGL40:GUS* and *AGL91:GUS* at later stages (6 DAP) was further confirmed by qRT-PCR analysis of the endogenous genes *AGL 40* and *AGL91* either in the cross using *nrpd1* as the maternal parent or in the manually crossed *nrpd1* mutant (*SI Appendix*, Fig. S15).

There is a difference in the silencing of *AGL91* and *AGL40*, however. Maternal *AGL91:GUS* was completely silenced at all regions and stages of the seeds tested in the cross using the Wt as a maternal parent (Fig. 4*B*), while silencing of the paternal
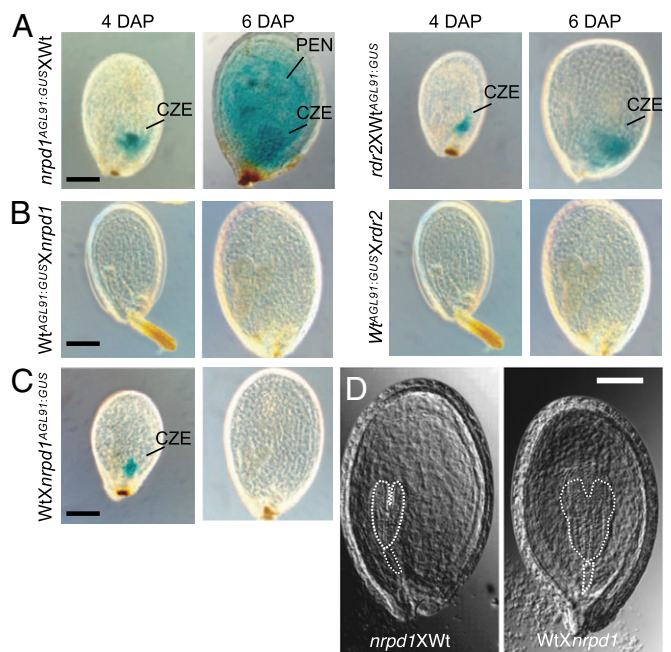

Fig. 4. *AGL91* is paternally expressed and dependent on maternal *NRPD1* or *RDR2*. (*A* and *B*) GUS-stained seed images at 4 and 6 DAP in (*Left*) reciprocal crosses of *nrpd1* $^{AGL91:GUS}$XWt (*A*) and Wt $^{AGL91:GUS}$X*nrpd1* (*B*) and (*Right*) reciprocal crosses of *rdr2*XWt $^{AGL91:GUS}$ (*A*) and Wt $^{AGL91:GUS}$X*rdr2* (*B*) (*n* = 3 × 100 seeds per cross). (*C*) GUS-stained seed images at 4 and 6 DAP in the cross of WtX *nrpd1* $^{AGL91:GUS}$. (Scale bars: 0.1 mm for all images.) (*D*) Images of cleared seeds in the crosses between Wt (Col-0) and *nrpd1* at 6 DAP; the area corresponding to the embryo and suspensor is outlined. (Scale bars: 0.1 mm.)

AGL91:GUS in the CZE was delayed until after 4 DAP in the reciprocal cross (Fig. 4*C*), as in the Wt (Fig. 3*B*). Similarly, maternal *AGL91:GUS* was completely silenced in the crosses using the *rdr2* mutant as a paternal parent (Fig. 4*B*), when no paternal *RDR2* would be present. This suggests that maternal siRNAs silence maternal rather than paternal *AGL91:GUS* more completely in all endosperm regions and stages. Consistent with this idea, in the WtX*nrpd1*$^{AGL91:GUS}$ seed, paternal *AGL91:GUS* was expressed in the CZE at 4 DAP but was silenced at 6 DAP (Fig. 4*C*).

For *AGL40*, *AGL40:GUS* expression was present in the CZE at 4 DAP in both the *nrpd1*$^{AGL91:GUS}$XWt cross (*SI Appendix*, Fig. S11*C*) and the reciprocal cross WtX*nrpd1*$^{AGL91:GUS}$ but was silenced at 6 DAP (*SI Appendix*, Fig. S11*D*). Thus, compared with *AGL91*, the maternal siRNAs silence maternal *AGL40:GUS* less completely in the CZE at early stages. This difference could be related to the different siRNA targets (TEs) present in *AGL91* and *AGL40* promoters (Fig. 3*A* and *SI Appendix*, Fig. S11*A*). Alternatively, maternal *AGL40* and *AGL91* alleles could be subjected to different modifications in the central cell through such mechanisms as demethylation by DEMETER (19). A third explanation associated with expression variation among transgenes is unlikely, because multiple lines expressing *AGL40:GUS* or *AGL91:GUS* exhibited the same differences.

