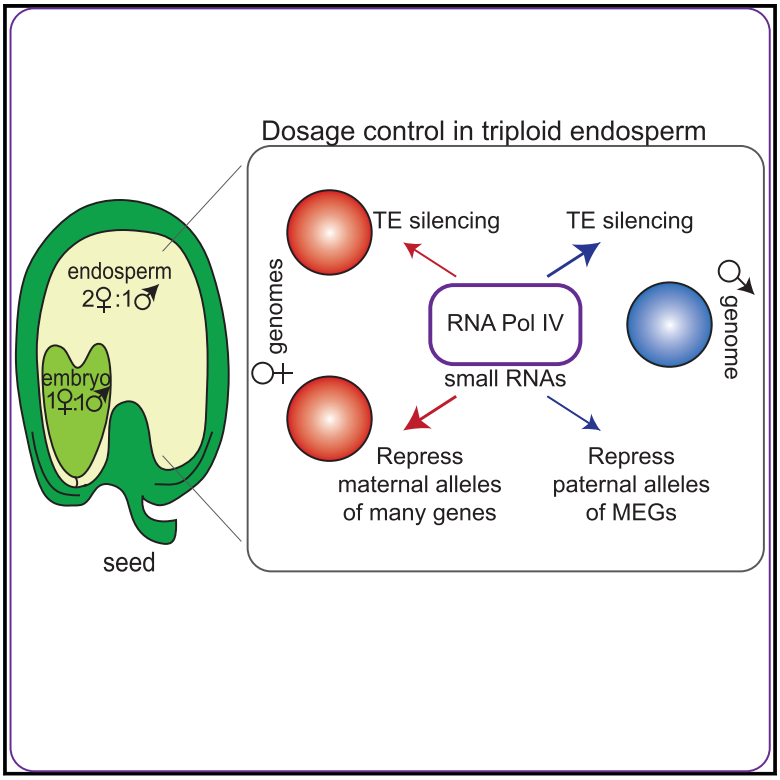
A Small RNA Pathway Mediates Allelic Dosage in Endosperm

Cell Reports

Report

A Small RNA Pathway Mediates Allelic Dosage in Endosperm

Graphical Abstract



Authors

Robert M. Erdmann, Prasad R.V. Satyaki, Maja Klosinska, Mary Gehring

Correspondence

mgehring@wi.mit.edu

In Brief

The endosperm is a triploid seed tissue critical for seed viability. Imprinted expression occurs there, but whether active regulation of maternal:paternal transcript ratios occurs on a broader scale is unknown. Erdmann et al. show that the *RNA Pol IV* small RNA pathway mediates dosage interactions between maternal and paternal genomes.

Highlights

- Genic 24-nt small RNAs (sRNAs) enriched in endosperm compared with other tissues
- Inverse relationship between allelic bias of mRNAs and sRNAs
- Maternal:paternal transcript ratio is actively regulated by *RNA Pol IV* pathway
- Mutations in *Pol IV* suppress negative effects of extra paternal genome dosage

Data and Software Availability

GSE94792

Erdmann et al., 2017, Cell Reports *21*, 3364–3372
December 19, 2017
https://doi.org/10.1016/j.celrep.2017.11.078

CellPress

OPEN ACCESS
CellPress

Cell Reports
Report

A Small RNA Pathway Mediates Allelic Dosage in Endosperm

Robert M. Erdmann,[1,2] Prasad R.V. Satyaki,[1] Maja Klosinska,[1] and Mary Gehring[1,2,3,*]
[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]Lead Contact
*Correspondence: mgehring@wi.mit.edu
https://doi.org/10.1016/j.celrep.2017.11.078

SUMMARY

Balance between maternal and paternal genomes within the triploid endosperm is necessary for normal seed development. The majority of endosperm genes are expressed in a 2:1 maternal:paternal ratio, reflecting genomic DNA content. Here, we find that the 2:1 transcriptional ratio is, unexpectedly, actively regulated. In *A. thaliana* and *A. lyrata*, endosperm 24-nt small RNAs are reduced in transposable elements and enriched in genes compared with the embryo. We find an inverse relationship between the parent of origin of sRNAs and mRNAs, with genes more likely to be associated with maternally than paternally biased sRNAs. Disruption of the Pol IV sRNA pathway causes a shift toward maternal allele mRNA expression for many genes. Furthermore, paternal inheritance of an *RNA Pol IV* mutation is sufficient to rescue seed abortion caused by excess paternal genome dosage. Thus, *RNA Pol IV* mediates the transcriptional balance between maternally and paternally inherited genomes in endosperm.

INTRODUCTION

Proper endosperm development is required for the formation of viable seeds. The seeds of flowering plants consist of three genetically and epigenetically distinct components: the two products of fertilization, diploid embryo and triploid endosperm, and the diploid seed coat, which is derived from maternal ovule tissues. Although the embryo has a 1:1 ratio of maternal to paternal genomes, the endosperm ratio of maternal to paternal genomes is 2:1. In the endosperm, genic maternal:paternal transcript ratios form a Gaussian distribution centered at 2:1 (Gehring et al., 2011). At the edges of this distribution are 100–200 imprinted genes that are expressed primarily from maternal or paternal alleles (Gehring et al., 2011; Hsieh et al., 2011). Differential DNA methylation and chromatin modifications have been shown to regulate the maternal to paternal transcript ratios of imprinted genes (Hsieh et al., 2011; Zhang et al., 2014; Moreno-Romero et al., 2016; Gehring and Satyaki, 2017). It is unknown whether there is active regulation of the maternal:paternal transcript ratios at genes that are expressed in endosperm but are not imprinted and, if so, what the identity of the regulators might be.

Small RNAs (sRNAs) have diverse functions in eukaryotes, including post-transcriptional gene silencing and transcriptional gene silencing via guided modifications to DNA or chromatin (Borges and Martienssen, 2015). In plants, DNA methylation in all sequence contexts (CG, CHG, and CHH, where H represents A, C, or T) is established and maintained by the RNA-directed DNA methylation (RdDM) pathway. The most abundant class of plant sRNAs are 24-nt sRNAs that primarily map to transposable elements (TEs) and other repetitive genomic regions (Matzke and Mosher, 2014). In the canonical RdDM pathway, the biogenesis of 24-nt sRNAs is initiated by RNA polymerase IV (Pol IV). Pol IV produces short non-coding transcripts that are converted into double-stranded RNA by the interacting RNA-dependent RNA polymerase RDR2 and are then diced by DCL3 into 24-nt sRNAs. 24-nt sRNAs can be loaded into AGO proteins to guide the DRM2 *de novo* DNA methyltransferase to matching DNA sequences, establishing transcriptional gene silencing by RdDM (Matzke and Mosher, 2014).

sRNAs are highly abundant within the reproductive tissues of *Arabidopsis thaliana* (Mosher et al., 2009). Previous studies have reported genome-wide, parent of origin-specific sRNA profiles from *Arabidopsis* seed pods (Mosher et al., 2009) or whole seeds (Pignatta et al., 2014). Endosperm maternally expressed imprinted genes (MEGs) are associated with paternally biased 24-nt sRNAs in whole seeds (Pignatta et al., 2014). Wild-type paternal *NRPD1*, which encodes the largest subunit of RNA Pol IV, is required in sporophytic tissues for paternal allele silencing of two MEGs in endosperm (Vu et al., 2013). Despite the possible connections to imprinting, the precise relationship between sRNAs in *Arabidopsis* endosperm and maternal or paternal allele expression or DNA methylation remains unknown. Here we investigated the distribution, parent of origin, and function of sRNAs in *Arabidopsis thaliana* and *Arabidopsis lyrata* endosperm. These experiments identified a genic sRNA population enriched in endosperm and uncovered a role for the sRNA pathway in widespread control of maternal:paternal transcript ratios.

