Resetting of 24-nt siRNA landscape is initiated before the first zygotic division in rice

Chenxin Li,[1,5] Jonathan I. Gent,[2,5,6] Hengping Xu,[3,5] Hong Fu,[3] Scott D. Russell,[3,6] and Venkatesan Sundaresan[1,4,6]

[1] Department of Plant Biology, University of California, Davis, California 95616, USA
[2] Department of Plant Biology, University of Georgia, Athens, Georgia 30602, USA
[3] Department of Microbiology and Plant Biology, University of Oklahoma, Norman, Oklahoma 73019, USA
[4] Department of Plant Sciences, University of California, Davis, California 95616, USA
[5] These authors contributed equally.
[6] Co-corresponding authors: sundar@ucdavis.edu, srussell@ou.edu, gent@uga.edu

ABSTRACT

The zygote, a totipotent stem cell, constitutes a critical sage of the life cycle of sexually reproducing organisms. It is produced by the fusion of two differentiated cells, egg and sperm, that in plants have radically different siRNA transcriptomes from each other and from multicellular embryos, implying extensive reprogramming of the epigenome in plant gametes. Here, we have characterized the small RNA transcriptome of rice zygotes and found widespread distribution of 24nt siRNA loci in zygotes, the majority of which are not detected in the egg cell. These newly detected zygote siRNA loci had similar genomic distribution to canonical siRNA loci. Unlike most egg cell siRNA loci, the newly detected zygote siRNA loci had abundant CHH methylation in the embryo that was dependent on the RNA-directed DNA methyltransferase DRM2. A small fraction of siRNA loci (~6%) called siren loci accounted for 75% of all zygote siRNAs, and corresponded to siren loci detected in egg cell and ovary, but not endosperm. The zygote siren loci likely represent maternal carryover and are not associated with embryo hypermethylation. Taken together, our results suggest that resetting of the gametic epigenome towards the canonical vegetative profile is initiated before the first embryonic division.

Introduction

Gametes and zygotes constitute critical developmental stages in the life cycle of all sexually reproducing organisms. During fertilization, the egg cell fuses with a sperm cell to form the zygote, which is an undifferentiated and totipotent stem cell that initiates embryogenesis. Flowering plants undergo double fertilization, with a second sperm cell that fuses with the central cell, which gives rise to the endosperm, a nutritive tissue that nurtures the developing embryo or germinating seedling [reviewed in (Lord and Russell, 2002)]. The maternal to zygotic transition (MZT) in animals consists of two steps: zygotic genome activation (ZGA) and maternal RNA degradation. In animals, early embryogenesis is controlled by maternal gene products pre-deposited in the egg cell. Depending on the organism, the zygotic genome does not become transcriptionally active until a number of cell divisions occur (Tadros and Lipshitz, 2009). Recent studies show that MZT in flowering plants differs markedly from most animals [reviewed in (Armenta-Medina and Gillmor, 2019)]. In rice zygotes, thousands of genes are upregulated in zygotes, many of which are undetected in the egg cell, consistent with similar observations in maize and *Arabidopsis* zygotes (Chen et al., 2017; Zhao et al., 2019). Furthermore, zygotic transcription was shown to be required for early embryogenesis (Kao and Nodine, 2019; Zhao et al., 2019). These observations suggest that in angiosperms, unlike most animals, zygotes are transcriptionally active, and plant ZGA occurs in the zygote. However, similar to animals, ZGA in plants is gradual. The initial transcriptome of zygote is dominated by egg cell RNA carryover, and although newly expressed genes in the zygote are widespread and represent a significant fraction of the zygote transcriptome, their expression levels are relatively low (Anderson et al., 2017; Chen et al., 2017; Zhao et al., 2019).

Along with dynamic changes in gene expression, epigenomic reprogramming has been observed during plant reproduction. In rice and maize, the egg cell is ~10 times larger than sperm in diameter, and thus ~1000 times larger than the sperm cell in volume (Anderson et al., 2013; Kranz et al., 1991), and its chromatin is diffused (Scholten et al., 2002). In contrast, the sperm cell chromatin undergoes global condensation, paralleling animal sperm chromatin in which protamines replace histones (Kimmins and Sassone-corsi, 2005). Many other sex-specific changes in chromatin occur during plant reproduction, as reported in the model plant *Arabidopsis thaliana* (Wang and Köhler 2017). For example, a male-germline specific histone H3 variant MGH3 (also termed H3.10) is present in the sperm cell (Borg and Berger, 2015; Okada et al., 2005), following the removal of H3.1 (Borg et al., 2020). H3.10 is resistant to trimethylation at

H3K27 (H3K27me3), thus priming the activation of key genes for sperm differentiation and embryogenesis (Borg et al., 2020). Upon karyogamy, H3.10 is removed from the paternal chromatin via a replication independent process (Ingouff et al., 2007). Other histone H3 variants, such as H3.3, are also removed from egg cell chromatin upon karyogamy, followed by loading of newly synthesized histones, again via a replication independent mechanism (Ingouff et al., 2010). In addition, other cells of both male and female gametophytes in *Arabidopsis* experience global chromatin changes as well. Heterochromatin is decondensed in the central cell, the cell which gives rise to endosperm (Pillot et al., 2010). A similar phenomenon occurs in the pollen vegetative cell, the cell which encapsulates the sperm cells and enables their migration through the style to the ovule (Schoft et al., 2009; Mérai et al., 2014; Hsieh et al., 2016). Relaxation of heterochromatin in the pollen vegetative cell has been reported to produce short interfering RNA (siRNA) that traffic into the sperm cells, and reinforce transposon silencing in the gametes (Slotkin et al., 2009; Calarco et al., 2011; Martínez et al., 2016; Park et al., 2016; Kim et al., 2019). Similarly, it has been proposed that siRNAs traffic from the central cell to the egg cell, as well as from the endosperm into the developing embryo (Hsieh et al., 2009; Ibarra et al., 2012; Martínez and Köhler 2017).

Concomitant with chromatin reprogramming, there is also evidence for changes in DNA methylation during plant reproduction, especially in the context of RNA-directed DNA methylation (RdDM) [reviewed in (Gehring, 2019)]. In plants, RdDM can function in both *de novo* and maintenance DNA methylation [reviewed in Cuerda-Gil, and Slotkin (2016)]. Briefly, 24-nt siRNAs are produced and loaded onto an argonaute protein (AGO). This siRNA-AGO complex base-pairs with the nascent transcript of RNA polymerase V (Pol V), using it as a scaffold to recruit the DNA methyltransferase Domains Rearranged Methyltransferase2 (DRM2). DRM2 leads to methylation in all sequence contexts, but methylation in the CHH context (mCHH), where H is A, C or T, is a strong indicator of RdDM in rice and maize (Tan et al., 2016, 2018; Gent et al., 2013), but not in all plants (Zemach et al., 2013). Methylated DNA is recognized by chromatin remodelers, which in turn lead to the deposition of repressive histone marks, such as H3 dimethylated at lysine 9 (H3K9me2). In specific genomic contexts, H3K9me2 recruits RNA polymerase IV (Pol IV), which produces the majority of 24-nt siRNA in plants [reviewed in (Matzke and Mosher, 2014)]. Multiple studies reported that disruption of RdDM leads to a variety of reproductive phenotypes, including aborted embryos (Autran et al., 2011;

Grover et al., 2018), arrested pollen (Wang et al., 2020), defective triploid block when the seeds were produced from a 2n maternal × 4n paternal cross (Borges et al., 2018; Erdmann et al., 2017; Martínez et al., 2018; Satyaki and Gehring, 2019) and defective floral development (Dorweiler et al., 2000; Moritoh et al., 2012). These observations suggest siRNAs and RdDM are important for normal plant reproduction.

In mammals, it has long been proposed that fusion of two epigenetically distinct gametes presents a challenge in reproduction, and resetting of the epigenome is required for the pluripotent state of the early embryo [reviewed in (Messerschmidt et al., 2014)]. Epigenome reprogramming in mammals includes large-scale erasure of somatic chromatin signatures in germ cell precursors, establishment of sex-specific signatures in gametes, and post-fertilization resetting towards pluripotency [reviewed in (Messerschmidt et al., 2014; Saitou et al., 2012; Tang et al., 2016)]. The functional consequences of epigenomic changes in gametic fate acquisition and subsequent zygotic totipotency in plants are unclear. It is clear, however, that in plants the majority of DNA methylation is stably transmitted both maternally and paternally [reviewed in (Gehring, 2019)]. In *C elegans*, siRNAs can serve as carriers of transgenerational epigenetic information, in which siRNAs can be inherited across a few generations [reviewed in (Houri-Zeevi and Rechavi, 2017)]. While multiple changes in siRNA profiles have been observed during plant reproduction (Calarco et al., 2012; Grover et al., 2020; Ibarra et al., 2012; Li et al., 2020; Papareddy et al., 2020; Schoft et al., 2009; Slotkin et al., 2009), transgenerational inheritance of siRNAs, or the lack thereof, has yet to be rigorously demonstrated in plants.

In vegetative tissues, such as seedling, 24-nt siRNAs coincide with mCHH islands, short regions with high CHH methylation, which are enriched around genes and mark the ends of TEs and euchromatin-heterochromatin boundaries (Gent et al., 2013; Li et al., 2015). Hereafter, we refer to such a 24-nt siRNA profile as the canonical siRNA profile, since 24-nt siRNAs are the most abundant length class in most plants including rice (Li et al., 2020, **Fig 1B**). We previously showed that the siRNA transcriptome is reprogramed in rice gametes (Li et al., 2020), where siRNA transcriptomes of egg and sperm were distinct from each other in genome-wide distribution, as well as distinct from that of the seedling (**Fig. 1**). The relative magnitude of the egg-borne and sperm-borne contribution of siRNAs to the zygote and the stage at which the embryo transitions toward a canonical siRNA profile are unknown. Since siRNA production is influenced by histone modifications and DNA methylation, and siRNAs in turn can direct

histone modifications and DNA methylation, the siRNA transcriptome is an output and indicator of the epigenome. Given the likely importance of epigenetic reprogramming during plant reproduction and lack of detailed studies on zygotes, we characterized the small RNA transcriptome of rice zygotes to investigate changes in the small RNA transcriptome that occur soon after fertilization. The differences in genome-wide distribution of siRNAs between gametes and zygote revealed that the siRNA transcriptome initiates a resetting towards the canonical profile before the first cell division, concurrently with zygotic genome activation.

Results

We collected unicellular rice zygotes ~9 hours after pollination (hap), which corresponds to the completion of S-phase, just prior to the first zygotic division (Anderson et al., 2017; Ding et al., 2009). We generated small RNA transcriptomes from 6 replicates, with ~50 zygotes in each replicate. As a maternal sporophytic control, we also collected ovary of the same developmental stage as zygote (9 hap). For post-fertilization ovary (9 hap), we produced small RNA transcriptomes from 3 replicates, with 10 ovaries in each replicate. For our analyses, we also included small RNA transcriptome data from rice gametes, pre-anthesis ovary (0 hr ovary) and seedlings (Li et al., 2020). Except where indicated otherwise, siRNAs used for analyses were small RNA reads (20-nt – 25-nt) not overlapping 90% or more of their lengths with known miRNAs [miRBase v22, (Kozomara et al., 2019)], 5S rRNA, tRNA, NOR, or phasiRNA loci [as detected in (Li et al., 2020)], and multi-mapped reads were included in all analyses unless indicated otherwise. Reproducibility of siRNA and miRNA accumulation patterns was confirmed by clustering of the cell types using principal component analyses (PCA, **Fig S1A - B**).

<u>The global siRNA pattern in zygote is determined by siRNA transcript carryover from the egg cell, with no detectable signature of sperm cell small RNAs</u>

As we previously reported, the sperm cell has an siRNA pattern complementary to the canonical pattern of vegetative tissues, in which its 24-nt siRNA spread out across wide heterochromatic regions, including centromeric tandem repeats. The egg cell and ovary have a pattern different from both, in which 24-nt siRNAs are concentrated at discrete loci (**Fig 1A**). We found that in a whole-genome view, the zygote had a very similar pattern to the egg cell (**Fig**

1A, zygote vs. egg track). This pattern was reproducible across all six zygote replicates, as shown by the tight clustering of zygote replicates on a principal component plot (**Fig S1A**). To confirm that the similarity between zygote and egg cell was not due to large numbers of residual unfertilized egg cells in the zygote samples, we performed a control pollination experiment under similar conditions, and we determined that 98 out of 101 pollinated rice florets produced mature seeds, implying that 3% or less of the rice florets were unfertilized (**Supplemental Table 2**). Thus, in our zygote samples, unfertilized egg cells might represent at most 3% of the total. We also performed differential expression analyses for miRNAs and detected 14 miRNAs that were lowly expressed in all six replicates of zygote but highly expressed in ovaries of the corresponding developmental stage, i.e., 9 hap (**Fig S1B**). Thus, the similarity between zygote and ovary (**Fig 1A**) is unlikely to be due to small RNA contamination from ovary. A similar analysis was previously used to show that the egg cell samples were free of pre-fertilization

ovary contamination as well (Li et al., 2020).

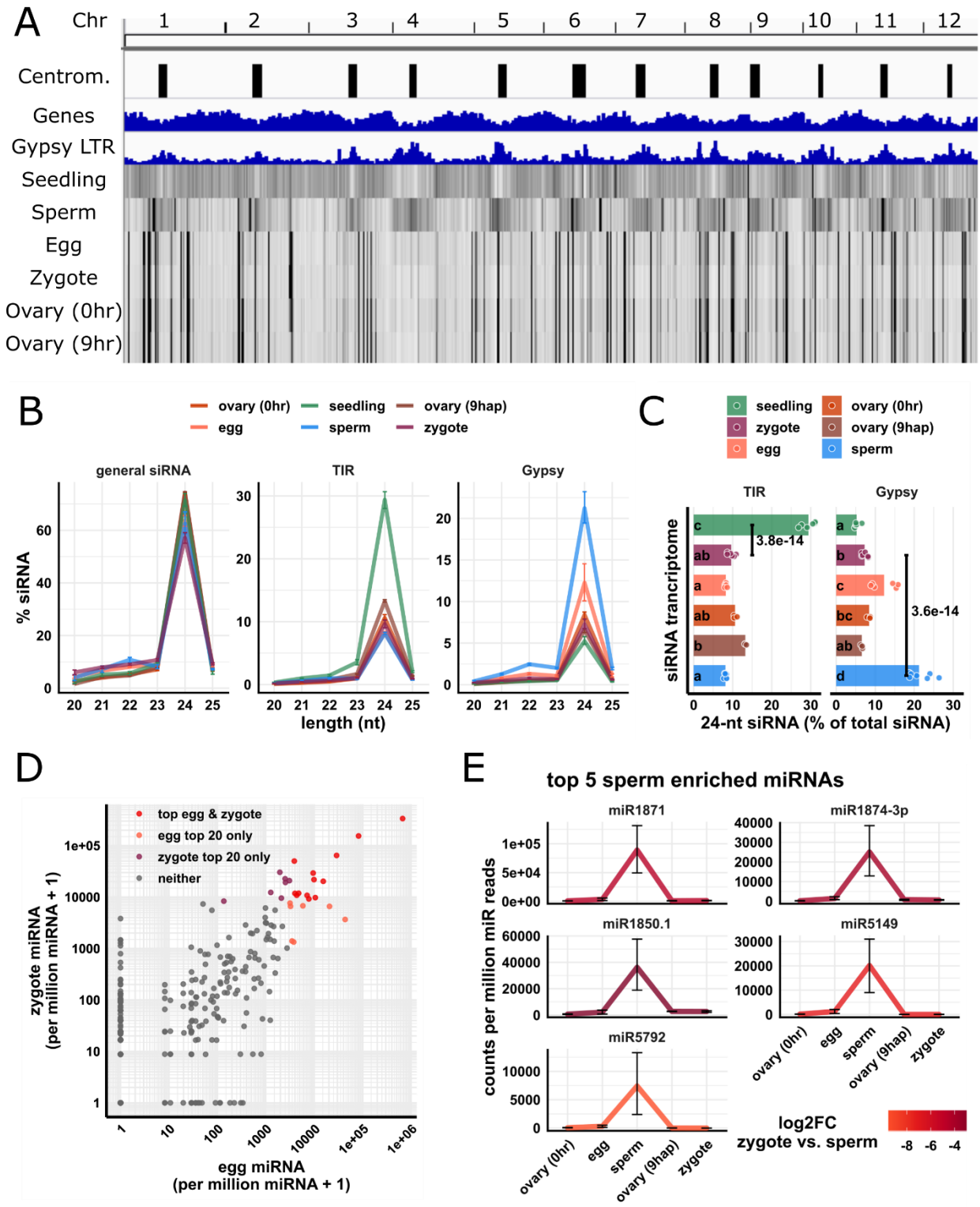


Fig 1: Overall pattern of zygote 24-nt siRNAs is similar to that of the egg cell.