Function of Endosperm-Expressed Genes (*AGLs*) in Seed Development. If *AGL91* and *AGL40* regulate endosperm cellularization (17, 18), we predict that repressing or overexpressing these genes would influence seed size. We tested this possibility in the T-DNA insertion lines for *AGL91* and *AGL40* (Fig. 5*A*), in which the corresponding transcript levels were disrupted and less abundant than in the Wt (Fig. 5*B*). Consequently, both seed size and weight were ~10% lower in the *agl91* and *agl40* mutants than in the Wt (Fig. 5 *C* and *D*). Similarly, when *AGL40* or *AGL91*
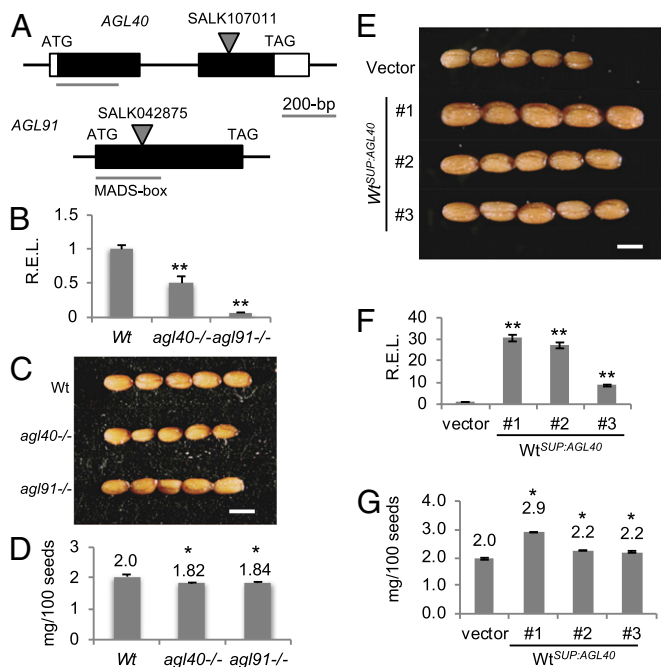
Fig. 5. Effects of maternal *NRPD1*-siRNA target genes on seed size. (*A*) Diagrams of *AGL40* and *AGL91* showing T-DNA insertion sites. (*B*) Relative expression levels (REL) of *AGL40* and *AGL91* in T-DNA insertion lines (mean ± SEM; $P = 0.01$). *$P = 0.05$; **$P = 0.01$. (*C*) Mature seed images in Wt, *agl40*, and *agl91* lines. (*D*) Seed weights in Wt, *agl40*, and *agl91* lines (mean ± SEM; $P = 0.05$). (*E*) Mature seeds in three independent Wt^SUP:GL40^ transgenic lines and one control (transformed with the vector alone) under the endosperm-specific promoter from AT5G27880. (*F*) RELs of *AGL40* in three Wt^SUP:GL40^ lines and one control (mean ± SEM; $P = 0.01$). (*G*) Seed weights in three Wt^SUP:GL40^ lines ($n = 3 \times 500$ seeds per line) and one control ($P = 0.05$). (Scale bars: 0.5 mm for all images.)

was expressed under a strong endosperm-specific promoter *SUP16* (At5g27880; *SI Appendix, SI Methods*) (20), the corresponding mRNA levels were dramatically increased relative to the vector transgenic control (Fig. 5E), and seed weight was increased by 10–45% in three independent Wt^SUP:AGL40^ transgenic lines (Fig. 5 *F* and *G*), although not in the Wt^SUP:AGL91^ lines.