RESULTS

Increased Accumulation of Genic sRNAs in Endosperm

To determine the identity and distribution of sRNAs in the two products of fertilization, we performed sRNA sequencing from *A. thaliana* embryos and endosperm isolated from multiple replicates of wild-type developing Columbia-0 (Col), Landsberg *erecta* (Ler), and Cvi seeds and from their F_1 hybrids, generating

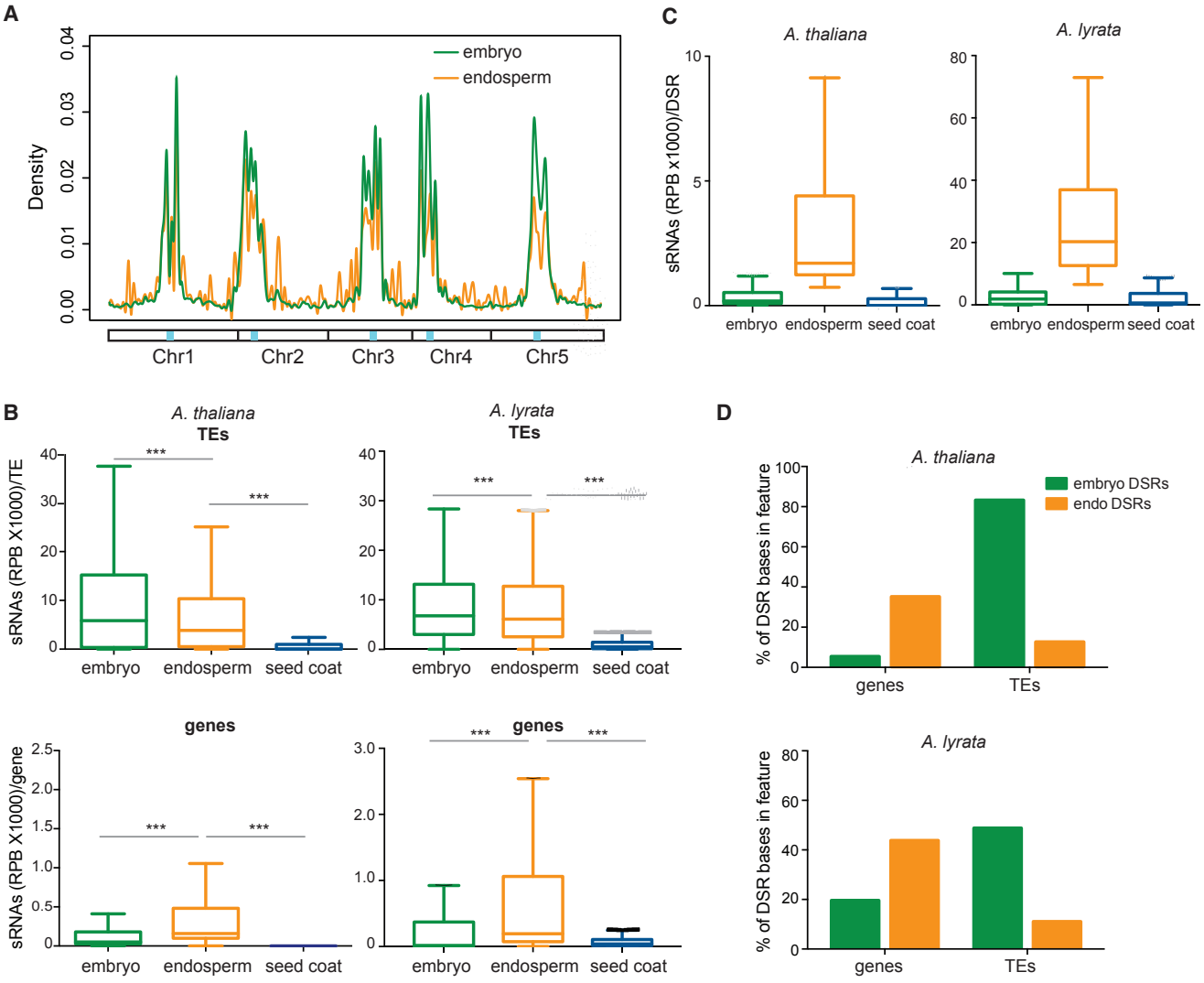


Figure 1. Endosperm Has a Unique sRNA Profile

(A) Kernel density plot of *A. thaliana* embryo and endosperm 24-nt sRNA abundance. Blue shading, centromeres.

(B) Boxplots of the average reads per bp (RPB) of 24-nt sRNAs per TE (top) or gene (bottom) for *A. thaliana* (left) and *A. lyrata* MN47 (right) seed tissues. Line, median; whiskers, 1.5× the interquartile range (IQR). Outliers are not shown. ***p < 2.2e−16 by two-sided Wilcoxon signed-rank test.

(C) 24-nt sRNA abundance in endosperm DSRs in seed tissues. Outliers are not shown.

(D) Overlap of DSRs with gene bodies and TEs.

See also Figures S1 and S2 and Tables S1, S2, S3, and S4.

around 250 million mapped sRNA reads for each tissue (Table S1). To determine the relevance of our findings outside of *A. thaliana*, we also profiled sRNAs in replicates of embryo and endosperm from the outcrossing perennial *A. lyrata* (Table S1). sRNAs were also sequenced from a limited number of *A. thaliana* and *A. lyrata* seed coat samples (Table S1). As expected, 24-nt sRNAs were the most abundant size class in embryo and endosperm, although endosperm had a somewhat higher fraction of 21-nt sRNAs (Figure S1). 24-nt sRNAs were relatively depleted in the seed coat in both species (Figure S1). sRNA profiles were highly correlated among biological replicates of the same genotype (for example, Pearson's R > 0.94 for *A. thaliana* endosperm replicates) and within but not between tissues (Table S2). Most embryo and endosperm sRNAs mapped to the areas of highest TE density, pericentric chromosome regions (Figure 1A), as observed in other *Arabidopsis* tissues (Greaves et al., 2012). However, the distribution of sRNAs was distinct between the two tissues (Figures 1A and 1B). TEs were associated with significantly lower levels of 24-nt sRNAs in endosperm compared with embryos, whereas gene bodies in endosperm were associated with significantly higher levels of 24-nt sRNAs compared with embryo, seed coat, or other tissues (Yang et al., 2017) in both *A. thaliana* and *A. lyrata* (Figure 1B; Figure S1).