(A) Heat map showing abundance of 24-nt siRNA across genome at 50-kb resolution. Centrom. Centrometric regions, as defined by (Mizuno et al., 2018).

(B) Length profiles of siRNAs. y-axis values are relative to total siRNA reads (20 – 25-nt siRNAs). TIR: terminal inverted repeat transposons, CACTA superfamily excluded. Gypsy: Gypsy retrotransposons. Error bars are 95% confidence intervals for each cell type.

(C) Quantification of TIR and Gypsy panels (B). Each data point is an siRNA transcriptome. Bar heights are averages. Letter grouping (α = 0.05) and P values are based on linear models with logit transformation followed by Tukey tests.

(D) Scatter plot showing miRNA relative abundances in egg and zygote. Each data point is a miRNA. Axes are relative to per million miRNA reads and log10 transformed. 'top egg & zygote' refers to intersection of the 20 highest abundant miRNAs in both egg and zygote.

(E) Top five sperm enriched miRNAs. Sperm enriched is classified as > 1000 reads per million miRNA reads in sperm and < 500 reads per million miRNA reads in egg. y-axis values are relative to per million miRNA reads. Color code reflects log2FC values for zygote vs. sperm, and negative values indicate higher in sperm. Error bars are 95% confidence intervals for each cell type. See **Fig S1D** for additional examples.

Zygote and 9 hap ovary data are from this study, all other data from Li et al., (2020).

We next looked at the length profile of siRNAs in zygotes, and compared that with our recent data from other cell and tissue types (Li et al., 2020). We found that in zygotes, as in all other tissues, 24-nt siRNAs predominated (**Fig 1B**). Since the abundance of siRNAs of other length classes were all relatively low, we focused on 24-nt siRNAs for further analysis. Based on relative abundance patterns, the zygote siRNAs appeared to resemble egg cell siRNAs. Like the egg cell and unlike seedling tissues, the zygote had a low abundance of siRNAs overlapping terminal inverted repeat (TIR) transposons (PIF/Harbinger, Tc1/Mariner, Mutator, or hAT superfamilies) than seedling (**Fig 1B – C**, seedling vs. zygote P = 3.8e-14, Tukey tests). Like the egg cell and unlike the sperm cell, the zygote had a low abundance of siRNAs overlapping *Gypsy* retrotransposons (**Fig 1B – C**, sperm vs. zygote P = 3.6e-14, Tukey tests). The similarity between egg and zygote siRNA profiles can be explained by carryover from the egg cell, since

the egg cell is ~1000-fold larger than the sperm cell by volume (Kranz, Bautor, and Lörz 1991; Anderson et al., 2013; Li et al., 2019). Although 24-nt siRNAs function in the nucleus, 24-nt siRNAs were found primarily in the cytoplasm of whole-plant homogenates (Ye et al., 2012). Thus, we predict that small RNAs already present in the egg cell before fertilization would contribute to much of the siRNAs present in the zygote. This is consistent with previous observations that the 50 most highly expressed genes in egg cell remained as most highly expressed in zygote, whereas the 50 most highly expressed genes in the sperm cell became much lower expressed in the zygote (Anderson et al., 2017, 2013). Indeed, 13 out of the 20 most abundant miRNAs in egg remained among the 20 most abundant miRNAs in zygote (P = 3e-14, Exact test, **Fig 1D**). The remaining 7 of the egg top-twenty miRNAs were still highly abundant in the zygote. However, we noted that the miRNA accumulation patterns were not identical between zygote and egg. 32 miRNAs were detected in the zygote (> 50 reads per million miRNA reads in zygote but undetected in egg cell) but not in the egg cell, and 7 miRNAs were detected in the egg cell but not in the zygote (> 50 reads per million miRNA reads in egg cell but undetected in zygote). The presence of 32 miRNAs detected in zygote but not egg suggests that ZGA is initiated at miRNA loci at this stage, which would be consistent with the known ZGA of other RNA polymerase II transcripts. Meanwhile, top sperm-enriched miRNAs were very much downregulated in the zygote, as expected from dilution after fertilization (**Fig 1E** and **Fig S1D**). Note that the expression values in the zygote were not used to define these sperm-enriched miRNAs, as we classified sperm-enriched miRNAs relative to egg alone. Specifically, we required >1000 reads per million miRNA reads in sperm, but < 500 reads per million miRNA reads in egg. The expression values of the full set of expressed miRNA genes [miRBase v22, (Kozomara et al., 2019)] are provided as a complementary transcriptomics resource (**Supplemental Dataset 1**). These results are again consistent with the prediction that sperm small RNAs were diluted by the egg cell cytoplasm, and that much of the siRNAs detected in the zygote were due to carryover from the egg cell.

<u>Changes in the zygote siRNA transcriptome are independent of changes in the ovary after fertilization, and newly detected siRNA loci in zygote are widespread</u>

Although the siRNA profile in the zygote was similar to the egg cell in terms of overall patterns of abundance, a deeper analysis revealed significant differences from the egg cell. We

produced metagene siRNA coverage plots for seedling, gametes, and zygote, as well as pre- and post-fertilization ovaries (**Fig 2A**). Seedling had a strong peak upstream of the transcription start site (TSS), corresponding to where TIR transposons are enriched in the genome, with the exception of the CACTA superfamily (Han et al., 2013), and such a peak was absent in gametes and ovaries. Zygote had a significant increase in 24-nt siRNA coverage at the peak of the metagene curve relative to egg (**Fig 2A – B**, P = 3e-8, Tukey tests). In contrast, there was no significant change between pre- and post-fertilization ovaries (**Fig 2A – B**, P = 0.98, Tukey tests). Thus, the differences between zygote and egg could not be explain by trafficking from the newly transcribed siRNAs in ovary. To analyze the abundance of siRNAs at individual genomic locus level, we defined siRNA loci from egg, sperm, and seedling using Shortstack (Axtell, 2013). We then classified seedling-specific loci as seedling siRNA loci that did not overlap any egg siRNA loci or sperm siRNA loci (seedling NOT egg NOT sperm loci). Likewise, we classified sperm-specific loci as sperm siRNA loci that did not overlap any egg siRNA loci or seedling siRNA loci (sperm NOT egg NOT seedling siRNA loci). At seedling-specific loci, zygote had more 24-nt siRNAs than egg cell (**Fig 2C**, P = 3e-3, Tukey tests). Since these seedling-specific loci did not overlap any egg siRNA loci or sperm siRNA loci, the increase of 24-nt siRNAs at seedling-specific loci in zygote was unlikely due to carryover from either gamete. At sperm-specific loci, zygote had very few 24-nt siRNAs (**Fig 2C**, zygote vs. sperm P = 0, Tukey tests), much like the results on miRNAs (**Fig 1E**), confirming small RNA contribution from sperm cell is very limited relative to egg. Both egg cell and zygote had few 24-nt siRNAs at sperm-specific loci (**Fig 2C**). However, zygote had even fewer 24-nt siRNAs at sperm-specific loci than the egg cell (**Fig 2C**, zygote vs. egg P = 0.019, Tukey tests). Lastly, there were no differences in the ovaries before and after fertilization for either of these locus categories (**Fig 2C**, P = 0.99 and P = 0.14 at seedling- and sperm-specific loci, respectively). Note that the zygote siRNA transcriptome was not used to define these locus categories. These results indicate that the zygote siRNA transcriptome differs from that of the egg cell, and further, that the changes in the zygote were independent of post-fertilization changes in the ovary.

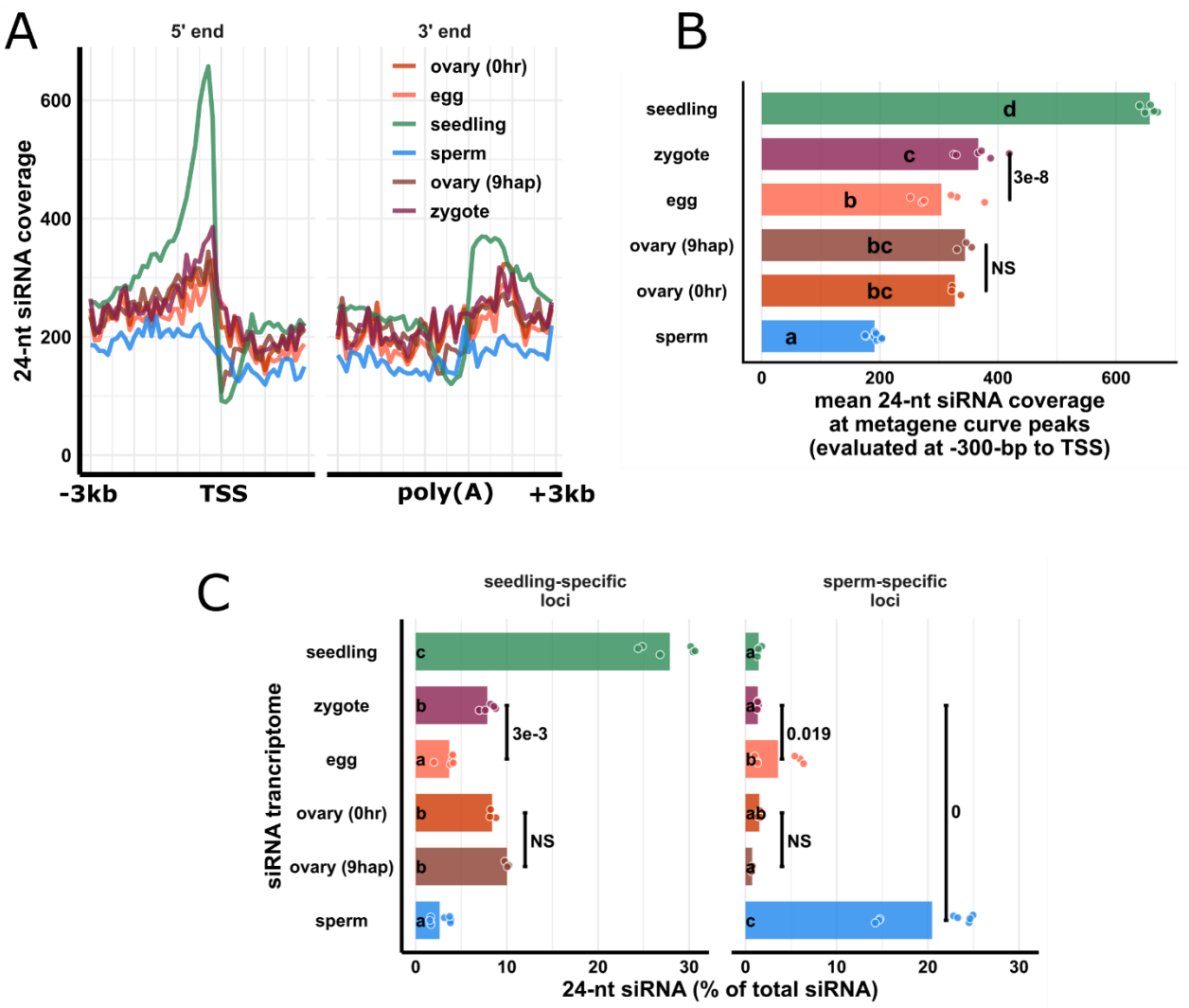


Fig 2: Changes in the zygote siRNA transcriptome are independent from the ovary after fertilization.

(A) Metagene coverage plot for 24-nt siRNAs. Coverage is measured over 100-bp intervals and normalized per 1000 24-nt siRNAs. Vertical grid lines are 500-bp intervals. TSS transcription start site, poly(A) polyadenylation site.

(B) Quantification of (A) at the interval from 300 to 200-bp upstream of TSS, corresponding to the peaks of metagene curves. Each data point is an siRNA transcriptome and bar heights are averages. x-axis values are normalized per 1000 24-nt siRNAs. Letter grouping ($\alpha = 0.05$) and P values are based on Tukey tests.

(C) Bar plot showing relative abundances of 24-nt siRNA across siRNA loci categories. Seedling-specific loci are seedling siRNA loci that do not overlap egg or sperm siRNA loci. Sperm-specific loci are sperm siRNA loci that do not overlap egg or seedling siRNA

loci. The zygote siRNA transcriptome was not used to define these locus categories. Each data point is an siRNA transcriptome. Bar heights are averages. x-axis values are normalized to per 1000 24-nt siRNAs. Letter grouping ($\alpha = 0.05$) and P values are based on linear models with logit transformation followed by Tukey tests.

Zygote and 9 hap ovary data are from this study, all other data from Li et al., (2020).

To further characterize the differences between the zygote siRNA transcriptome and that of the egg cell, we used Shortstack with siRNAs from zygote to define zygote siRNA loci as well. We classified zygote NOT egg loci as zygote siRNA loci that did not overlap any egg cell siRNA loci, egg NOT zygote loci as egg siRNA loci that did not overlap any zygote siRNA loci, and zygote/egg intersect as zygote siRNA loci that overlapped egg siRNA loci. We compared the relative abundances of 24-nt siRNAs at these loci in egg cell and zygote (**Fig 3A**). Most of the 24-nt siRNAs were accounted for by a small number of loci that had high siRNA abundances. Loci with expression values between 1000 reads per million (RPM) and 10,000 RPM accounted for ~50% of the 24-nt siRNAs in all siRNA loci but only accounted for 0.16% of the siRNA loci. This observation suggests siRNA distribution is highly uneven among siRNA loci in egg and zygote, explaining why 24-nt siRNAs in egg and zygote appeared to be concentrated at discreet sites across the genome (**Fig 1A**, see also **Fig 5**). This observation further supports the idea that much of the siRNAs in zygote were carryover from the egg cell, as loci with the most abundant siRNAs in the egg cell continued to be the most abundant in the zygote, much like the results on miRNAs (**Fig 1D**). Related to this, loci with the highest siRNA abundances (RPM > 1000) were highly correlated between zygote and egg ($r = 0.98$, $P = 5e\text{-}53$, correlation test), explaining the similarities between the siRNA profiles of zygote and egg cell (**Fig 1A**).