These data indicate that *NRPD1*-siRNAs guide RdDM, mediating the spatial-temporal regulation of their siRNA target genes, including *AGL40* and *AGL91*, in the endosperm during seed development (Fig. 6). Activation of RdDM genes was delayed in the CZE, which consists of polyploid endosperm nuclei with a delayed developmental trajectory relative to other endosperm domains (14). The spatial-temporal role of RdDM could control gene expression in the endosperm, regulating endosperm cellularization and seed mass (4, 14, 17). Indeed, embryo development and endosperm cellularization were delayed, and seeds were larger in the cross in which *nrpd1* was used as the maternal parent compared with the reciprocal cross (Fig. 4D); *nrpd1* seeds are slightly larger but nearly normal probably because an imprinting effect is erased in the mutant (see *Discussion*).

Discussion

Our findings reveal a spatial-temporal mechanism for the *NRPD1*-dependent RdDM pathway to regulate gene expression during the endosperm and seed development (Fig. 6), and they clarify conflicting data in the field. For example, the data derived from the whole seed or manually dissected samples could confound gene expression information in the endosperm with the seed coat (4, 5, 8, 13). Similarly, the little overlap of the imprinted genes identified among several studies (13) could be related to the mixture of endosperm with the seed coat and other tissues or to the lack of spatial-temporal information. According

to a comparative study of gene expression data (13), LCM datasets are superior to other datasets generated using manually dissected tissues at different stages of seed development. However, only a few reports on small RNA datasets from seed tissues are available (8). Before the LCM approach, we also manually dissected endosperm and seed coats for small RNA sequencing as a pilot experiment. We compared the pilot data and those from published manually dissected samples (8) with our LCM dataset. Among the manually dissected tissues, the endosperm and seed coat samples were similar and grouped with the LCM seed coat sample in the WtX*n* cross (*SI Appendix*, Fig. S16A). This finding implies contamination between endosperm and seed coat tissues in the manually dissected samples. However, in the LCM samples, the endosperm was free of seed coat siRNAs in the WtX*n* cross and of maternal siRNAs in the *n*XWt cross, because *nrpd1* was used as the maternal parent. These results were further confirmed by the analysis of correlation coefficients in these samples (*SI Appendix*, Fig. S16B).

The contamination issue in the manually dissected tissues can be somewhat mitigated by filtering datasets to exclude seed coat siRNA loci from the endosperm, but this approach would miss the siRNA loci present in both the seed coat and endosperm, which represents an important group in our study. Thus, using LCM, or better yet, a single-cell analysis is the best approach for revealing spatial and temporal information on siRNA regulation during seed development.

Our LCM and sensor datasets have clarified not only the location of 24-nt siRNAs produced in the seed coat or endosperm, avoiding seed coat contamination (13, 21, 22), but also the increased resolution of ambiguous spatial-temporal effects of the RdDM pathway on imprinting in the endosperm (22). In our study, maternal alleles of *NRPD1*-siRNA–associated genes are expressed at early developmental (preglobular to heart) stages, during which the RdDM pathway is inactive (*SI Appendix*, Fig. S13). The RdDM is established after fertilization, when maternal *NRPD1* is required for RdDM. The basis for activating RdDM pathways in the late stages remains unknown, however.

This concept of spatial-temporal regulation could explain the different views of the relative contributions of maternal and paternal alleles to embryogenesis and seed development (23). The cause of this spatial-temporal regulation is largely unknown. Based on our data (Figs. 3 and 4), it is regulated by the RdDM pathway, involving *NRPD1*, *RPD2*, *AGO4*, and other factors, including *AGO9* (16). Notably, the expression of RdDM pathway genes is also spatiotemporally regulated (*SI Appendix*, Fig. S13), suggesting that additional unidentified factors or pathways are involved in the spatial-temporal regulation of gene expression in the endosperm and possibly the embryo.

In the endosperm, maternal *NRPD1*-siRNA loci (ECec group) were enriched with the genes in the Gene Ontology (GO) groups of cytosine methylation and methyltransferase activity (Dataset S6); coincidently, the *NRPD1*-independent siRNA loci (EcEc
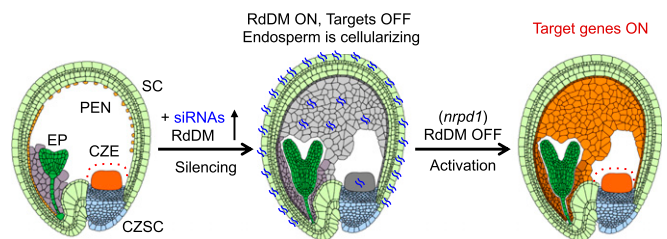