To identify regions of tissue-enriched sRNA expression, termed here differential sRNA regions or DSRs, we binned the

24-nt fraction of the *A. thaliana* or *A. lyrata* embryo or endosperm sRNA population into overlapping 300-bp windows. DSRs were defined as at least 5.7-fold (log_2 = 2.5) more enriched in the embryo or endosperm at an adjusted p < 1e−10. Potential endosperm DSRs were filtered for possible seed coat contamination by removing regions of the genome with equivalent or greater sRNA abundance in the seed coat compared with endosperm (Figure 1C). This yielded 1.6 Mb of endosperm DSRs and 0.5 Mb of embryo DSRs in *A. thaliana* (Tables S3 and S4). The genomic distribution of embryo and endosperm DSRs was distinct and consistent with genome-wide trends (Figure 1A): embryo DSRs showed greater overlap with TEs, and endosperm DSRs showed greater overlap with genes (Figure 1D). The same patterns were observed in *A. lyrata* (Figures 1C and 1D; Figure S1; Tables S3 and S4). Although less prevalent, 21- to 22-nt DSRs were highly overlapping with 24-nt DSRs of the same type, with 60%–77% of bases in 21- to 22-nt DSRs also in 24-nt DSRs (Figure S1), indicating that these regions produce sRNAs of multiple sizes.

We examined whether endosperm DSRs were associated with any specific functional class of genes. In *A. thaliana*, 1,578 genes were overlapped by a 24-nt endosperm DSR (DSR within the gene body or within 500 bp). These genes were highly enriched for the gene ontology (GO) term regulation of transcription, DNA-dependent (Bonferroni-corrected p = 2.01e−44), particularly members of the MADS box and homeodomain-like transcription factor families. The 835 genes overlapped by *A. lyrata* endosperm DSRs were also enriched for the GO term regulation of transcription, DNA-dependent (Bonferroni-adjusted p = 5.82e−7). Indeed, homologs of 27% of the genes overlapped by *A. lyrata* endosperm DSRs were also associated with endosperm DSRs in *A. thaliana*, suggesting that accumulation of sRNAs in these genes is evolutionarily conserved (Table S3). In addition to enrichment in transcriptional regulators (Bonferroni-adjusted p = 1.52e−4), conserved genes were also enriched for chromatin modifiers (Bonferroni-adjusted p = 7.91e−5). Thus, we conclude that widespread accumulation of 24-nt sRNAs in genic regions is an evolutionarily conserved feature of endosperm.

Endosperm-Enriched 24-nt sRNAs Are Associated with Substantially Increased CHH DNA Methylation in *A. lyrata* but Not in *A. thaliana*

24-nt sRNAs are often associated with DNA methylation of matching sequences, a hallmark of which is the accumulation of CHH methylation. The reduced accumulation of 24-nt TE-associated sRNAs in the endosperm compared with the embryo (Figure 1) is consistent with the previously observed hypomethylation of TEs in the endosperm (Gehring et al., 2009; Ibarra et al., 2012; Pignatta et al., 2014). We determined whether DSRs were differentially methylated between embryo and endosperm. As expected, regions corresponding to embryo DSRs were more methylated in the embryo than endosperm in all cytosine contexts in *A. thaliana* and *A. lyrata* (Figure S2). In *A. lyrata,* CHG and CHH methylation in endosperm DSRs was much higher in the endosperm than embryo, also consistent with the role of 24-nt sRNAs in guiding DNA methyltransferases. By contrast, in *A. thaliana*, statistically significant but minor increases in CHH methylation were observed in endosperm DSRs, with no significant difference in CHG methylation (Figure S2). Even homologous genes that were associated with endosperm DSRs in both species displayed distinct methylation profiles (Figure S2). The greater CHH hypermethylation in *A. lyrata* endosperm DSRs was consistent with higher 24-nt sRNAs abundance in endosperm DSRs in *A. lyrata* compared with *A. thaliana* (Figure 1C; Figure S2). We previously reported that *A. lyrata* endosperm is hypermethylated in the CHG context on the maternal allele in a subset of gene bodies, which is associated with reduced maternal allele expression (Klosinska et al., 2016). Indeed, endosperm DSRs had higher non-CG methylation on maternal alleles compared with paternal alleles in *A. lyrata* (Figure S2), suggesting that the methylation patterns might be related to allele-specific sRNA populations.

Parental sRNA Bias Is Inversely Correlated with Parental mRNA Bias

Several of the profiled sRNA populations were derived from a cross between two wild-type polymorphic parents (Table S1), allowing examination of the contribution of maternally and paternally inherited genomes to sRNA biogenesis. In *A. thaliana,* 60% of endosperm sRNAs and 50% of embryo sRNAs were derived from the maternally inherited genome, as were 58% and 46% of *A. lyrata* endosperm and embryo sRNAs (Figure 2A). The expectation is that 67% of sRNAs should be maternal; thus, these data reveal a paternal bias for endosperm 24-nt sRNAs at the level of the entire genome. A previous report indicated that sRNAs in *A. thaliana* endosperm were exclusively maternally derived but did not sequence sRNAs from isolated endosperm material (Mosher et al., 2009). Our findings indicate that parent of origin-specific sRNA populations in *Arabidopsis* endosperm are consistent with sRNA data from rice and maize endosperm (Rodrigues et al., 2013; Xin et al., 2014) and our previous whole-seed data (Pignatta et al., 2014), in which both maternally and paternally derived sRNAs were detected. We found that maternally and paternally derived embryo sRNAs had identical genomic distributions and were concentrated proximal to the centromeres, whereas, in the endosperm, maternally derived sRNAs were relatively more evenly distributed across the chromosomes (Figure 2B). Examination of the fraction of maternally derived sRNAs in genes and TEs in *A. thaliana* Col-L*er*, Col-Cvi, and *A. lyrata* Karhumäki (Kar) × MN F_1 endosperm indicated that TEs were associated with a lower fraction of maternal sRNAs than genes (p < 2.2e−16 in all comparisons) (Figure 2C). Thus, the higher than expected fraction of paternal sRNAs in endosperm (Figure 2A) is likely caused by a reduction of maternally derived sRNAs in TEs (Figure 2C).