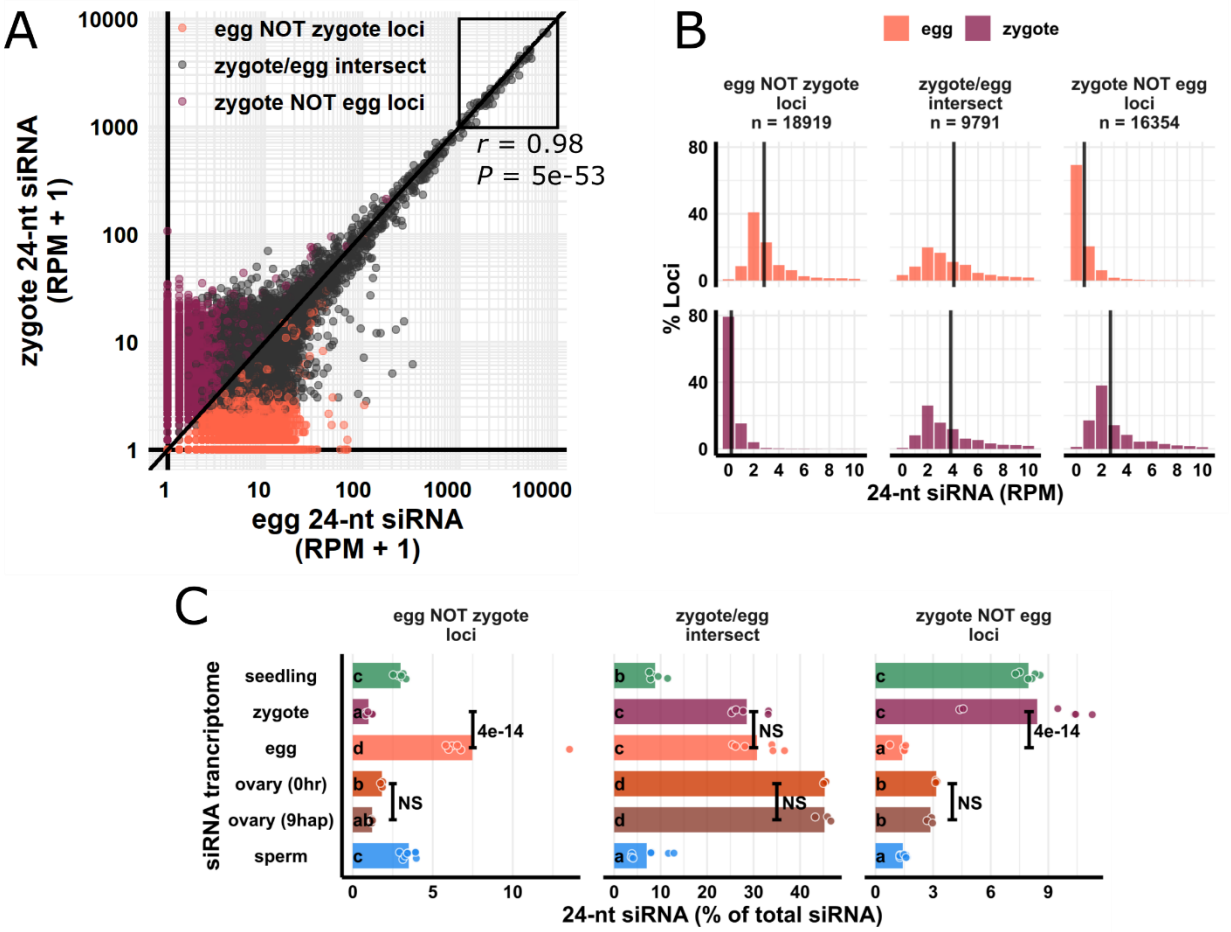


Fig 3: Widespread newly detected siRNA loci in zygote.

(A) Scatter plot showing relative abundances of 24-nt siRNAs across siRNA loci in egg and zygote. Each data point is an siRNA locus. Zygote NOT egg loci are zygote siRNA loci that do not overlap egg siRNA loci. Egg NOT zygote loci are egg siRNA loci that do not overlap zygote siRNA loci. Zygote/egg intersect are zygote siRNA loci that overlap egg siRNA loci. Axes are relative to per million total siRNA reads (20 – 25-nt) and log10 transformed. Correlation coefficient (*r*) and P value are for locus with RPM > 1000.

(B) Quantification of (A). Vertical lines are medians. x-axis values are relative to per million total siRNA reads (20 – 25-nt) and cut off at RPM = 10. Bar heights represent percentage of siRNA loci in each locus category.

(C) Quantification of 24-nt siRNA relative abundances for (A). Each data point is a siRNA transcriptome. Bar heights are averages. x-axis-values are relative to total siRNA reads

(20 – 25-nt). Letter grouping (α = 0.05) and P values are based on linear model with logit transformation followed by Tukey tests.

Zygote and 9 hap ovary data are from this study, all other data from Li et al., (2020).

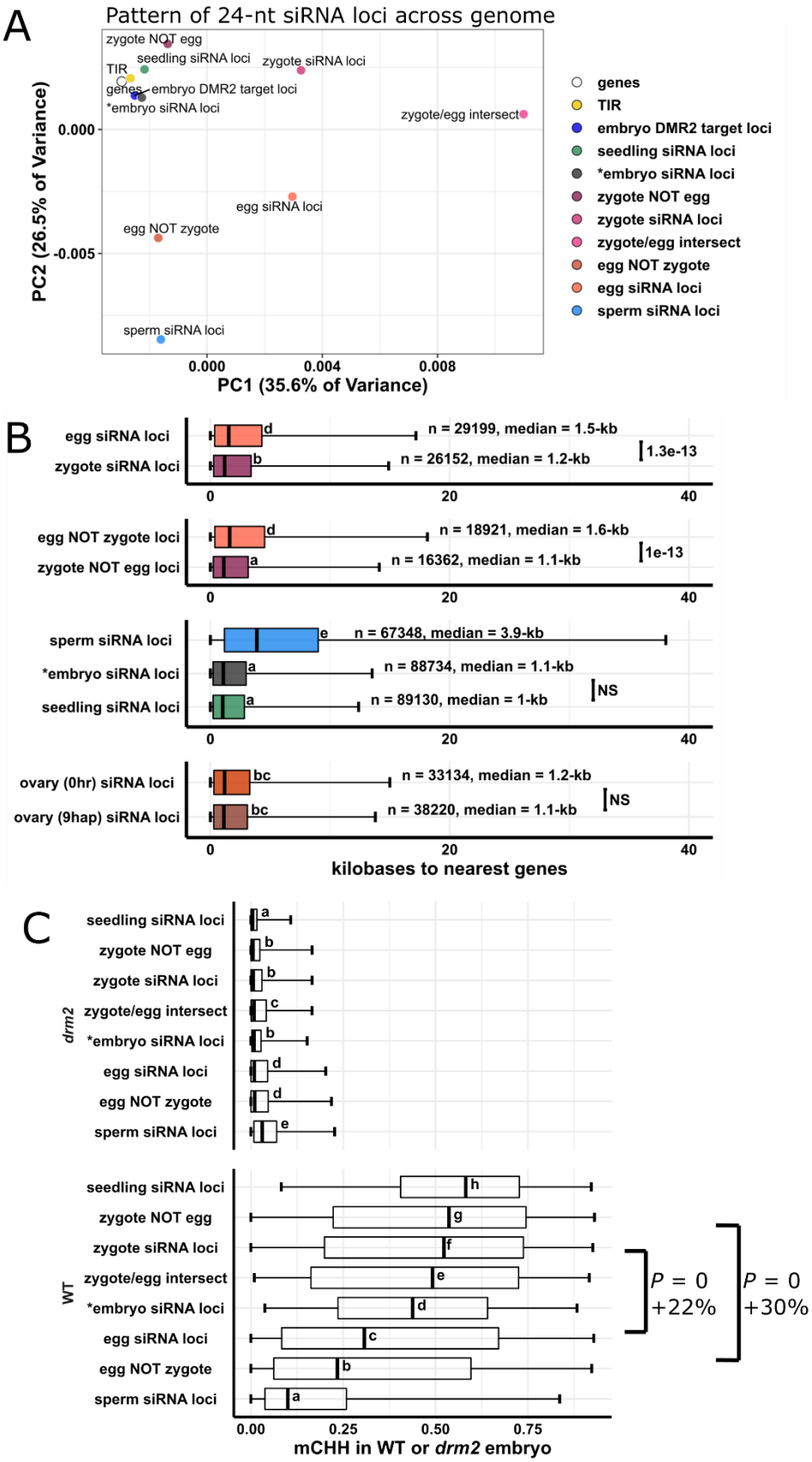
Despite the similarities between egg and zygote at the high abundance siRNA loci, there were widespread distinct siRNA loci that were detected in one cell type but not the other. There were 18919 egg NOT zygote loci, 16354 zygote NOT egg loci, but only 9791 zygote/egg intersect loci. During ZGA of mRNA transcriptomes, genes expressed in zygote but not in egg cell all had initially low expression relative to a background of abundant maternal transcripts (Anderson et al., 2017; Chen et al., 2017; Zhao et al., 2019). Thus, if the siRNA transcriptome is transitioning similarly in the zygote, one would expect to see an initial widespread detection of low abundance 24nt siRNAs at new loci, relative to a background of more abundant maternal carryover siRNAs corresponding to egg siRNA loci. Indeed, in contrast to the high abundance siRNAs of intersect loci, zygote NOT egg and egg NOT zygote loci overall had much lower siRNA abundances than zygote/egg intersect loci (**Fig 3A**).

The validity of our classification method in capturing distinct siRNA loci between zygote and egg was supported by detailed statistical analysis illustrated in **Fig 3B** and **Fig 3C**. We found a clear separation of median RPM between egg and zygote at zygote NOT egg and egg NOT zygote loci, while no separation of medians at zygote/egg intersect (**Fig 3B**). Further, ~80% of egg NOT zygote loci had near 0 siRNAs in zygote (< 1 RPM after merging 6 replicates), and ~70% of zygote NOT egg loci had near 0 siRNAs in egg (< 1 RPM after merging 6 replicates). When all 24-nt siRNA reads at individual loci were tallied and normalized to total siRNAs, as expected, we found that at egg NOT zygote loci, egg had more 24-nt siRNAs than the zygote (**Fig 3C**, P = 4e-14, Tukey tests); at zygote NOT egg loci, zygote had more 24-nt siRNAs than egg (**Fig 3C**, P = 4e-14, Tukey tests); and no difference at zygote/egg intersect. There were no differences between ovaries before and after fertilization in any of the three locus categories (**Fig 3C**), again suggesting changes in the zygote siRNA transcriptome were not coupled with the ovary (**Fig 2**). In addition, the abundance of seedling siRNAs in zygote NOT egg loci and scarcity in egg NOT zygote loci revealed a non-random trend toward a seedling-like siRNA pattern in zygote NOT egg loci (**Fig 3C**). Since the seedling siRNA transcriptome was not used to classify zygote NOT egg loci, this observation suggests that the zygote has initiated a return to

the canonical siRNA profile and is consistent with the increase in 24-nt siRNAs at seedling-specific loci in zygote (**Fig 2C**).

Newly detected siRNA loci in zygote resemble canonical siRNA loci in terms of genomic location and DNA methylation

To investigate the pattern and characteristics of zygote NOT egg loci, the newly detected siRNA loci in zygote, we compared the genomic distribution of zygote NOT egg loci against a set of different siRNA loci categories, including egg NOT zygote loci, total egg siRNA loci, embryo siRNA loci (Rodrigues et al., 2013), seedling siRNA loci, sperm siRNA loci, zygote/egg intersect, and total zygote siRNA loci. The genomic distribution of embryo DRM2 targets were included in this analysis as well. We previously generated a rice *drm2* mutant using CRISPR-Cas9 genome editing, and compared DNA methylation patterns in CHH context between mature wildtype and *drm2* embryos to identify DRM2 targets (Li et al., 2020). As reference points, the distribution of genes and TIR transposons were also included (Kawahara et al., 2013). We used principal component analysis (PCA) to cluster the genomic distributions of the above 11 loci categories based on their abundance in 50-kb windows genome-wide (**Fig 4A**). The distribution of TIR transposons (PIF/Harbinger, Tc1/Mariner, Mutator, or hAT superfamilies) closely clustered with genes, canonical siRNA loci (seedling and embryo), and DRM2 targets, consistent with the prior knowledge that RdDM and TIR transposon are enriched at gene borders in cereal genomes (Gent et al., 2013; Han et al., 2013; Li et al., 2015)**.** The above results suggest that this analysis was successful in capturing the similarities between genomic distributions of multiple locus categories. The genomic distribution of zygote NOT egg loci also clustered with the same loci, but egg NOT zygote loci did not. The probability of observing such a difference by chance over 18919 egg NOT zygote loci and 16354 zygote NOT egg loci was very low ($P = 0$, bootstrapping with 1000 iterations). Total zygote siRNA loci were closer to the canonical siRNA loci patterns than total egg siRNA loci were. These observations indicate that the newly detected siRNA loci are similar to canonical siRNA loci in terms of their genomic distribution.


A
Pattern of 24-nt siRNA loci across genome
zygote NOT egg
seedling siRNA loci
zygote siRNA loci
TIR
genes
embryo DMR2 target loci
*embryo siRNA loci
zygote/egg intersect
egg siRNA loci
egg NOT zygote
sperm siRNA loci
PC2 (26.5% of Variance)
0.000
-0.005
0.000
0.004
0.008
PC1 (35.6% of Variance)
genes
TIR
embryo DMR2 target loci
seedling siRNA loci
*embryo siRNA loci
zygote NOT egg
zygote siRNA loci
zygote/egg intersect
egg NOT zygote
egg siRNA loci
sperm siRNA loci
B
egg siRNA loci
zygote siRNA loci
n = 29199, median = 1.5-kb
n = 26152, median = 1.2-kb
1.3e-13
egg NOT zygote loci
zygote NOT egg loci
n = 18921, median = 1.6-kb
n = 16362, median = 1.1-kb
1e-13
sperm siRNA loci
*embryo siRNA loci
seedling siRNA loci
n = 67348, median = 3.9-kb
n = 88734, median = 1.1-kb
n = 89130, median = 1-kb
NS
ovary (0hr) siRNA loci
ovary (9hap) siRNA loci
n = 33134, median = 1.2-kb
n = 38220, median = 1.1-kb
NS
0
20
40
kilobases to nearest genes
C
drm2
seedling siRNA loci
zygote NOT egg
zygote siRNA loci
zygote/egg intersect
*embryo siRNA loci
egg siRNA loci
egg NOT zygote
sperm siRNA loci
WT
P = 0
+22%
P = 0
+30%
0.00
0.25
0.50
0.75
mCHH in WT or drm2 embryo

367

Fig 4: Newly detected siRNA loci in zygote reset to the canonical siRNA profile and predict CHH methylation in embryo in an RdDM-dependent manner.