Fig. 6. A model for RdDM in seed development. A model for siRNA regulation of gene (*AGL*) expression in the endosperm. Early in development, RdDM activity is low, spatially restricting *AGL* expression to the CZE. Later in development, siRNAs and the RdDM pathway are activated to silence *AGL* expression. Loss of the RdDM pathway results in the failure to restrict *AGL* expression spatially to the CZE and temporally to early stages of development. CZSC, chalazal seed coat; EP, embryo proper; PEN, peripheral endosperm; SC, distal seed coat.

group) were also enriched with similar GO terms (*SI Appendix, Fig. S12C*). The *NRPD1*-independent siRNA loci in other groups are small or have no obvious hint of biological relevance (Dataset S6). We speculate that *NRPD1*-dependent and -independent DNA methylation pathways could exert feedback regulation to mediate spatial and temporal gene expression in the endosperm.

Although functions of many newly discovered siRNA loci in the seed coat are unknown (Dataset S6), these loci could potentially influence biological processes involved in endosperm and seed development through regulation of gene expression and TE activities in the seed coat and/or through their movement from the seed coat to the endosperm and other seed regions.

The differences in allelic expression variation between *AGL91* and *AGL40* in the endosperm is likely associated with the maternal siRNAs that act in *cis* (on the maternal allele) more effectively than in *trans* (on the paternal allele) (Fig. 4). The molecular basis for this difference is unknown. One possibility is that maternal *NRPD1*-siRNA regulation of target loci strikingly resembles the parental conflict model in mammals, in which maternally expressed factors inhibit growth and paternally expressed factors promote growth (24). Consequently, the mutant phenotype of paternally expressed *Igf2* (*Insulin-like growth factor2*) is a 40% reduction in growth (25), whereas the mutation of maternally expressed *Igf2r* (*Igf2 receptor*) leads to overgrowth and death (26). However, the double mutant is normal-sized and viable (27). Consistent with this model in mammals, PolIV may produce maternal or paternal siRNAs to silence maternal or paternal alleles in plants. When both maternal and paternal imprints are erased in the *nrpd1* mutant, seed size remains normal. The effect of reduced maternal *NRPD1*-siRNAs on the seed size increase (Fig. 4D) is reminiscent of a similar effect in the crosses when the maternal parent is demethylated (28).

Interestingly, *NRPD1*-siRNAs have evolved concurrently with the double-fertilization process in angiosperms, because no gymnosperm species produces *NRPD1*-siRNAs (29, 30). Apart from providing nutrients, the endosperm plays key roles in embryogenesis, seed development, and speciation (2, 3). The endosperm is associated with seed-size variation in *Arabidopsis* interploidy crosses (4) and seed lethality in interspecific hybrids (31). The former is correlated with the maternal siRNA dosage (4), while the latter is predicted as the inability to silence TEs because of the divergence between maternal siRNAs and paternal TEs or genes in related species (32). This prediction could be tested to elucidate the evolutionary role for maternal siRNAs in seed development, as well as the genetic basis for selection of crop seeds. Notably, most imprinted genes that affect seed development in *Arabidopsis* are maternally expressed or biparentally expressed (2, 3). Expression of a paternally imprinted gene could exert relatively small effects on seed phenotypes, as observed in the Wt$^{SUP:AGL91}$ lines and interspecific hybrids (31). Alternatively, disrupting either maternally or paternally expressed genes could affect seed size as a mechanism for optimizing seed development and offspring fitness to facilitate parental-offspring coadaptation, as has been reported in rice (33). Elucidating the spatial-temporal role of siRNAs and DNA methylation in imprinting and endosperm development should help improve seed production in agricultural crops.

Materials and Methods

Detailed descriptions of the plant materials and methods used in this study are provided in *SI Appendix, SI Methods*. Small RNA sequencing data have been deposited in the Gene Expression Omnibus database repository (accession no. GSE47722), and the plasmid pFGUS2a sequence has been deposited in GenBank database (accession no. KC920577).