We identified imprinted sRNA regions across the entire genome for Col-L*er* and Col-Cvi reciprocal cross datasets (Figure 2D; Table S5), using the same strategy we have previously employed to identify imprinted genes (Gehring et al., 2011; Pignatta et al., 2014). The Col-L*er* and Col-Cvi results were highly similar, with about 40% of bases in imprinted sRNA regions identified in one dataset also identified in the other (Figure 2D). sRNAs in endosperm DSRs were maternally biased (Figure 2C). Approximately 50%–70% of bases in maternally biased imprinted sRNA regions overlapped endosperm DSRs, in

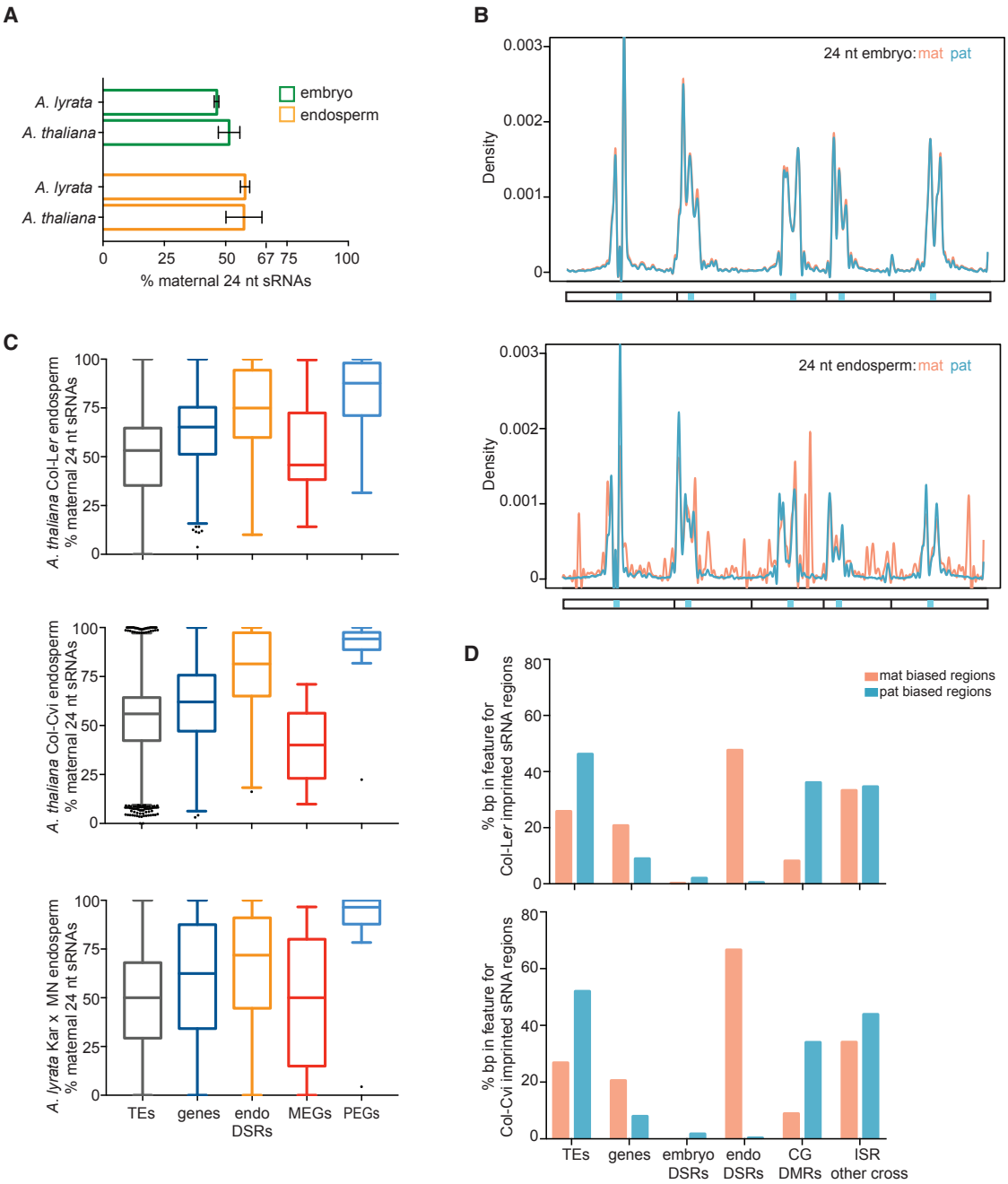


Figure 2. Endosperm sRNAs Are Produced from Both Maternally and Paternally Inherited Genomes, with Some Regions of Parental Bias

(A) Average percent sRNAs derived from maternally inherited genomes. Mean ± SD is plotted.

(B) Kernel density plot of maternally and paternally derived 24-nt sRNAs in *A. thaliana* embryo and endosperm.

(C) Percent maternal 24-nt sRNAs in features in endosperm from *A. thaliana* Col-L*er* reciprocal crosses (top), *A. thaliana* Col-Cvi reciprocal crosses (center), and *A. lyrata* Kar × MN (bottom). Only features with at least 20 allele-specific reads and a seed coat/endosperm RPB ratio of less than 1 were plotted. Whiskers, 1.5 IQR.

(D) Percent of maternally or paternally biased sRNA regions that overlap the indicated feature for Col-L*er* (top) and Col-Cvi (bottom) reciprocal cross datasets. Values add to greater than 100% because some features are coincident. CG differentially methylated regions (DMRs), regions less methylated in endosperm than embryo. ISR, imprinted sRNA region.

See also Table S5.

contrast to the almost complete lack of overlap of paternally biased imprinted sRNA regions with endosperm DSRs (Figure 2D). Similarly, paternally expressed imprinted genes (PEGs) were biased toward sRNAs derived from the maternally inherited genome ($p < 5e{-}5$) (Figure 2C), and 73% of PEGs were associated with an endosperm DSR in the gene body or within 500 bp (Table S3). We previously reported that MEGs were enriched for paternally derived sRNAs in whole seeds (Pignatta et al., 2014). Consistent with this finding, in the endosperm the fraction of maternally derived 24-nt sRNAs in MEGs was reduced compared with all genes (Figure 2C). Additionally, a number of MEGs, including *FWA*, *SDC*, and *MYB3R2*, were associated with paternally biased imprinted sRNA regions within the gene or in proximal flanking regions (Table S5). In *A. lyrata,* MEGs and PEGs were also associated with reduced and increased fractions of sRNA derived from the maternally inherited genome, respectively (Figure 2C). These data suggest that sRNAs are associated with the silenced allele of imprinted genes in both *A. thaliana* and *A. lyrata*. However, endosperm DSRs and parentally biased sRNA regions were not only associated with imprinted genes.

Thus, to determine whether there was a general relationship between the maternal:paternal ratios of endosperm sRNAs and mRNAs, we examined the percent maternal allele mRNAs per gene as a function of the maternal fraction of overlapping 24-nt sRNAs for 829 genes with sufficient allele-specific information (Figure 3A). This analysis revealed a more general negative association between percent maternal allele sRNA and percent maternal allele mRNA—the maternal allele mRNA fraction for genes decreased as the percentage of maternal allele sRNAs associated with those genes increased (Figure 3A). These data suggest that there is a reciprocal relationship between sRNA and mRNA production that is not restricted to imprinted genes.