(A) Principal component analysis of siRNA loci distributions across genome. Locus distribution are evaluated with 50-kb resolution. Each data point represents the genomic distribution of a siRNA loci category.

(B) Boxplots showing distance of siRNA loci to nearest genes. Middle lines are median. Boxes span interquartile range. Whiskers span 2.5th and 97.5th percentiles. Letter groupings ($\alpha = 0.05$) and P values are based on linear model with log10(distance + 1-bp) followed by Tukey tests.

(C) Boxplots showing CHH methylation level in mature wildtype and *drm2* mutant embryos. Middle lines are median. Boxes span interquartile range. Whiskers span 2.5th and 97.5th percentiles. Letter groupings ($\alpha = 0.05$) and P values are based on generalized linear model with logit link followed by Tukey tests. Zygote NOT egg loci n = 16354, zygote/egg intersect n = 9791, egg NOT zygote loci n = 18919.

*Embryo siRNA data from Rodrigues et al., (2013), which was based on a single replicate. Seedling, gametes and pre-fertilization ovary data from Li et al., (2020).

Genomic distributions of siRNA loci, as quantified by distance to nearest genes, also revealed that a resetting of the siRNA loci pattern has been initiated in the zygote, and before the first embryonic division (**Fig 4B**). Half of seedling siRNA loci were within 1-kb to their nearest genes (**Fig 4B**). This was also the case for zygote NOT egg loci (seedling siRNA loci vs zygote NOT egg loci P = 1, Tukey tests), where half of them were within 1.1-kb to their nearest genes (**Fig 4B**) and are significantly closer to genes than egg NOT zygote loci (P = 1e-13, Tukey tests). Total zygote siRNA loci were closer to genes than total egg siRNA loci (P = 1.3e-13, Tukey tests), again suggesting the gene-proximal property of zygote NOT egg loci was not due to potential technical artifacts regarding how zygote NOT egg or zygote NOT zygote loci were classified. From egg cell to zygote, there was a 20% decrease in median distance (1.5-kb vs. 1.2-kb), which occurred over the course of less than one cell cycle (8 – 9 hrs after fertilization). In contrast, from zygote to embryo (7 days after fertilization, data from (Rodrigues et al., 2013), there was a 8% decrease (1.2-kb vs. 1.1-kb), and such a change occurred over days, over the course of numerous cell cycles. These observations are consistent with the results where zygote

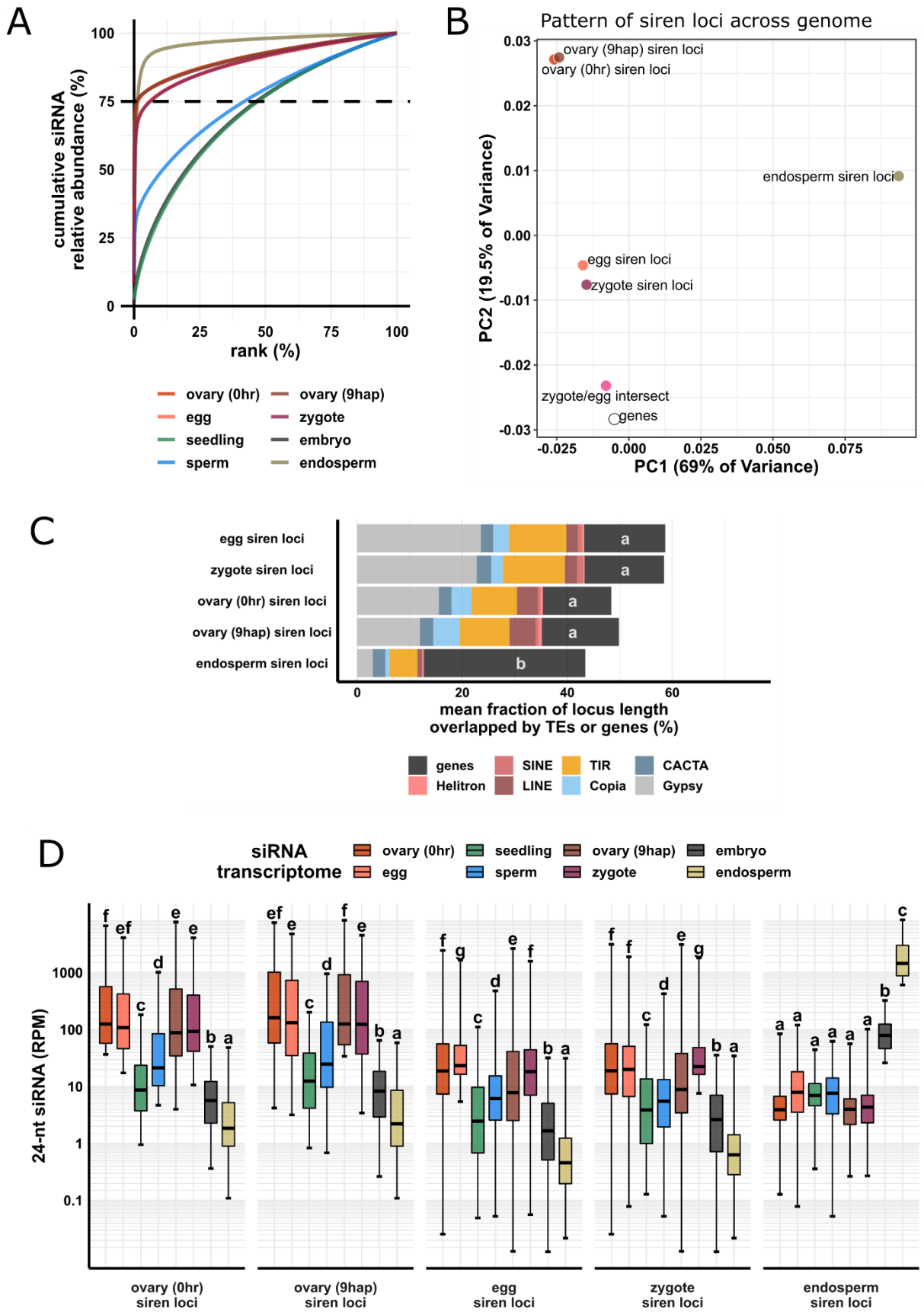
had more gene-proximal 24-nt siRNAs (**Fig 2A – B**), zygote had increased siRNAs in seedling-specific loci (**Fig 2C**), and seedling had comparable siRNA level with zygote in zygote NOT egg loci (**Fig 3C**).

Resetting to the canonical siRNA pattern might suggest that the newly detected siRNA loci in the zygote are targeted for CHH methylation during embryogenesis in an RdDM-dependent manner. We compared DNA methylation levels across different siRNA loci categories in mature wildtype and *drm2* embryos (**Fig 4C**, see also **Fig S2**). In wildtype embryos, zygote NOT egg loci had high mCHH levels, closely tailing the median mCHH levels at seedling siRNA loci. In contrast, egg NOT zygote loci had much lower mCHH levels (**Fig 4C**, median 54 vs. 23 mCHHs per 100 CHH). The probability of observing such a difference by chance over 18919 egg NOT zygote loci and 16354 zygote NOT egg loci was very low ($P = 0$, Tukey tests). Total zygote siRNA loci had higher mCHH levels than total egg siRNA loci (**Fig 4C**, median 52 vs. 30 mCHHs per 100 CHH, $P = 0$). Meanwhile, in *drm2* embryo, the mCHH levels for both zygote NOT egg loci and egg NOT zygote loci were low **(Fig 4C**). The magnitudes of any differences among siRNA loci categories in *drm2* embryos were also much lower. In addition, while there were less differences in mCG and mCHG across locus categories in wildtype embryo, the full extent of CG and CHG methylation at seedling siRNA loci and zygote NOT egg loci also depended on DRM2 (**Fig S2**). Taken together, these results suggest not only did zygote NOT egg loci reset to the canonical siRNA pattern during embryogenesis, but also that this pattern of siRNA distribution is functional in targeting CHH methylation, and to a lesser extent CG and CHG methylation, in a DRM2-dependent manner.

<u>Unusual features of siRNA loci with abundant siRNAs in egg, ovary, zygote and endosperm</u>

Zygote, like egg cell and ovary, had a very uneven siRNA distribution across the genome, where siRNAs appeared to be concentrated at discrete sites, without a clear relationship to gene density (**Fig 1A**). It has been previously reported that rice developing endosperm (7-8 days after fertilization) has a unique siRNA profile in which a small number of loci accounted for the majority of siRNAs (Rodrigues et al., 2013). These siRNA loci were termed siren loci (<u>siR</u>NA in the <u>en</u>dosperm). A similar phenomenon was recently reported in *Brassica rapa* and *Arabidopsis* ovules and endosperm (Grover et al., 2020). The term ‘siren loci’ was also used by Grover et al to describe these loci. To further investigate this phenomenon in zygote as well as

egg, ovary, and endosperm, we ranked siRNA loci according to siRNA abundance in each cell type (**Fig 5A**). In endosperm and ovaries (pre- and post-fertilization), ~1% (*n* = 119, 483 and 358, respectively) of the siRNA loci accounted for 75% of the total siRNA accumulation in all siRNA loci for each tissue type (**Fig 5A**). Similarly, in egg cell and zygote, ~6% (*n* = 1750 and 1622, respectively) of the siRNA loci accounted for 75% of the total siRNA accumulation in all siRNA loci for each cell type (**Fig 5A**). We call these loci that account for 75% of the siRNAs siren loci, independently of siRNAs in endosperm. In fact, the siren loci in rice ovaries, egg, and zygote showed little correlation with the siren loci in rice endosperm, at least at the specific endosperm stage examined, 7-8 days after fertilization (Li et al., 2020; Rodrigues et al., 2013). These observations explain why loci with the highest siRNA abundances accounted for much of the siRNAs in egg and zygote (**Fig 3A**), and the differences between 24-nt siRNA distribution (**Fig 1D**, **Fig S1A**) and siRNA loci distribution (**Fig 4A**).


A
cumulative siRNA relative abundance (%)
rank (%)
ovary (0hr)
ovary (9hap)
egg
zygote
seedling
embryo
sperm
endosperm
B
Pattern of siren loci across genome
ovary (9hap) siren loci
ovary (0hr) siren loci
endosperm siren loci
egg siren loci
zygote siren loci
zygote/egg intersect
genes
PC2 (19.5% of Variance)
PC1 (69% of Variance)
C
egg siren loci
zygote siren loci
ovary (0hr) siren loci
ovary (9hap) siren loci
endosperm siren loci
mean fraction of locus length overlapped by TEs or genes (%)
genes
SINE
TIR
CACTA
Helitron
LINE
Copia
Gypsy
D
siRNA transcriptome
ovary (0hr)
seedling
ovary (9hap)
embryo
egg
sperm
zygote
endosperm
24-nt siRNA (RPM)
ovary (0hr) siren loci
ovary (9hap) siren loci
egg siren loci
zygote siren loci
endosperm siren loci

442

Fig 5: zygote siren loci are similar to siren loci detected in ovary and egg cell, but dissimilar to siren loci detected in endosperm.

(A) x-axis is the rank order of siRNA loci. siRNA loci with highest siRNA abundances are ranked first. y-axis is cumulative relative abundance of siRNA in all siRNA loci. Axis values are scaled between 0 and 100%. 1% of siRNA loci accounted for 75% of siRNA reads in all siRNA loci in endosperm and ovary. 6% of siRNA loci accounted for 75% of siRNA reads in all siRNA loci in egg and zygote.

(B) Principal component plot siren loci distribution across the genome. Distributions are evaluated at 50-kb resolution across the genome. Each data point is the distribution of a siren loci category.

(C) Stacked bar plots showing mean fraction of locus length overlapped by TEs or genes. Letter grouping ($\alpha = 0.05$) is based generalized linear model with logit link followed by Tukey tests.

(D) Boxplots showing 24-nt siRNA relative abundances across siren classes across cell types. Middle lines are median. Boxes span interquartile range. y-axis values are relative to per million total 24-nt siRNA in each siRNA transcriptome. Whiskers span 2.5^{th} and 97.5^{th} percentiles. Letter groupings ($\alpha = 0.05$) are based on linear model with log(RPM + 1) followed by Tukey tests.

Embryo and endosperm data from Rodrigues et al., (2013). Seedling, gametes and pre-fertilization ovary data from Li et al., (2020).

Next, we compared the similarity among different siren loci categories based on their genomic distributions. We used principal component analysis (PCA) to cluster the genomic distributions of the loci categories based on their abundance in 50-kb windows genome-wide (**Fig 5B**). For comparison with **Fig 4A**, we also included genes and zygote/egg intersect loci, which were furthest separated along PC1 in **Fig 4A**. We found that the five siren loci categories (endosperm siren loci, ovary (0 hr) siren loci, ovary (9 hap) siren loci, egg siren loci and zygote siren loci) can be classified into two larger classes: the endosperm siren class and ovary/egg/zygote siren class (**Fig 5B**). On the PC plot, endosperm siren loci were well separated from all the others along PC1, which accounts for 69% of the variance in their genomic distributions. The rest of the siren loci categories were separated along PC2, which accounts for

19.5% of the variance, much less than what was explained by PC1. Ovary siren loci categories (pre- and post-fertilization) had nearly identical genomic distributions, clustering closely together (**Fig 5B**). Egg and zygote siren loci also had nearly the same genomic distribution, clustering closely together (**Fig 5B**). Consistent with its unique genomic distribution, endosperm siren loci were more likely to overlap a gene (**Fig 5C**). On average, ~30% of the locus length of an endosperm siren locus was covered by a gene, while all the other siren categories have similar fraction of locus length covered by genes (~15%, $P < 1.4e\text{-}4$, Tukey tests). Lastly, we compared the relative abundances of 24-nt siRNAs at different siren categories across different cell types. Based on the pattern of 24-nt siRNA accumulation, the five types of siren loci can again be divided into two larger siren classes: the endosperm siren class and ovary/egg/zygote siren class (**Fig 5D**). At endosperm siren loci, endosperm had the highest 24-nt siRNA expression, ~10-fold higher than the level in embryo and more than 100-fold than the levels in all other cell types we examined (**Fig 5D**). In contrast, the ovary/egg/zygote siren class shared a siRNA accumulation pattern across cell types (**Fig 5D**). Ovaries (pre- and post-fertilization), egg cell and zygote all have high abundances of 24-nt siRNAs at ovary/egg/zygote siren loci, while seedling, sperm, embryo and endosperm all had low abundances of 24-nt siRNAs at these siren loci. Taken together, these distinct siRNA accumulation patterns explain why in zygote, like egg and ovary, siRNAs were concentrated at discrete sites (**Fig 1A**).