1. Borlaug NE (1973) Civilization's future: A call for international granaries. *Science Public Affairs* 24:7–15.
2. Jiang H, Köhler C (2012) Evolution, function, and regulation of genomic imprinting in plant seed development. *J Exp Bot* 63:4713–4722.
3. Raissig MT, Baroux C, Grossniklaus U (2011) Regulation and flexibility of genomic imprinting during seed development. *Plant Cell* 23:16–26.
4. Lu J, Zhang C, Baulcombe DC, Chen ZJ (2012) Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of *Arabidopsis* seeds. *Proc Natl Acad Sci USA* 109:5529–5534.
5. Mosher RA, et al. (2009) Uniparental expression of PolIV-dependent siRNAs in developing endosperm of *Arabidopsis*. *Nature* 460:283–286.
6. Onodera Y, et al. (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* 120:613–622.
7. Herr AJ, Jensen MB, Dalmay T, Baulcombe DC (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science* 308:118–120.
8. Erdmann RM, Satyaki PRV, Klosinska M, Gehring M (2017) A small RNA pathway mediates allelic dosage in endosperm. *Cell Rep* 21:3364–3372.
9. Pélissier T, Wassenegger M (2000) A DNA target of 30 bp is sufficient for RNA-directed DNA methylation. *RNA* 6:55–65.
10. Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11:204–220.
11. Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. *Nat Rev Genet* 15:394–408.
12. Martinez G, et al. (2018) Paternal easiRNAs regulate parental genome dosage in *Arabidopsis*. *Nat Genet* 50:193–198.
13. Schon MA, Nodine MD (2017) Widespread contamination of *Arabidopsis* embryo and endosperm transcriptome datasets. *Plant Cell* 29:608–617.
14. Belmonte MF, et al. (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the *Arabidopsis* seed. *Proc Natl Acad Sci USA* 110: E435–E444.
15. Hsieh TF, et al. (2009) Genome-wide demethylation of *Arabidopsis* endosperm. *Science* 324:1451–1454.
16. Havecker ER, et al. (2010) The *Arabidopsis* RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell* 22:321–334.
17. Kang IH, Steffen JG, Portereiko MF, Lloyd A, Drews GN (2008) The AGL62 MADS domain protein regulates cellularization during endosperm development in *Arabidopsis*. *Plant Cell* 20:635–647.
18. Bemer M, Heijmans K, Airoldi C, Davies B, Angenent GC (2010) An atlas of type I MADS box gene expression during female gametophyte and seed development in *Arabidopsis*. *Plant Physiol* 154:287–300.
19. Gehring M, et al. (2006) DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* 124:495–506.
20. Wang D, et al. (2010) Identification of transcription-factor genes expressed in the *Arabidopsis* female gametophyte. *BMC Plant Biol* 10:110.
21. Autran D, et al. (2011) Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. *Cell* 145:707–719.
22. Vu TM, et al. (2013) RNA-directed DNA methylation regulates parental genomic imprinting at several loci in *Arabidopsis*. *Development* 140:2953–2960.
23. Nodine MD, Bartel DP (2012) Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. *Nature* 482:94–97.
24. Ferguson-Smith AC (2011) Genomic imprinting: The emergence of an epigenetic paradigm. *Nat Rev Genet* 12:565–575.
25. DeChiara TM, Efstratiadis A, Robertson EJ (1990) A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. *Nature* 345:78–80.
26. Lau MM, et al. (1994) Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. *Genes Dev* 8: 2953–2963.
27. Filson AJ, Louvi A, Efstratiadis A, Robertson EJ (1993) Rescue of the T-associated maternal effect in mice carrying null mutations in Igf-2 and Igf2r, two reciprocally imprinted genes. *Development* 118:731–736.
28. Adams S, Vinkenoog R, Spielman M, Dickinson HG, Scott RJ (2000) Parent-of-origin effects on seed development in *Arabidopsis thaliana* require DNA methylation. *Development* 127:2493–2502.
29. Morin RD, et al. (2008) Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa*. *Genome Res* 18:571–584.
30. Lee EK, et al. (2011) A functional phylogenomic view of the seed plants. *PLoS Genet* 7: e1002411.
31. Kirkbride RC, et al. (2015) An epigenetic role for disrupted paternal gene expression in postzygotic seed abortion in *Arabidopsis* interspecific hybrids. *Mol Plant* 8: 1766–1775.
32. Ng DW, Lu J, Chen ZJ (2012) Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Curr Opin Plant Biol* 15:154–161.
33. Yuan J, et al. (2017) Both maternally and paternally imprinted genes regulate seed development in rice. *New Phytol* 216:373–387.