The Ratio of Maternal:Paternal Transcripts Shifts in *RNA Pol IV* Mutants

If the negative association between the parent of origin of 24-nt sRNAs and mRNAs is causal, eliminating *Pol IV*-dependent sRNAs should affect the relative ratio of maternal to paternal allele transcripts. *NRPD1* encodes the largest subunit of RNA Pol IV (Onodera et al., 2005). We examined parent of origin specific gene expression in wild-type (WT) and *nrpd1* mutant endosperm in L*er* x Col crosses (Table S6). Consistent with endosperm having a unique 24-nt sRNA pattern, we found broader misregulation of gene expression in *nrpd1* endosperm than in *nrpd1* seedlings (Yang et al., 2017), with the most significant change being an increase in gene expression compared with the WT (Figure S3; Table S7). In *nrpd1* mutant endosperm, the expression of most genes that could be assayed shifted toward expression from the maternal allele by, on average, 5% compared with the WT (Figure 3B; Figure S3). There was an inverse relationship between the parent of origin of mRNAs and the degree of maternal shift in *nrpd1* endosperm; the most paternally biased mRNAs had the largest shifts toward maternal allele expression in *nrpd1* (Figure 3C). Additionally, genes with larger increases in maternal allele transcripts also had increased total gene expression and higher levels of associated 24-nt sRNA in the WT (Figure S3). To ensure that contamination from the maternal seed coat (Schon and Nodine, 2017) was not responsible for the shift toward expression from maternal alleles in *nrpd1* mutant endosperm, we examined allelic expression in a set of genes specifically expressed in the endosperm (Belmonte et al., 2013) and also observed the shift (Figure 3D; Table S8). Together, these results are consistent with genic 24-nt sRNAs being associated with a large fraction of genes in WT endosperm (Figure 1; Figure S1; Table S3), with more of those 24-nt sRNAs being biased toward maternal allele expression than paternal allele expression (Figures 2A and 2D; Table S5), and the removal of sRNAs resulting in allelic derepression (Figures 3B and 3C; Figure S3).

NRPD1 expression is paternally biased in endosperm 6 days after pollination (DAP) in *A. thaliana*, with, on average, 55% (range of 43%–69%) of transcripts from the paternally inherited allele compared with the expected 33% (Gehring et al., 2011; Pignatta et al., 2014). To determine whether mutation of the paternally inherited *NRPD1* allele was sufficient to alter the allelic ratios, we also performed mRNA sequencing (mRNA-seq) on L*er* × Col and L*er* × Col *nrpd1* F_1 endosperm. The shift toward increased expression from maternal alleles was again observed, although to a lesser extent than when the endosperm was homozygously mutant for *nrpd1* (Figure 3E; Figure S3). These results indicate that mutations in *Pol IV* result in a shift in transcriptional balance toward maternally inherited alleles for many genes throughout the genome, likely through both direct and indirect effects.

Mutations in *RNA Pol IV* Suppress Paternal Genomic Excess Phenotypes

Shifting the dosage of maternal and paternal genomes in endosperm is thought to be a driving force behind the seed abortion that results when diploid females are crossed to tetraploid males, a situation referred to as "paternal genomic excess" (Lin, 1984; Scott et al., 1998; Pennington et al., 2008). Given the excess maternal allele expression in *nrpd1* mutant endosperm (Figure 3; Figure S3), we reasoned that *nrpd1* mutants might suppress the phenotypic effects of increased paternal genomes. Tetraploid *nrpd1* mutants in the Col background were created by colchicine treatment. Seed abortion and seed germination phenotypes from crosses between WT or *nrpd1* mutant parents were assessed (Figure 4). As expected, crosses between WT diploid females and WT tetraploid males (producing endosperm with a 1:1 maternal:paternal ratio) resulted in 81% aborted seeds, with another 10% exhibiting an abnormal phenotype (Figures 4A and 4B). By contrast, when diploid *nrpd1* females were crossed to tetraploid *nrpd1* males, seed abortion was reduced to 24% (Figures 4A and 4B), indicating substantial rescue. 85% of tested seeds from this cross germinated (Figure 4C) and were confirmed to be triploid. We also investigated whether having one parent mutant for *NRPD1* was sufficient to suppress the seed abortion phenotype. Seeds lacking the paternal *NRPD1* allele displayed reduced seed abortion to an even greater extent, 13%, and were able to germinate and form true leaves (Figure 4). Suppression of seed abortion was also observed in crosses between WT L*er* females and tetraploid Col *nrpd1*

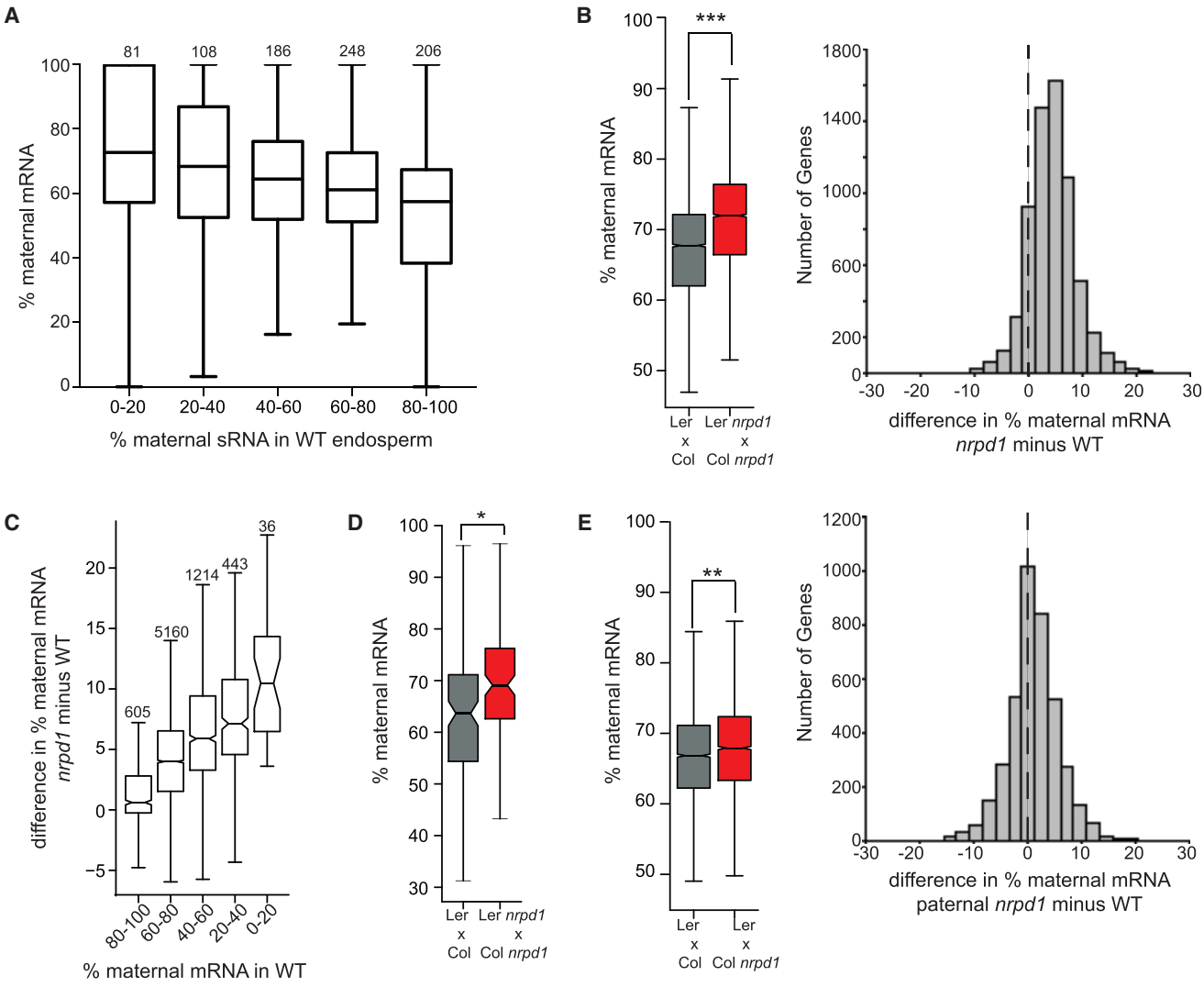


Figure 3. Loss of Pol *IV* Is Associated with a Genome-wide Shift toward Maternal Allele mRNA Expression in Endosperm

(A) Maternal sRNAs are negatively associated with maternal mRNAs. Boxplots show the percent maternal allele mRNAs per gene in each sRNA category for wild-type (WT) L*er* × Col crosses. The number of genes in each sRNA category is shown above the plot. Genes had at least 10 allele-specific mRNA and sRNA reads.