The fact that siren loci were defined by abundant 24-nt siRNAs led us to speculate that they would have high mCHH. In fact, we found they had lower mCHH levels relative to other siRNA loci in each tissue (**Fig S3**). In ovary, canonical siRNA loci (seedling siRNA loci) and ovary siRNA loci that were not siren loci had the highest mCHH levels, while ovary siren loci had lower mCHH level (**Fig S3A**). Similar results were found for egg cell and embryo as well (**Fig S3B – C**). At siren loci as well as siRNA loci that were not siren loci, DMR2 was required for mCHH in embryo. Lastly, both wildtype and *drm2* endosperm had overall low mCHH methylation, and endosperm siren loci did not correspond to high mCHH level in the endosperm (**Fig S3D**). Unlike mCHH, mCG and mCHG did not produce any notable pattern across cell types (**Fig S3**). Taken together, these results suggest that although the highly abundant siRNAs produced by siren loci may target DNA methylation in a DRM2-dependent manner, they are not as efficient as siRNAs from 24nt siRNA loci that were not siren loci.

Discussion

We report here the first small RNA transcriptome characterization of plant zygotes. Overall, the small RNA transcriptome of the zygote is similar to that of the egg cell (**Fig 1A – D**), which can be explained by transcript carryover from the egg cell, and the dilution of sperm cell small RNAs. This is supported by our finding that the most abundant miRNAs in egg continued to be the most abundant in zygote (**Fig 1D**), and siRNA loci with the most abundant siRNAs in egg continue to be most abundant in zygote (**Fig 3A**). In contrast, top sperm-enriched miRNAs became orders of magnitudes less abundant in zygote (**Fig 1E** and **Fig S1D**), and zygote had very few 24-nt siRNAs at sperm-specific siRNA loci **(Fig 2C**). It has been proposed that sperm-transmitted siRNAs regulate TEs and balance parental contribution in the endosperm, as RdDM-mutants affect endosperms from 2n maternal × 4n paternal crosses (Borges et al., 2018; Erdmann et al., 2017; Martinez et al., 2018; Satyaki and Gehring, 2019). Our data indicate that any effects of sperm-transmitted siRNAs on embryos are likely to be indirect. Assuming sperm-derived siRNAs are also diluted by the larger central cell, we speculate that the effect of the sperm-transmitted siRNAs may act through sperm histone modifications and not siRNAs themselves. In the newly formed endosperm, there is lack of an active replacement of histone variants, and sperm-derived histone variants are passively diluted through nuclear divisions (Ingouff et al., 2007, 2010). In contrast, in the zygote, histone variants are actively replaced in a replication-independent manner before the first embryonic division (Ingouff et al., 2007, 2010).

We detected widespread new zygote siRNA loci relative to the egg cell, representing ~63% of all zygote siRNA loci: There were 16354 zygote NOT egg loci, as compared to 9791 zygote/egg intersect siRNA loci (**Fig 3B**). In relative abundance, most of the siRNA reads were accounted for by egg siRNA carryover (**Fig 3A**), and thus siRNA abundance was lower at zygote NOT egg loci than at zygote/egg intersect (**Fig 3A**). This low relative abundance can be understood in the context of the zygotic transition, which involves a new genomic program initiated within that one cell, so that production of new siRNAs will be occurring against the backdrop of egg cell RNA carryover. Similar observations have been made for zygote mRNA transcriptomes from multiple independent laboratories from different plant species (Anderson et al., 2017; Chen et al., 2017; Zhao et al., 2019), where zygote *de novo* expressed genes, including those with key functions in embryogenesis, were lowly expressed in the zygote.

Several lines of evidence indicate the zygote has initiated a resetting towards the canonical siRNA pattern, and that such resetting is independent from the ovary. First, the zygote had increased 24-nt siRNAs at seedling-specific loci (**Fig 2C**), as compared to the egg cell. In contrast, there were no significant changes to the 24nt siRNAs at seedling loci in ovaries pre- and post-fertilization (**Fig 2C**). Moreover, seedling had comparable 24-nt siRNAs to zygote at zygote NOT egg loci (**Fig 3C**). As the zygote siRNA transcriptome was not used to define seedling-specific loci, and the seedling siRNA transcriptome was not used to define zygote NOT egg loci, these results serve as an objective indication that the zygote shifted towards a more seedling-like siRNA transcriptome. Second, zygote had increases relative to the egg cell in 24-nt siRNAs at the TSS region upstream of genes, while there was lack of a corresponding change in the ovary after fertilization (**Fig 2A – B**). High 24-nt siRNA coverage upstream of genes around the TSS is a feature of a canonical siRNA transcriptome, as exemplified by seedling (**Fig 2A**). Third, the genomic distribution of zygote NOT egg loci clustered closely with those of TIR, seedling siRNA loci, embryo siRNA loci and embryo DRM2 targets, while the genomic distribution of egg NOT zygote loci did not (**Fig 4A**). Consistent with the major contribution of the zygote NOT egg siRNA loci to the zygote siRNA distribution, the total set of zygote siRNA loci also displayed a closer relationship to the canonical siRNA distribution than did the total egg siRNA loci (**Fig 4A**). Fourth, the similarities in genomic distribution were confirmed by distances to the nearest genes (**Fig 4B**). Notably, there was a 20% decrease in median distance to genes from egg to zygote, which took place over the course of less than one cell cycle, and only an 8% decrease from zygote to embryo (7 days after fertilization), which occur over the course of many cell cycles. These results indicate a resetting was initiated in the zygote before the first embryogenic division. Lastly and importantly, newly detected siRNA loci have abundant CHH methylation in embryos that is dependent on the RdDM methyltransferase DRM2 (**Fig 4C**). Like canonical siRNA loci (seedling siRNA loci), zygote NOT egg loci had high mCHH level, whereas egg NOT zygote loci had much lower mCHH level in mature wildtype embryo. In *drm2* embryo, mCHH levels were low at both zygote NOT egg and egg NOT zygote loci (**Fig 4C**). These results support newly detected siRNA loci in zygote not only reset to canonical siRNA pattern, but also that the corresponding 24-nt siRNAs function in targeting CHH methylation during embryogenesis.

Similar to *Brassica rapa* and *Arabidopsis* ovules (Grover et al., 2020), a small number of loci accounted for most of the siRNAs in zygote (**Fig 5A**), and these loci were termed siren loci. Siren loci were first discovered in rice endosperm (Rodrigues et al., 2013); however, siren loci in the zygote was distinct from endosperm siren loci in endosperm collected 7-8 days after fertilization, but instead coincided with siren loci detected in ovary and egg cell (**Fig 5B – D**). However, it remains possible that the central cell and earlier stages of endosperm have an siRNA transcriptome more like that of the zygote. It has been proposed that the embryo receives siRNAs from the endosperm (Hsieh et al., 2009; Martínez and Köhler, 2017). This does not appear to be the case in 7-8 day rice seeds, since rice embryos had low siRNA abundance at endosperm siren loci at this stage (Rodrigues et al., 2013). A recent publication demonstrated that trans-acting siRNAs from ARFs (tasiR-ARF) traffic across ovule cell layers to regulate megaspore mother cell (MMC) identity in *Arabidopsis* (Su et al., 2020). It has also been proposed that siRNAs may traffic from the seed coat into the embryo during seed development (Grover et al., 2020, 2018). Likewise, it is possible that siren siRNAs in the egg cell and zygotes are produced in the ovary tissue instead. While siRNAs at siren loci may direct some CHH methylation in ovary or during embryogenesis, they appeared to have lower mCHH level than their non-siren siRNA loci counterparts (**Fig S3**). This raises the possibility that the 24-nt siRNAs of siren loci function in chromatin modification or post transcriptional silencing independently of DNA methylation in the zygote or egg, regardless of their cell type of origin.

Due to the extreme difficulties associated with zygote isolation and low input sequencing, epigenome profiling for plant zygotes has been challenging. Plant gametes are highly dimorphic in terms of size, chromatin (Wang and Köhler 2017; Borg and Berger 2015; Ingouff et al. 2010), and gene expression (Anderson et al., 2013), consistent with a differential reprogramming of gamete epigenomes prior to fertilization inferred from their siRNA profiles (Li et al., 2020). However, changes in the gametes must be followed by resetting to the canonical somatic profile during the next generation. Here, using the siRNA transcriptome of rice zygotes, we infer that the resetting to canonical siRNA transcriptome is initiated along with the zygotic transition, soon after fertilization, and sets the stage for the methylation pattern of the embryo. This conclusion is consistent with previous observations in *Arabidopsis* that replacement of H3 variants occurs in the zygote before the first cell division (Ingouff et al., 2007; 2010). Lastly, as siRNA expression is influenced by chromatin structure, and siRNAs can either reinforce or initiate DNA

methylation and histone modifications, the siRNA transcriptome is an indicator and output of the epigenome. Thus, it is likely that resetting of the gametic epigenomes for the zygotic transition, including features such as histone modifications and chromatin conformation, is also initiated in the unicellular zygote in plants.

Methods

Plant growth condition and zygote collection

Rice (*Oryza sativa*) variety *Kitaake* was grown in soil in greenhouse under natural light condition. Zygote isolation was performed as described (Anderson et al., 2017; Li et al., 2019). Briefly, rice flowers were hand pollinated. At eight to nine hours post pollination, ovaries were dissected. A transverse cut was made at the middle region of the ovary in a droplet of 0.3M mannitol. The lower part of the cut ovary was gently pushed by an acupuncture needle under a phase contrast inverted microscope. Once the zygote floated out of the ovary incision, it was captured by a fine glass capillary and immediately frozen in liquid nitrogen. 50 zygotes were collected for each replicate, and six replicates were collected. Intact ovaries at 8-9 hours after pollination were collected separately for the ovary small RNA analysis. 10 ovaries were collected for each replicate, and three replicates were collected (**Supplemental Table 1**).

RNA extraction and small RNA library construction

RNA extractions were performed using Ambion RNAqueous Total RNA kit (AM1931), including an on-column DNase I treatment using Qiagen DNase I (79254). Total RNA was run on a Bioanalyzer (Agilent) to check for RNA integrity, using the eukaryotic total RNA-pico program. RNA input for library construction was ~30ng. Small RNA libraries were made using the NEXTflex small RNA-seq kit v3 (PerkinElmer NOVA-5132-05), with the following modifications. ¼ dilution of adapters was used. The 3’ adapter ligation step was done at 20°C overnight. Zygote libraries were amplified at 24 cycles. Post-fertilization ovary libraries were amplified at 20 cycles, as pre-fertilization ovaries (Li et al., 2020). The library product was size selected using PippinHT (Sage Science) 3% agarose gel cassettes.

Small RNA sequencing analysis

Analyses were based on the Os-Nipponbare-Reference-IRGSP-1.0 reference genome (Kawahara et al., 2013b). Genome annotations for transposable elements, genes, miRNAs, 5S rRNA, tRNA, NOR, CentO repeats and phasiRNA loci were performed as described (Li et al. 2020). Quality filtering, adapter trimming, PCR duplicate removal and alignment were performed as described (Li et al. 2020). Small RNA-seq reads were quality filtered and trimmed of adapters using cutadapt (Martin, 2011), parameters "-q 20 -a TGGAATTCTCGGGTGCCAAGG -e .05 -O 5 --discard-untrimmed -m 28 -M 33". PCR duplicates were then removed using PRINSEQ, parameters "prinseq-lite.pl -fastq out_format 3 -out_good -derep 1" (Schmieder and Edwards, 2011). The four random nucleotides at each end were then removed using cutadapt "-u4" followed by cutadapt "-u -4". Reads were aligned to the genome with BWA-backtrack (version 0.7.15) (Li and Durbin 2009), parameters "aln -t 8 -l 10." Except where indicated otherwise, multi-mapping reads were included in all analyses. The uniquely mapping subset of siRNAs was defined by having MAPQ values of at least 20 using SAMtools (Li et al. 2009). Except where indicated otherwise, siRNAs used for analyses were small RNA reads (20 – 25-nt) not overlapping 90% or more of their lengths with miRNA, 5S rRNA, tRNA, NOR and phasiRNA loci as determined by the BEDTools intersection tool (Quinlan and Hall, 2010). For analysis of overlaps of siRNAs and *Gypsy* retrotransposons, the CentO centromeric tandem repeat, Terminal Inverted Repeat (TIR) DNA transposons, and 24-nt siRNA loci, only siRNAs that overlapped by at least 50% of their lengths were counted. CACTA elements were excluded from the TIR DNA transposons. Distances to closest genes were obtained using the BEDTools closest tool. Whole-genome small RNA heatmaps were made on 50-kb intervals using IGVtools (Thorvaldsdottir et al., 2013). For better visualization of midrange values, heatmap intensity was maxed out at 1.25X coverage (per 10 million 24-nt siRNAs).

<u>miRNA analysis</u>

To measure miRNA accumulation, the BEDTools coverage tool was used to count the number of 20 – 25-nt reads that overlapped at least 90% of their length with annotated miRNA positions (**Supplemental Dataset 1)**. R package EdgeR was used to analyze miRNA accumulation (McCarthy et al., 2012). Individual miRNA counts were normalized by total mapped small RNAs and filtered for >1 counts per million reads (CPM) in at least three libraries.

Differential expression analyses were performed under |log2FC| > 1 and FDR < 0.05 cutoffs. Differential expressing miRNA genes were visualized under counts per million miRNAs. Principal component analyses were performed using log-transformed CPM values.

Definition of siRNA loci

Small RNA loci were identified from the initial 20 – 25-nt total small RNA alignment bam files using Shortstack (Axtell, 2013) after merging replicates using default parameters. For each tissue type (pre- and post-fertilization ovary, egg cell, sperm cell, zygote, seedling, embryo and endosperm), siRNA loci were defined as RPM > 2, 24-nt-dominant and not detected as a miRNA locus ("DicerCall=24; MIRNA=N"). Endosperm siren loci were defined as the highest expressing loci that accounted for 75% of the cumulative RPM in the endosperm. Similarly, pre- and post-fertilization ovary siren loci as well as egg and zygote siren loci were defined as the highest expressing loci that accounted for 75% of the cumulative RPM in the ovary. The 75% cutoff was selected based on the turning point of cumulative expression vs. percentage rank plot of ovary (**Fig 5A**). Seedling-specific loci were identified as seedling siRNA loci that did not overlap any sperm siRNA loci or egg siRNA loci (seedling NOT egg NOT sperm loci) using the BEDTools intersect tool (Quinlan and Hall, 2010). Similarly, sperm-specific loci were identified as sperm siRNA loci that did not overlap any egg siRNA loci or sperm siRNA loci (sperm NOT egg NOT seedling loci). Zygote NOT egg loci were zygote siRNA loci that did not overlap egg siRNA loci. Egg NOT zygote loci were egg siRNA loci that did not overlap zygote siRNA loci. Zygote/egg intersect were zygote siRNA loci that overlapped egg siRNA loci.

DNA methylation analyses

Methylation values were calculated for each locus using the mtr function of CGmapTools v0.1.2 (Guo et al., 2018) using the CGmap files generated in our previous study as input (Li et al., 2020). Only loci with more than 3 (mC + C) calls were included in the analyses.

Statistical analyses

Tukey tests were performed using the R package emmeans (Searle et al., 1980) with multiple comparison correction using Tukey's method. Letter groupings were done at $\alpha = 0.05$, where the differences between means sharing the same letter were not statistically significant.

For multifactorial analyses, multiple comparisons were applied to families of tests at each interacting factor level: at the level of each TE/locus category for **Fig 1C**, **Fig 2C**, **Fig 3C**, and **Fig 5D**, and at the level of genotype for **Fig 4C**, **Fig S2** and **Fig S3**. For analyses of siRNA relative abundances across siRNA locus category across siRNA transcriptomes, a linear model was fitted using logit transformation to correct for heteroscedasticity. For analyses of siRNA relative abundances across individual siRNA locus, a linear model was fitted using log(RPM + 1) transformation to correct for heteroscedasticity. For analyses of distances to nearest genes, a linear model was fitted using log(distance + 1-bp) transformation to correct for heteroscedasticity. For analyses of DNA methylation levels across different locus categories, a generalized linear model of quasibinomial family with logit link was fitted to accommodate the mean-error relationship of proportion data. For analyses of fraction of locus length covered by genes, a generalized linear model of quasibinomial family with logit link was fitted to accommodate the mean-error relationship of fractional data.

Data Access

All small RNA data have been deposited in the Sequence Read Archive, BioProject PRJNA533115.

Code Access

All R codes regarding data visualization and statistical analyses were deposited in https://github.com/cxli233/zygote_smRNA/

Acknowledgements

We thank Zachary Liechty and Christian Santos for assistance in R programming; and Alina Yalda, Jake Anichowski, and Michelle Binyu Cui for greenhouse maintenance and technical assistance. The UC Davis Genome Center provided Illumina sequencing, library quality control and size selection services. CL also acknowledges the partial supported by Elsie Taylor Stocking Memorial Fellowship from the Department of Plant Biology at University of California, Davis. This study was supported in part by resources and technical expertise from the Georgia Advanced Computing Resource Center, a partnership between the University of Georgia's Office of the Vice President for Research and Office of the Vice President for

Information Technology. This research was funded by the National Science Foundation (IOS-1547760) and the U.S. Department of Agriculture (USDA) Agricultural Experiment Station (CA-D-XXX-6973-H).

Author contributions

CL, JIG, SDR and VS designed the study. HX and HF collected zygotes. SDR supervised zygote collections. CL produced small RNA sequencing libraries. CL and JIG analyzed data. VS supervised data collection and analyses. CL wrote the manuscript with input from all authors.

References

Anderson, S.N., Johnson, C.S., Chesnut, J., Jones, D.S., Khanday, I., Woodhouse, M., Li, C., Conrad, L.J., Russell, S.D., Sundaresan, V., 2017. The Zygotic Transition Is Initiated in Unicellular Plant Zygotes with Asymmetric Activation of Parental Genomes. Dev. Cell 43, 349-358.e4. https://doi.org/10.1016/j.devcel.2017.10.005

Anderson, S.N., Johnson, C.S., Jones, D.S., Conrad, L.J., Gou, X., Russell, S.D., Sundaresan, V., 2013. Transcriptomes of isolated Oryza sativa gametes characterized by deep sequencing: evidence for distinct sex-dependent chromatin and epigenetic states before fertilization. Plant J. 76, 729–741. https://doi.org/10.1111/tpj.12336

Armenta-Medina, A., Gillmor, C.S., 2019. Genetic, molecular and parent-of-origin regulation of early embryogenesis in flowering plants, in: Current Topics in Developmental Biology. Elsevier, pp. 497–543. https://doi.org/10.1016/bs.ctdb.2018.11.008

Autran, D., Baroux, C., Raissig, M.T., Lenormand, T., Wittig, M., Grob, S., Steimer, A., Barann, M., Klostermeier, U.C., Leblanc, O., Vielle-Calzada, J.-P., Rosenstiel, P., Grimanelli, D., Grossniklaus, U., 2011. Maternal Epigenetic Pathways Control Parental Contributions to Arabidopsis Early Embryogenesis. Cell 145, 707–719. https://doi.org/10.1016/j.cell.2011.04.014

Axtell, M.J., 2013. ShortStack: Comprehensive annotation and quantification of small RNA genes. RNA 19, 740–751. https://doi.org/10.1261/rna.035279.112

Borg, M., Berger, F., 2015. Chromatin remodelling during male gametophyte development. Plant J. 83, 177–188. https://doi.org/10.1111/tpj.12856

Borg, M., Jacob, Y., Susaki, D., LeBlanc, C., Buendía, D., Axelsson, E., Kawashima, T., Voigt, P., Boavida, L., Becker, J., Higashiyama, T., Martienssen, R., Berger, F., 2020. Targeted reprogramming of H3K27me3 resets epigenetic memory in plant paternal chromatin. Nat. Cell Biol. 22, 621–629. https://doi.org/10.1038/s41556-020-0515-y

Borges, F., Parent, J., Ex, F. Van, Wolff, P., Martínez, G., Köhler, C., Martienssen, R.A., 2018. Transposon-derived small RNAs triggered by miR845 mediate genome dosage response in Arabidopsis. Nat. Genet. 50. https://doi.org/10.1038/s41588-017-0032-5

Calarco, J.P., Borges, F., Donoghue, M.T.A., Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijó, J.A., Becker, J.D., Martienssen, R.A., 2012. Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA. Cell 151, 194–205. https://doi.org/10.1016/j.cell.2012.09.001

Chen, J., Strieder, N., Krohn, N.G., Cyprys, P., Sprunck, S., Engelmann, J.C., Dresselhaus, T., 2017. Zygotic Genome Activation Occurs Shortly after Fertilization in Maize. Plant Cell 29, 2106–2125. https://doi.org/10.1105/tpc.17.00099

Cuerda-gil, D., Slotkin, R.K., 2016. Non-canonical RNA-directed DNA methylation. https://doi.org/10.1038/NPLANTS.2016.163

Ding, J., Shen, J., Li, W., Yang, H., 2009. Cytological Observation of Double Fertilization and Its Duration in Oryza sativa. Chi N Bull Bot 473–483.

Dorweiler, J.E., Carey, C.C., Kubo, K.M., Hollick, J.B., Kermicle, J.L., Chandler, V.L., 2000. mediator of paramutation1 Is Required for Establishment and Maintenance of Paramutation at Multiple Maize Loci 19.

Erdmann, R.M., Satyaki, P.R. V, Klosinska, M., Gehring, M., Erdmann, R.M., Satyaki, P.R. V, Klosinska, M., Gehring, M., 2017. A Small RNA Pathway Mediates Allelic Dosage in A Small RNA Pathway Mediates Allelic Dosage in Endosperm. CellReports 21, 3364–3372. https://doi.org/10.1016/j.celrep.2017.11.078

Gehring, M., 2019. Epigenetic dynamics during flowering plant reproduction: evidence for reprogramming? New Phytol. 224, 91–96. https://doi.org/10.1111/nph.15856

Gent, J. I., Ellis, N.A., Guo, L., Harkess, A.E., Yao, Y., Zhang, X., Dawe, R.K., 2013. CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. Genome Res. 23, 628–637. https://doi.org/10.1101/gr.146985.112

Grover, J.W., Burgess, D., Kendall, T., Baten, A., Pokhrel, S., King, G.J., Meyers, B.C., Freeling, M., Mosher, R.A., 2020. Abundant expression of maternal siRNAs is a conserved feature of seed development. Proc. Natl. Acad. Sci. 202001332. https://doi.org/10.1073/pnas.2001332117

Grover, J.W., Kendall, T., Baten, A., Burgess, D., Freeling, M., King, G.J., Mosher, R.A., 2018. Maternal components of RNA -directed DNA methylation are required for seed development in *Brassica rapa*. Plant J. 94, 575–582. https://doi.org/10.1111/tpj.13910

Guo, W., Zhu, P., Pellegrini, M., Zhang, M.Q., Wang, X., Ni, Z., 2018. CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. Bioinformatics 34, 381–387. https://doi.org/10.1093/bioinformatics/btx595

Han, Y., Qin, S., Wessler, S.R., 2013. Comparison of class 2 transposable elements at superfamily resolution reveals conserved and distinct features in cereal grass genomes. BMC Genomics 14, 71. https://doi.org/10.1186/1471-2164-14-71

Houri-Zeevi, L., Rechavi, O., 2017. A Matter of Time: Small RNAs Regulate the Duration of Epigenetic Inheritance. Trends Genet. 33, 46–57. https://doi.org/10.1016/j.tig.2016.11.001

Hsieh, P., He, S., Buttress, T., Gao, H., Couchman, M., Fischer, R.L., 2016. Arabidopsis male sexual lineage exhibits more robust maintenance of CG methylation than somatic tissues. PNAS 113. https://doi.org/10.1073/pnas.1619074114

Hsieh, T.-F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., Zilberman, D., 2009. Genome-Wide Demethylation of Arabidopsis Endosperm. Science 324, 1451–1454. https://doi.org/10.1126/science.1172417

Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.-F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., Rojas, D., Fischer, R.L., Tamaru, H., Zilberman, D., 2012. Active DNA Demethylation in Plant Companion Cells Reinforces

Transposon Methylation in Gametes. Science 337, 1360–1364. https://doi.org/10.1126/science.1224839

Ingouff, M., Hamamura, Y., Gourgues, M., Higashiyama, T., 2007. Distinct Dynamics of HISTONE3 Variants between the Two Fertilization Products in Plants. Curr. Biol. 1032–1037. https://doi.org/10.1016/j.cub.2007.05.019

Ingouff, M., Rademacher, S., Holec, S., Xin, N., Readshaw, A., Foo, S.H., 2010. Report Zygotic Resetting of the HISTONE 3 Variant Repertoire Participates in Epigenetic Reprogramming in Arabidopsis. Curr. Biol. 2137–2143. https://doi.org/10.1016/j.cub.2010.11.012

Kao, P., Nodine, M.D., 2019. Transcriptional Activation of Arabidopsis Zygotes Is Required for Initial Cell Divisions. Sci. Rep. 9, 17159. https://doi.org/10.1038/s41598-019-53704-2

Kawahara, Y., de la Bastide, M., Hamilton, J.P., Kanamori, H., McCombie, W.R., Ouyang, S., Schwartz, D.C., Tanaka, T., Wu, J., Zhou, S., Childs, K.L., Davidson, R.M., Lin, H., Quesada-Ocampo, L., Vaillancourt, B., Sakai, H., Lee, S.S., Kim, J., Numa, H., Itoh, T., Buell, C.R., Matsumoto, T., 2013a. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. Rice 6, 4. https://doi.org/10.1186/1939-8433-6-4

Kim, M.Y., Ono, A., Scholten, S., Kinoshita, T., Zilberman, D., Okamoto, T., Fischer, R.L., 2019. DNA demethylation by ROS1a in rice vegetative cells promotes methylation in sperm. Proc. Natl. Acad. Sci. 116, 9652–9657. https://doi.org/10.1073/pnas.1821435116

Kimmins, S., Sassone-corsi, P., 2005. Chromatin remodelling and epigenetic features of germ cells. Nature 583–589.

Kozomara, A., Birgaoanu, M., Griffiths-Jones, S., 2019a. miRBase: from microRNA sequences to function. Nucleic Acids Res. 47, D155–D162. https://doi.org/10.1093/nar/gky1141

Kranz, E., Bautor, J., Lörz, H., 1991. In vitro fertilization of single, isolated gametes of maize mediated by electrofusion. Sex. Plant Reprod. 4. https://doi.org/10.1007/BF00194565

Li, C., Xu, H., Fu, F.-F., Russell, S.D., Sundaresan, V., Gent, J.I., 2020. Genome-wide redistribution of 24-nt siRNAs in rice gametes. Genome Res. 30, 173–184. https://doi.org/10.1101/gr.253674.119

Li, C., Xu, H., Russell, S.D., Sundaresan, V., 2019. Step-by-step protocols for rice gamete isolation. Plant Reprod. 32, 5–13. https://doi.org/10.1007/s00497-019-00363-y

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, Q., Gent, J.I., Zynda, G., Song, J., Makarevitch, I., Hirsch, C.D., Hirsch, C.N., Dawe, R.K., Madzima, T.F., McGinnis, K.M., Lisch, D., Schmitz, R.J., Vaughn, M.W., Springer, N.M., 2015. RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. Proc. Natl. Acad. Sci. 112, 14728–14733. https://doi.org/10.1073/pnas.1514680112

Lord, E.M., Russell, S.D., 2002. The Mechanisms of Pollination and Fertilization in Plants. Annu. Rev. Cell Dev. Biol. 18, 81–105. https://doi.org/10.1146/annurev.cellbio.18.012502.083438

Martin, M., 2011. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. EMBnetjournal 17, 3.

Martínez, G., Köhler, C., 2017. Role of small RNAs in epigenetic reprogramming during plant sexual reproduction. Curr. Opin. Plant Biol. 36, 22–28. https://doi.org/10.1016/j.pbi.2016.12.006

Martínez, G., Panda, K., Köhler, C., Slotkin, R.K., 2016. Silencing in sperm cells is directed by RNA movement from the surrounding nurse cell. Nat. Plants 2, 1–8. https://doi.org/10.1038/nplants.2016.30

Martínez, G., Wolff, P., Wang, Z., Moreno-romero, J., Santos-gonzález, J., Conze, L.L., Defraia, C., Slotkin, R.K., Köhler, C., 2018. Paternal easiRNAs regulate parental genome dosage in Arabidopsis. Nat. Genet. 50. https://doi.org/10.1038/s41588-017-0033-4

Matzke, M.A., Mosher, R.A., 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat. Rev. Genet. 15, 394–408. https://doi.org/10.1038/nrg3683

McCarthy, D.J., Chen, Y., Smyth, G.K., 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 40, 4288–4297. https://doi.org/10.1093/nar/gks042

Mérai, Z., Chumak, N., García-Aguilar, M., Hsieh, T.-F., Nishimura, T., Schoft, V.K., Bindics, J., Ślusarz, L., Arnoux, S., Opravil, S., Mechtler, K., Zilberman, D., Fischer, R.L., Tamaru, H., 2014. The AAA-ATPase molecular chaperone Cdc48/p97 disassembles sumoylated centromeres, decondenses heterochromatin, and activates ribosomal RNA genes. Proc. Natl. Acad. Sci. 111, 16166–16171. https://doi.org/10.1073/pnas.1418564111

Messerschmidt, D.M., Knowles, B.B., Solter, D., 2014. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes Dev. 28, 812–828. https://doi.org/10.1101/gad.234294.113

Mizuno, H., Matsumoto, T., Wu, J., 2018. Composition and Structure of Rice Centromeres and Telomeres, in: Sasaki, T., Ashikari, M. (Eds.), Rice Genomics, Genetics and Breeding. Springer Singapore, Singapore, pp. 37–52. https://doi.org/10.1007/978-981-10-7461-5_3

Moritoh, S., Eun, C.-H., Ono, A., Asao, H., Okano, Y., Yamaguchi, K., Shimatani, Z., Koizumi, A., Terada, R., 2012. Targeted disruption of an orthologue of DOMAINS REARRANGED METHYLASE 2, OsDRM2, impairs the growth of rice plants by abnormal DNA methylation: osdrm2 affects DNA methylation and development. Plant J. 71, 85–98. https://doi.org/10.1111/j.1365-313X.2012.04974.x