(B) Left: increased maternal allele expression in *nrpd1* mutants (L*er nrpd1* × Col *nrpd1*) compared with WT (L*er* × Col). ***$p < 2.2e{-}16$. Right: histogram of the difference between *nrpd1* and WT maternal mRNA fraction for individual genes (n = 6,675) in endosperm.

(C) Extent of shift to maternal allele expression in *nrpd1* endosperm is inversely correlated to percent maternal allele mRNA in the WT. $p < 0.003$ for all comparisons.

(D) Same as in (B) except for a subset of 125 genes expressed in the endosperm and not the seed coat. *p = 0.001246.

(E) Same as in (B) except for crosses between WT females and *nprd1* mutant males (L*er* × Col *nrpd1*) (n = 4,019 genes). **$p = 2.77e{-}10$. For all boxplots, the line is median, and whiskers are 1.5 IQR. All p values were determined with two-tailed Wilcoxon test.

See also Figure S3 and Tables S6, S7, and S8.

males (Figure S4). By contrast, inheritance of the *nrpd1* allele from the diploid female parent only did not suppress seed abortion in a cross to a WT tetraploid male (Figures 4B and 4C). Although the molecular mechanism of triploid seed rescue by *nrpd1* mutations remains unknown, together these data suggest that the negative effects of an extra paternal endosperm genome might be suppressed by the relatively higher maternal allele expression caused by mutations in *NRPD1*, either at the level of the entire genome or at specific genes. Alternatively, or in addition, seed rescue may be dependent on expression differences of individual genes (Table S7).

DISCUSSION

Transcriptional dosage control has been described with regard to the ratio of sex chromosomes to autosomes and in polyploids and aneuploids (Disteche, 2016). The original application of allele-specific mRNA-seq to endosperm suggested that the

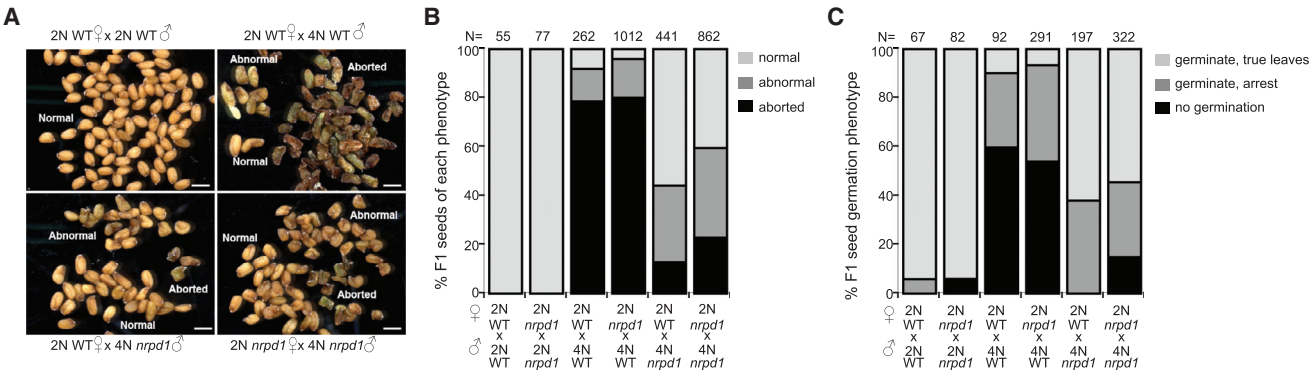


Figure 4. Mutations in *NRPD1* Suppress Paternal Genomic Excess Phenotypes

(A) Seed phenotypes observed in crosses among *A. thaliana* WT Col-0 diploids, WT tetraploids, and diploid and tetraploid *nrpd1* mutants. The female in the cross is listed first. Scale bar, 500 μm.

(B) Quantification of F_1 seed phenotypes.

(C) F_1 seed germination. Seeds that germinated but did not develop beyond the cotyledon stage were classified as arrested. N, number of seeds analyzed. For (B) and (C), the two-tailed p value for chi-square test is < 1.0e−4 for comparisons between 2N WT × 4N WT and 2N WT × 4N *nrpd1* or 2N *nrpd1* × 4N *nrpd1*. See also Figure S4.

endosperm was not subject to parent of origin specific global dosage control; most genes were expressed in a ratio reflective of genomic DNA content (Gehring et al., 2011; Hsieh et al., 2011). This study has revealed that maintaining the expected 2:1 ratio at many genes is an active process mediated by *RNA Pol IV*. In its absence, expression shifts toward more expression from maternally inherited genomes. Thus, our results have revealed previously hidden transcriptional dosage control. Several questions remain unanswered, foremost among them why an active process is required to maintain a seemingly default transcriptional ratio. One possibility is that this is a necessary response to the alterations of the maternal epigenome that are initiated before fertilization. Maternally inherited endosperm DNA is hypomethylated because of active DNA demethylation before fertilization, an essential process that establishes gene imprinting (Huh et al., 2008; Gehring and Satyaki, 2017). Additionally, endosperm chromatin is decondensed compared with chromatin from other cell types, and this might be specific to the maternally inherited genomes (Baroux et al., 2007; Pillot et al., 2010). The combined effect of DNA demethylation and chromatin decondensation could render the maternally inherited genomes more permissive to transcription, and thus, in the absence of an active mechanism, expression from maternal genomes would be favored. Our findings raise the possibility that hidden dosage control may also exist in other systems, particularly when genomes or chromosomes are in distinct epigenetic states.

The precise mechanism by which 24-nt sRNAs exert their effects on genes in the endosperm remains unknown and will require further experimentation. Although *A. lyrata* genes associated with endosperm DSRs are hypermethylated, an epigenetic state associated with transcriptional repression, the methylation gains in *A. thaliana* endosperm are quite modest (Figure S2). One possibility is that transcription by the Pol IV complex reduces the ability of Pol II to access the same regions.

The overall significance of our findings is highlighted by the ability of mutations in a gene encoding a Pol IV subunit to suppress seed abortion caused by excess paternal genomes (Figure 4)—the only developmental phenotype so far ascribed to *nrpd1* mutants. Our findings place *NRPD1* among a relatively small number of genes that, when mutated, can paternally suppress triploid seed abortion (Kradolfer et al., 2013; Schatlowski et al., 2014; Wolff et al., 2015; Huang et al., 2017). Although the precise molecular basis for seed abortion suppression by *NRPD1* mutations remains unknown, our data indicate that a critical function of *Pol IV* is during reproductive development.