Okada, T., Endo, M., Singh, M.B., Bhalla, P.L., 2005. Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3: Histone H3 gene family in Arabidopsis. Plant J. 44, 557–568. https://doi.org/10.1111/j.1365-313X.2005.02554.x

Papareddy, R.K., Páldi, K., Paulraj, S., Kao, P., Lutzmayer, S., Nodine, M.D., 2020. Chromatin regulates expression of small RNAs to help maintain transposon methylome homeostasis in Arabidopsis. Genome Biol. 21, 251. https://doi.org/10.1186/s13059-020-02163-4

Park, K., Kim, M.Y., Vickers, M., Park, J.-S., Hyun, Y., Okamoto, T., Zilberman, D., Fischer, R.L., Feng, X., Choi, Y., Scholten, S., 2016. DNA demethylation is initiated in the central cells of *Arabidopsis* and rice. Proc. Natl. Acad. Sci. 113, 15138–15143. https://doi.org/10.1073/pnas.1619047114

Pillot, M., Baroux, C., Arteaga Vazquez, M., Autran, D., Leblanc, O., Vielle-calzada, J.P., Grossniklaus, U., Grimanelli, D., 2010. Embryo and Endosperm Inherit Distinct

Chromatin and Transcriptional States from the Female Gametes in Arabidopsis. Plant Cell 22, 307–320. https://doi.org/10.1105/tpc.109.071647

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rodrigues, J.A., Ruan, R., Nishimura, T., Sharma, M.K., Sharma, R., Ronald, P.C., Fischer, R.L., 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. 110, 7934–7939. https://doi.org/10.1073/pnas.1306164110

Saitou, M., Kagiwada, S., Kurimoto, K., 2012. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. Development 139, 15–31. https://doi.org/10.1242/dev.050849

Satyaki, P.R. V, Gehring, M., 2019. Paternally Acting Canonical RNA-Directed DNA Methylation Pathway Genes Sensitize Arabidopsis Endosperm to Paternal. Plant Cell 31, 1563–1578. https://doi.org/10.1105/tpc.19.00047

Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863–864. https://doi.org/10.1093/bioinformatics/btr026

Schoft, V.K., Chumak, N., Mosiolek, M., Slusarz, L., Komnenovic, V., Brownfield, L., Twell, D., Kakutani, T., Tamaru, H., 2009. Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. Sci. Rep. 10, 1015–1021. https://doi.org/10.1038/embor.2009.152

Scholten, S., Lörz, H., Kranz, E., 2002. Paternal mRNA and protein synthesis coincides with male chromatin decondensation in maize zygotes. Plant J. 32, 221–231. https://doi.org/10.1046/j.1365-313X.2002.01418.x

Searle, S.R., Speed, F.M., Milliken, G.A., 1980. Population Marginal Means in the Linear Model: An Alternative to Least Squares Means. Am. Stat. 34, 216–221. https://doi.org/10.1080/00031305.1980.10483031

Slotkin, R Keith, Vaughn, M., Borges, F., Feijo, A., Becker, D., Martienssen, R.A., 2009. Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. Cell 461–472. https://doi.org/10.1016/j.cell.2008.12.038

Su, Z., Wang, N., Hou, Z., Li, B., Li, D., Liu, Y., Cai, H., Qin, Y., Chen, X., 2020. Regulation of Female Germline Specification via Small RNA Mobility in Arabidopsis. Plant Cell tpc.00126.2020. https://doi.org/10.1105/tpc.20.00126

Tadros, W., Lipshitz, H.D., 2009. The maternal-to-zygotic transition: a play in two acts. Development 136, 3033–3042. https://doi.org/10.1242/dev.033183

Tan, F., Lu, Y., Jiang, W., Zhang, R., Zhao, Y., Zhou, D., 2018. DDM1 Represses Noncoding RNA Expression and RNA-Directed DNA Methylation in Heterochromatin 177, 1187–1197. https://doi.org/10.1104/pp.18.00352

Tan, F., Zhou, C., Zhou, Q., Zhou, S., Yang, W., Zhao, Y., Li, G., 2016. Analysis of Chromatin Regulators Reveals Speci fi c Features of Rice DNA Methylation Pathways. Plant Physiol. 171, 2041–2054. https://doi.org/10.1104/pp.16.00393

Tang, W.W.C., Kobayashi, T., Irie, N., Dietmann, S., Surani, M.A., 2016. Specification and epigenetic programming of the human germ line. Nat. Rev. Genet. 17, 585–600. https://doi.org/10.1038/nrg.2016.88

Thorvaldsdottir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192. https://doi.org/10.1093/bib/bbs017

Wang, G., Köhler, C., 2017. Epigenetic processes in flowering plant reproduction. J. Exp. Bot. erw486. https://doi.org/10.1093/jxb/erw486

Wang, Z., Butel, N., Santos-González, J., Borges, F., Yi, J., Martienssen, R.A., Martinez, G., Köhler, C., 2020. Polymerase IV Plays a Crucial Role in Pollen Development in Capsella. Plant Cell 32, 950–966. https://doi.org/10.1105/tpc.19.00938

Ye, R., Wang, W., Iki, T., Liu, C., Wu, Y., Ishikawa, M., Zhou, X., Qi, Y., 2012. Cytoplasmic Assembly and Selective Nuclear Import of Arabidopsis ARGONAUTE4/siRNA Complexes. Mol. Cell 46, 859–870. https://doi.org/10.1016/j.molcel.2012.04.013

Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., Zilberman, D., 2013. The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. Cell 153, 193–205. https://doi.org/10.1016/j.cell.2013.02.033

Zhao, P., Zhou, X., Shen, K., Liu, Z., Cheng, T., Liu, D., Cheng, Y., Peng, X., Sun, M., 2019. Two-Step Maternal-to-Zygotic Transition with Two-Phase Parental Genome Contributions. Dev. Cell 49, 882-893.e5. https://doi.org/10.1016/j.devcel.2019.04.016

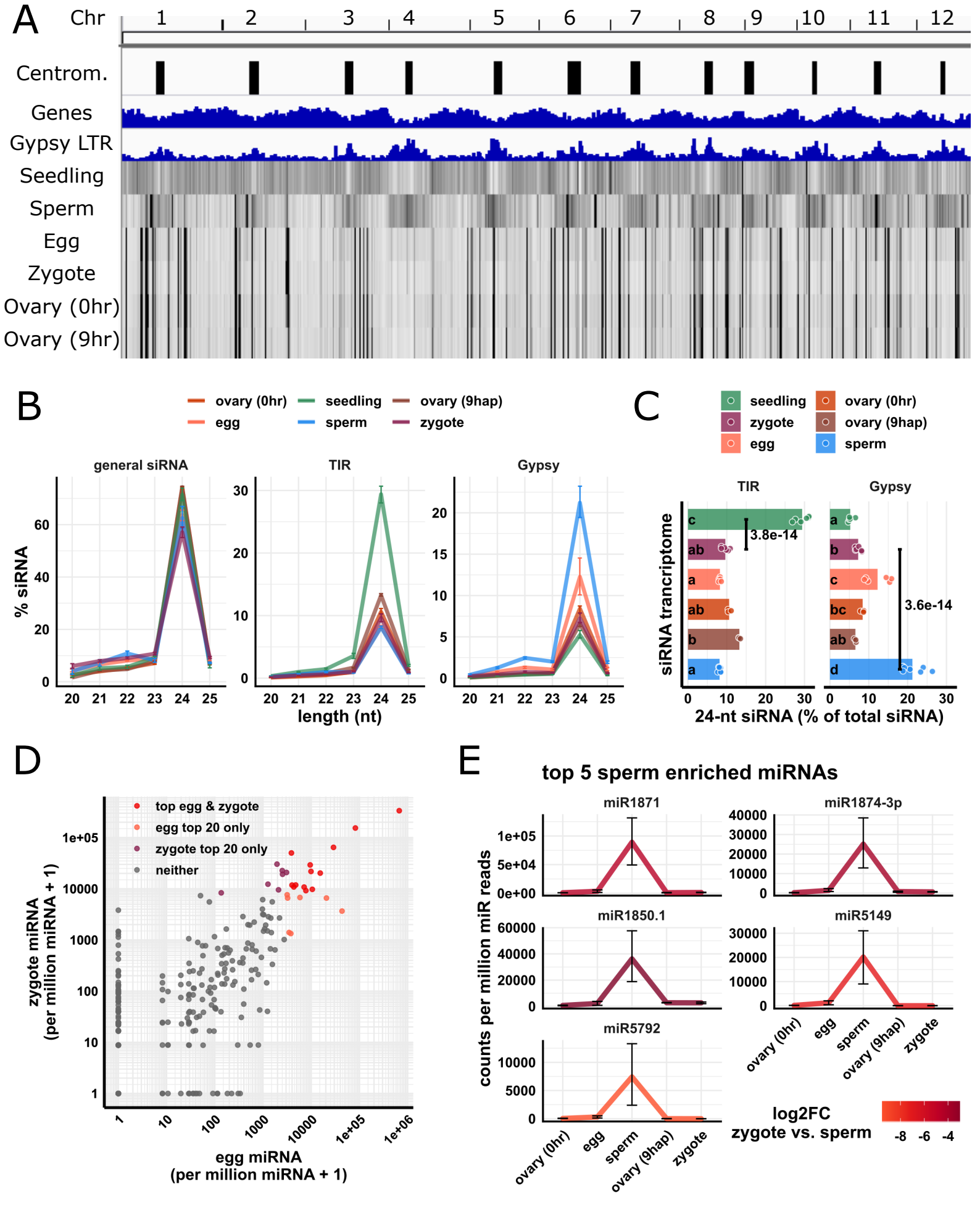

A
Chr
1
2
3
4
5
6
7
8
9
10
11
12
Centrom.
Genes
Gypsy LTR
Seedling
Sperm
Egg
Zygote
Ovary (0hr)
Ovary (9hr)
B
ovary (0hr)
egg
seedling
sperm
ovary (9hap)
zygote
general siRNA
TIR
Gypsy
% siRNA
length (nt)
C
seedling
zygote
egg
ovary (0hr)
ovary (9hap)
sperm
TIR
Gypsy
siRNA trancriptome
3.8e-14
3.6e-14
24-nt siRNA (% of total siRNA)
D
top egg & zygote
egg top 20 only
zygote top 20 only
neither
zygote miRNA
(per million miRNA + 1)
egg miRNA
(per million miRNA + 1)
E
top 5 sperm enriched miRNAs
miR1871
miR1874-3p
miR1850.1
miR5149
miR5792
counts per million miR reads
ovary (0hr)
egg
sperm
ovary (9hap)
zygote
log2FC
zygote vs. sperm
-8
-6
-4

Fig 1: Overall pattern of zygote 24-nt siRNAs is similar to that of the egg cell.

(A) Heat map showing abundance of 24-nt siRNA across genome at 50-kb resolution. Centrom. Centrometric regions, as defined by (Mizuno et al., 2018).

(B) Length profiles of siRNAs. y-axis values are relative to total siRNA reads (20 – 25-nt siRNAs). TIR: terminal inverted repeat transposons, CACTA superfamily excluded. Gypsy: Gypsy retrotransposons. Error bars are 95% confidence intervals for each cell type.

(C) Quantification of TIR and Gypsy panels (B). Each data point is an siRNA transcriptome. Bar heights are averages. Letter grouping ($\alpha = 0.05$) and P values are based on linear models with logit transformation followed by Tukey tests.

(D) Scatter plot showing miRNA relative abundances in egg and zygote. Each data point is a miRNA. Axes are relative to per million miRNA reads and log10 transformed. 'top egg & zygote' refers to intersection of the 20 highest abundant miRNAs in both egg and zygote.

(E) Top five sperm enriched miRNAs. Sperm enriched is classified as > 1000 reads per million miRNA reads in sperm and < 500 reads per million miRNA reads in egg. y-axis values are relative to per million miRNA reads. Color code reflects log2FC values for zygote vs. sperm, and negative values indicate higher in sperm. Error bars are 95% confidence intervals for each cell type. See **Fig S1D** for additional examples.

Zygote and 9 hap ovary data are from this study, all other data from Li et al., (2020).

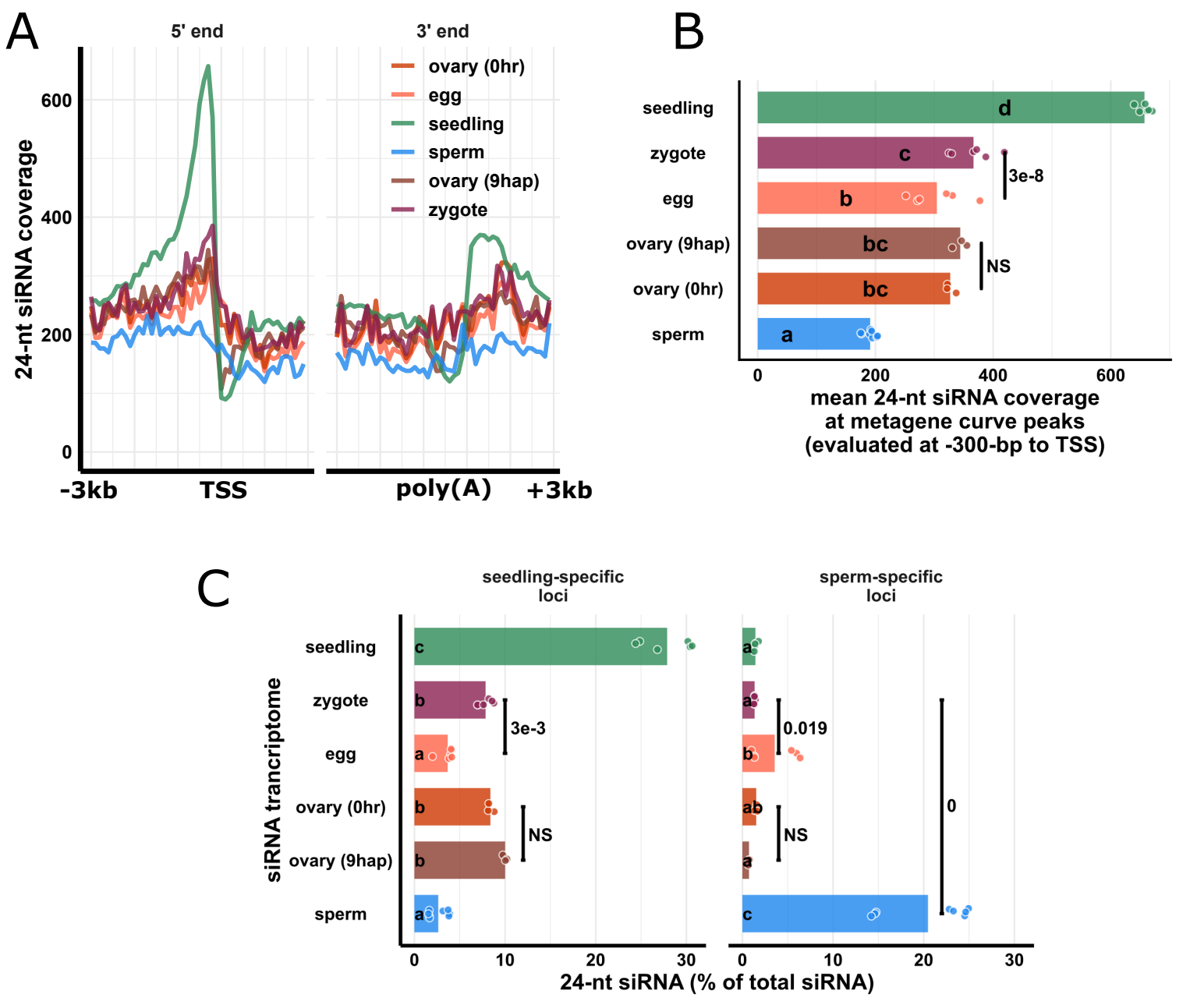

A
5' end
3' end
ovary (0hr)
egg
seedling
sperm
ovary (9hap)
zygote
24-nt siRNA coverage
600
400
200
0
-3kb
TSS
poly(A)
+3kb
B
seedling
zygote
egg
ovary (9hap)
ovary (0hr)
sperm
d
c
b
bc
bc
a
3e-8
NS
0
200
400
600
mean 24-nt siRNA coverage
at metagene curve peaks
(evaluated at -300-bp to TSS)
C
seedling-specific
loci
sperm-specific
loci
siRNA trancriptome
seedling
zygote
egg
ovary (0hr)
ovary (9hap)
sperm
c
b
a
b
b
a
3e-3
NS
a
a
b
ab
a
c
0.019
NS
0
0
10
20
30
24-nt siRNA (% of total siRNA)

Fig 2: Changes in the zygote siRNA transcriptome are independent from the ovary after fertilization.