Parental conflict is thought to underlie the evolution of gene imprinting in flowering plants and mammals (Haig, 2013). The idea behind this theory is that maternally and paternally inherited genomes have differing interests with regard to maternal investment in offspring (in plants, endosperm), the outcome of which is parent of origin specific (imprinted) gene expression. Our data suggest that there are two outcomes of conflict in endosperm, operating at different scales: the first in the form of imprinting at individual genes, and the second acting more globally on a substantial portion of the genome. RdDM pathway genes, which themselves are imprinted, function at both of these scales. The phenomenon described here is likely to be broadly relevant. For example, endosperm from interspecific hybrid tomato seed have phenotypes suggestive of maternal:paternal genomic imbalance and display a global shift toward expression from maternally inherited genomes (Florez-Rueda et al., 2016). In conclusion, our results demonstrate that the *Pol IV* sRNA pathway mediates dosage interactions between maternal and paternal genomes, highlighting an active tug of war between parents with regard to transcriptional outcomes in the endosperm.

EXPERIMENTAL PROCEDURES

Plant Material

A. thaliana plants were grown in a greenhouse with 16-hr days at approximately 21°C. For crosses, flowers were emasculated 2 days prior to pollination. To introgress the *nrpd1a-4* allele into the L*er* background, we performed 4 backcrosses of the SALK_083051 line from Col into L*er*, followed by self-fertilization to obtain homozygous *nrpd1a-4* mutants. In Figure 3B and

Figure S3B, each of the three replicates was derived from different backcross progeny. A description of Col and *nrpd1* tetraploids can be found in the Supplemental Experimental Procedures. *A. lyrata* plants of the Kar and MN47 strains were grown in a growth chamber with 16-hr days at 20°C (Klosinska et al., 2016).

Seed Dissection and RNA Isolation

A. thaliana seeds were dissected 6 DAP, corresponding to the torpedo stage of embryogenesis. *A. lyrata* seed dissections were performed at 15 DAP, at the late torpedo/walking stick stage of development. Details on RNA isolation can be found in the Supplemental Experimental Procedures.

sRNA Library Construction and Analysis

Libraries for Illumina sequencing were constructed using the NEXTflex sRNA-Seq Kit v2 as directed (Bioo Scientific). 40-base single-end sequencing of sRNA libraries was performed on an Illumina HiSeq machine (Table S1).

Low-quality read ends were trimmed (fastq_quality_trimmer, -t 20 and -l 25). Adapters were removed with cutadapt (−0 6 -m 26–discard-untrimmed) (Martin, 2011). sRNA libraries were appended with 4-base randomized barcodes immediately 3′ and 5′ of the sRNA read, which were used to remove PCR duplicates (any read of the same sRNA sequence with identical flanking barcodes on both ends) before being removed themselves. 18- to 26-nt reads were aligned using Bowtie 1.1.1 (Langmead et al., 2009), with two mismatches allowed but perfect matches favored (-v 2 –best), using a metagenomic approach to reduce mapping bias as described in the Supplemental Experimental Procedures. With a final set of aligned reads, a single base read depth value for sRNAs of any desired size class was created; each size of sRNA was normalized separately so that each base of a read n nucleotides in length contributed 1/n of a read to an overlapping genomic position. Alignment files were run through bedtools genomecov (-d -scale: adjusted to account for individual library size; bedtools v. 2.23.0). Per-position values for all libraries pooled within a given analysis were averaged, and the resulting per-position average values could then be intersected with any genomic locus of interest using bedtools intersect (bedtools v. 2.23.0). This single-base averaging method was used for all analyses that reported average sRNA reads per base pair (RPB).

Identifying DSRs

A. thaliana 24-nt sRNAs or 21- to 22-nt sRNAs were binned into 300-bp windows with 150-bp overlaps using bedtools coverage (bedtools v. 2.23.0). Embryo and endosperm sRNA window read values were used as input for DESeq2, with the sizeFactors command used to provide RPM normalization between libraries (Love et al., 2014). For *A. thaliana*, the 14-embryo and 16-endosperm individual libraries were considered biological replicates. For *A. lyrata*, DSRs were defined separately for MN × MN and Kar × MN libraries. Embryo DSRs were defined as having a $\log_2$ embryo/endosperm count value of greater than 2.5, with a false discovery rate (FDR)-adjusted $p < 1 \times 10^{-10}$. Endosperm DSRs have a $\log_2$ embryo/endosperm count value of less than −2.5, with the same p value requirement. Endosperm DSRs with a seed coat/endosperm RPB ratio of ≥ 1 were removed.

Identifying *A. thaliana* Imprinted sRNA Regions

We used our previously described imprinting pipeline to define statistically significant parentally biased 24-nt sRNA windows (Pignatta et al., 2014). Additional details can be found in the Supplemental Experimental Procedures.

mRNA Library Construction and Data Analysis

Details can be found in the Supplemental Experimental Procedures. For both sets of mRNA-seq data, only genes that had at least 15 reads in all replicates of both WT and mutant genotypes and varied between replicates of a genotype by an SD of 10% or less were tested for a maternal shift (Figures 3B, 3D, and 3E).

Statistical Procedures

Statistical significance of differences in the distribution of sRNA abundance in two tissues was determined with the Wilcoxon signed-rank test (non-parametric, paired, two-tailed test). Tests for differences in percent maternal mRNA expression between two samples were performed with a two-tailed Wilcoxon test (wilcox.test in R). Differences in DNA methylation values in regions of interest were evaluated with a one-tailed t test. To identify DSRs, 14 embryo replicates were compared with 16 endosperm replicates (*A. thaliana*), three embryo replicates were compared with three endosperm replicates (*A. lyrata* Kar × MN), and three embryo replicates were compared with two endosperm replicates (*A. lyrata* MN x MN) using DESeq2. For imprinted sRNA regions, statistical significance ($p < 0.01$, Benjamini-corrected for multiple hypothesis testing) of deviation from a 2:1 maternal:paternal ratio was determined with Fisher's exact test. The chi-square test was used to test the null hypothesis that crosses where one of the parents was mutant for *nrpd1* produced the same ratio of normal:abnormal:aborted seeds as crosses between diploid females and tetraploid males.

DATA AND SOFTWARE AVAILABILITY

The accession number for the sequencing data reported in this paper is NCBI GEO: GSE94792.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and eight tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2017.11.078.

ACKNOWLEDGMENTS

We thank Colette L. Picard and Prat Thiru for advice and assistance with bioinformatics analyses. This research was funded by NSF CAREER grant 1453459 (to M.G.). R.M.E. was supported by an NSF graduate research fellowship.

AUTHOR CONTRIBUTIONS

R.M.E., P.R.V.S., and M.G. conceived and designed the experiments. R.M.E., P.R.V.S., and M.K. performed the experiments. All authors analyzed the data. R.M.E., P.R.V.S., and M.G. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 4, 2017
Revised: September 18, 2017
Accepted: November 21, 2017
Published: December 19, 2017

REFERENCES

Baroux, C., Pecinka, A., Fuchs, J., Schubert, I., and Grossniklaus, U. (2007). The triploid endosperm genome of Arabidopsis adopts a peculiar, parental-dosage-dependent chromatin organization. Plant Cell *19*, 1782–1794.

Belmonte, M.F., Kirkbride, R.C., Stone, S.L., Pelletier, J.M., Bui, A.Q., Yeung, E.C., Hashimoto, M., Fei, J., Harada, C.M., Munoz, M.D., 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.

Borges, F., and Martienssen, R.A. (2015). The expanding world of small RNAs in plants. Nat. Rev. Mol. Cell Biol. *16*, 727–741.

Disteche, C.M. (2016). Dosage compensation of the sex chromosomes and autosomes. Semin. Cell Dev. Biol. *56*, 9–18.

Florez-Rueda, A.M., Paris, M., Schmidt, A., Widmer, A., Grossniklaus, U., and Städler, T. (2016). Genomic imprinting in the endosperm is systematically perturbed in abortive hybrid tomato seeds. Mol. Biol. Evol. *33*, 2935–2946.

Gehring, M., and Satyaki, P.R. (2017). Endosperm and imprinting, inextricably linked. Plant Physiol. *173*, 143–154.

Gehring, M., Bubb, K.L., and Henikoff, S. (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science *324*, 1447–1451.

Gehring, M., Missirian, V., and Henikoff, S. (2011). Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. PLoS ONE *6*, e2387.

Greaves, I.K., Groszmann, M., Ying, H., Taylor, J.M., Peacock, W.J., and Dennis, E.S. (2012). Trans chromosomal methylation in Arabidopsis hybrids. Proc. Natl. Acad. Sci. USA *109*, 3570–3575.

Haig, D. (2013). Kin conflict in seed development: an interdependent but fractious collective. Annu. Rev. Cell Dev. Biol. *29*, 189–211.

Hsieh, T.F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada, J.J., Zilberman, D., and Fischer, R.L. (2011). Regulation of imprinted gene expression in Arabidopsis endosperm. Proc. Natl. Acad. Sci. USA *108*, 1755–1762.

Huang, F., Zhu, Q.H., Zhu, A., Wu, X., Xie, L., Wu, X., Helliwell, C., Chaudhury, A., Finnegan, E.J., and Luo, M. (2017). Mutants in the imprinted PICKLE RELATED 2 gene suppress seed abortion of fertilization independent seed class mutants and paternal excess interploidy crosses in Arabidopsis. Plant J. *90*, 383–395.

Huh, J.H., Bauer, M.J., Hsieh, T.F., and Fischer, R.L. (2008). Cellular programming of plant gene imprinting. Cell *132*, 735–744.

Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., et al. (2012). Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science *337*, 1360–1364.

Klosinska, M., Picard, C.L., and Gehring, M. (2016). Conserved imprinting associated with unique epigenetic signatures in the Arabidopsis genus. Nat. Plants *2*, 16145.

Kradolfer, D., Wolff, P., Jiang, H., Siretskiy, A., and Köhler, C. (2013). An imprinted gene underlies postzygotic reproductive isolation in *Arabidopsis thaliana*. Dev. Cell *26*, 525–535.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lin, B.Y. (1984). Ploidy barrier to endosperm development in maize. Genetics *107*, 103–115.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12.

Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat. Rev. Genet. *15*, 394–408.

Moreno-Romero, J., Jiang, H., Santos-González, J., and Köhler, C. (2016). Parental epigenetic asymmetry of PRC2-mediated histone modifications in the *Arabidopsis* endosperm. EMBO J. *35*, 1298–1311.

Mosher, R.A., Melnyk, C.W., Kelly, K.A., Dunn, R.M., Studholme, D.J., and Baulcombe, D.C. (2009). Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. Nature *460*, 283–286.

Onodera, Y., Haag, J.R., Ream, T., Costa Nunes, P., Pontes, O., and Pikaard, C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell *120*, 613–622.

Pennington, P.D., Costa, L.M., Gutierrez-Marcos, J.F., Greenland, A.J., and Dickinson, H.G. (2008). When genomes collide: aberrant seed development following maize interploidy crosses. Ann. Bot. *101*, 833–843.

Pignatta, D., Erdmann, R.M., Scheer, E., Picard, C.L., Bell, G.W., and Gehring, M. (2014). Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. eLife *3*, e03198.

Pillot, M., Baroux, C., Vazquez, M.A., Autran, D., Leblanc, O., Vielle-Calzada, J.P., Grossniklaus, U., and Grimanelli, D. (2010). Embryo and endosperm inherit distinct chromatin and transcriptional states from the female gametes in Arabidopsis. Plant Cell *22*, 307–320.

Rodrigues, J.A., Ruan, R., Nishimura, T., Sharma, M.K., Sharma, R., Ronald, P.C., Fischer, R.L., and Zilberman, D. (2013). Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. Proc. Natl. Acad. Sci. USA *110*, 7934–7939.

Schatlowski, N., Wolff, P., Santos-González, J., Schoft, V., Siretskiy, A., Scott, R., Tamaru, H., and Köhler, C. (2014). Hypomethylated pollen bypasses the interploidy hybridization barrier in Arabidopsis. Plant Cell *26*, 3556–3568.

Schon, M.A., and Nodine, M.D. (2017). Widespread contamination of Arabidopsis embryo and endosperm transcriptome data sets. Plant Cell *29*, 608–617.

Scott, R.J., Spielman, M., Bailey, J., and Dickinson, H.G. (1998). Parent-of-origin effects on seed development in *Arabidopsis thaliana*. Development *125*, 3329–3341.

Vu, T.M., Nakamura, M., Calarco, J.P., Susaki, D., Lim, P.Q., Kinoshita, T., Higashiyama, T., Martienssen, R.A., and Berger, F. (2013). RNA-directed DNA methylation regulates parental genomic imprinting at several loci in Arabidopsis. Development *140*, 2953–2960.

Wolff, P., Jiang, H., Wang, G., Santos-González, J., and Köhler, C. (2015). Paternally expressed imprinted genes establish postzygotic hybridization barriers in *Arabidopsis thaliana*. eLife *4*, 10074.

Xin, M., Yang, R., Yao, Y., Ma, C., Peng, H., Sun, Q., Wang, X., and Ni, Z. (2014). Dynamic parent-of-origin effects on small interfering RNA expression in the developing maize endosperm. BMC Plant Biol. *14*, 192.

Yang, R., Zheng, Z., Chen, Q., Yang, L., Huang, H., Miki, D., Wu, W., Zeng, L., Liu, J., Zhou, J.X., et al. (2017). The developmental regulator PKL is required to maintain correct DNA methylation patterns at RNA-directed DNA methylation loci. Genome Biol. *18*, 103.

Zhang, M., Xie, S., Dong, X., Zhao, X., Zeng, B., Chen, J., Li, H., Yang, W., Zhao, H., Wang, G., et al. (2014). Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. Genome Res. *24*, 167–176.