(A) Metagene coverage plot for 24-nt siRNAs. Coverage is measured over 100-bp intervals and normalized per 1000 24-nt siRNAs. Vertical grid lines are 500-bp intervals. TSS transcription start site, poly(A) polyadenylation site.

(B) Quantification of (A) at the interval from 300 to 200-bp upstream of TSS, corresponding to the peaks of metagene curves. Each data point is an siRNA transcriptome and bar heights are averages. x-axis values are normalized per 1000 24-nt siRNAs. Letter grouping ($\alpha = 0.05$) and P values are based on Tukey tests.

(C) Bar plot showing relative abundances of 24-nt siRNA across siRNA loci categories. Seedling-specific loci are seedling siRNA loci that do not overlap egg or sperm siRNA loci. Sperm-specific loci are sperm siRNA loci that do not overlap egg or seedling siRNA loci. The zygote siRNA transcriptome was not used to define these locus categories. Each data point is an siRNA transcriptome. Bar heights are averages. x-axis values are normalized to per 1000 24-nt siRNAs. Letter grouping ($\alpha = 0.05$) and P values are based on linear models with logit transformation followed by Tukey tests.

Zygote and 9 hap ovary data are from this study, all other data from Li et al., (2020).

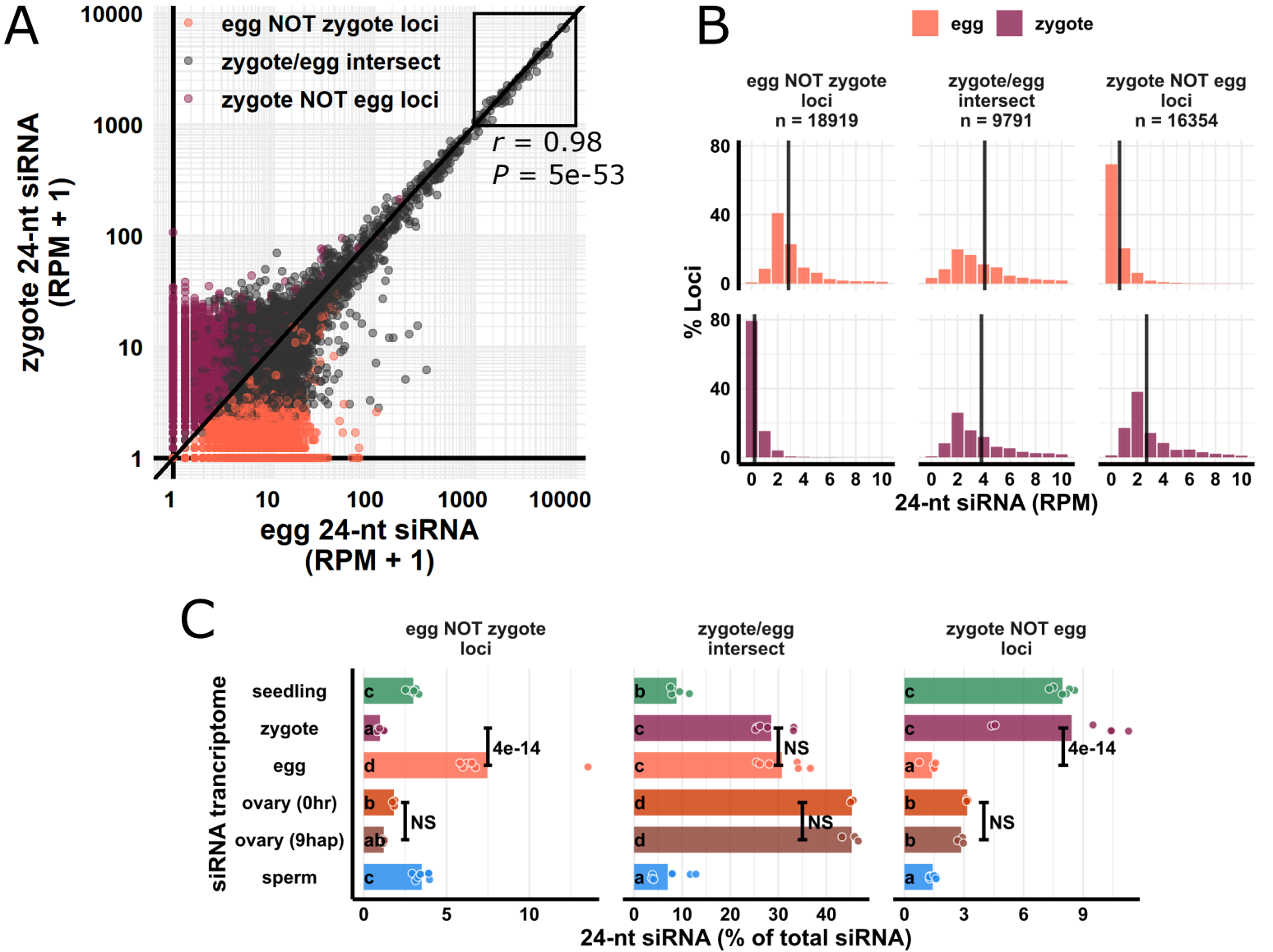

A
egg NOT zygote loci
zygote/egg intersect
zygote NOT egg loci
r = 0.98
P = 5e-53
zygote 24-nt siRNA (RPM + 1)
egg 24-nt siRNA (RPM + 1)
B
egg
zygote
egg NOT zygote loci n = 18919
zygote/egg intersect n = 9791
zygote NOT egg loci n = 16354
% Loci
24-nt siRNA (RPM)
C
egg NOT zygote loci
zygote/egg intersect
zygote NOT egg loci
siRNA trancriptome
seedling
zygote
egg
ovary (0hr)
ovary (9hap)
sperm
4e-14
NS
24-nt siRNA (% of total siRNA)

Fig 3: Widespread newly detected siRNA loci in zygote.

(A) Scatter plot showing relative abundances of 24-nt siRNAs across siRNA loci in egg and zygote. Each data point is an siRNA locus. Zygote NOT egg loci are zygote siRNA loci that do not overlap egg siRNA loci. Egg NOT zygote loci are egg siRNA loci that do not overlap zygote siRNA loci. Zygote/egg intersect are zygote siRNA loci that overlap egg siRNA loci. Axes are relative to per million total siRNA reads (20 – 25-nt) and log10 transformed. Correlation coefficient (*r*) and P value are for locus with RPM > 1000.

(B) Quantification of (A). Vertical lines are medians. x-axis values are relative to per million total siRNA reads (20 – 25-nt) and cut off at RPM = 10. Bar heights represent percentage of siRNA loci in each locus category.

(C) Quantification of 24-nt siRNA relative abundances for (A). Each data point is a siRNA transcriptome. Bar heights are averages. x-axis-values are relative to total siRNA reads (20 – 25-nt). Letter grouping ($\alpha = 0.05$) and P values are based on linear model with logit transformation followed by Tukey tests.

Zygote and 9 hap ovary data are from this study, all other data from Li et al., (2020).

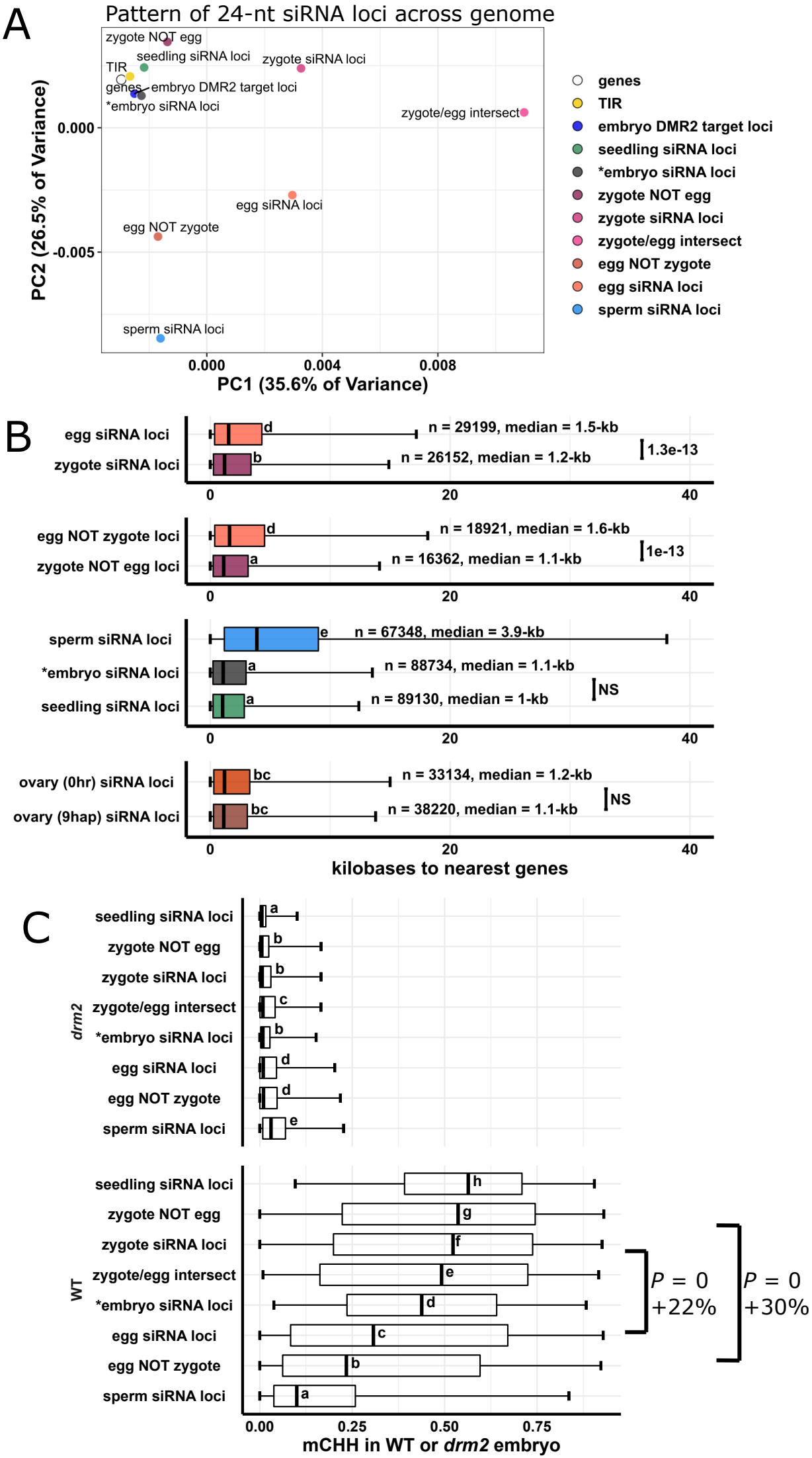

A
Pattern of 24-nt siRNA loci across genome
zygote NOT egg
seedling siRNA loci
zygote siRNA loci
TIR
genes
embryo DMR2 target loci
*embryo siRNA loci
zygote/egg intersect
egg siRNA loci
egg NOT zygote
sperm siRNA loci
PC2 (26.5% of Variance)
0.000
-0.005
0.000
0.004
0.008
PC1 (35.6% of Variance)
genes
TIR
embryo DMR2 target loci
seedling siRNA loci
*embryo siRNA loci
zygote NOT egg
zygote siRNA loci
zygote/egg intersect
egg NOT zygote
egg siRNA loci
sperm siRNA loci
B
egg siRNA loci
zygote siRNA loci
n = 29199, median = 1.5-kb
n = 26152, median = 1.2-kb
1.3e-13
egg NOT zygote loci
zygote NOT egg loci
n = 18921, median = 1.6-kb
n = 16362, median = 1.1-kb
1e-13
sperm siRNA loci
*embryo siRNA loci
seedling siRNA loci
n = 67348, median = 3.9-kb
n = 88734, median = 1.1-kb
n = 89130, median = 1-kb
NS
ovary (0hr) siRNA loci
ovary (9hap) siRNA loci
n = 33134, median = 1.2-kb
n = 38220, median = 1.1-kb
NS
0
20
40
kilobases to nearest genes
C
drm2
seedling siRNA loci
zygote NOT egg
zygote siRNA loci
zygote/egg intersect
*embryo siRNA loci
egg siRNA loci
egg NOT zygote
sperm siRNA loci
WT
P = 0
+22%
P = 0
+30%
0.00
0.25
0.50
0.75
mCHH in WT or drm2 embryo

Fig 4: Newly detected siRNA loci in zygote reset to the canonical siRNA profile and predict CHH methylation in embryo in an RdDM-dependent manner.

(A) Principal component analysis of siRNA loci distributions across genome. Locus distribution are evaluated with 50-kb resolution. Each data point represents the genomic distribution of a siRNA loci category.

(B) Boxplots showing distance of siRNA loci to nearest genes. Middle lines are median. Boxes span interquartile range. Whiskers span 2.5^{th} and 97.5^{th} percentiles. Letter groupings ($\alpha = 0.05$) and P values are based on linear model with log10(distance + 1-bp) followed by Tukey tests.

(C) Boxplots showing CHH methylation level in mature wildtype and *drm2* mutant embryos. Middle lines are median. Boxes span interquartile range. Whiskers span 2.5^{th} and 97.5^{th} percentiles. Letter groupings ($\alpha = 0.05$) and P values are based on generalized linear model with logit link followed by Tukey tests. Zygote NOT egg loci $n = 16354$, zygote/egg intersect $n = 9791$, egg NOT zygote loci $n = 18919$.

*Embryo siRNA data from Rodrigues et al., (2013), which was based on a single replicate. Seedling, gametes and pre-fertilization ovary data from Li et al., (2020).

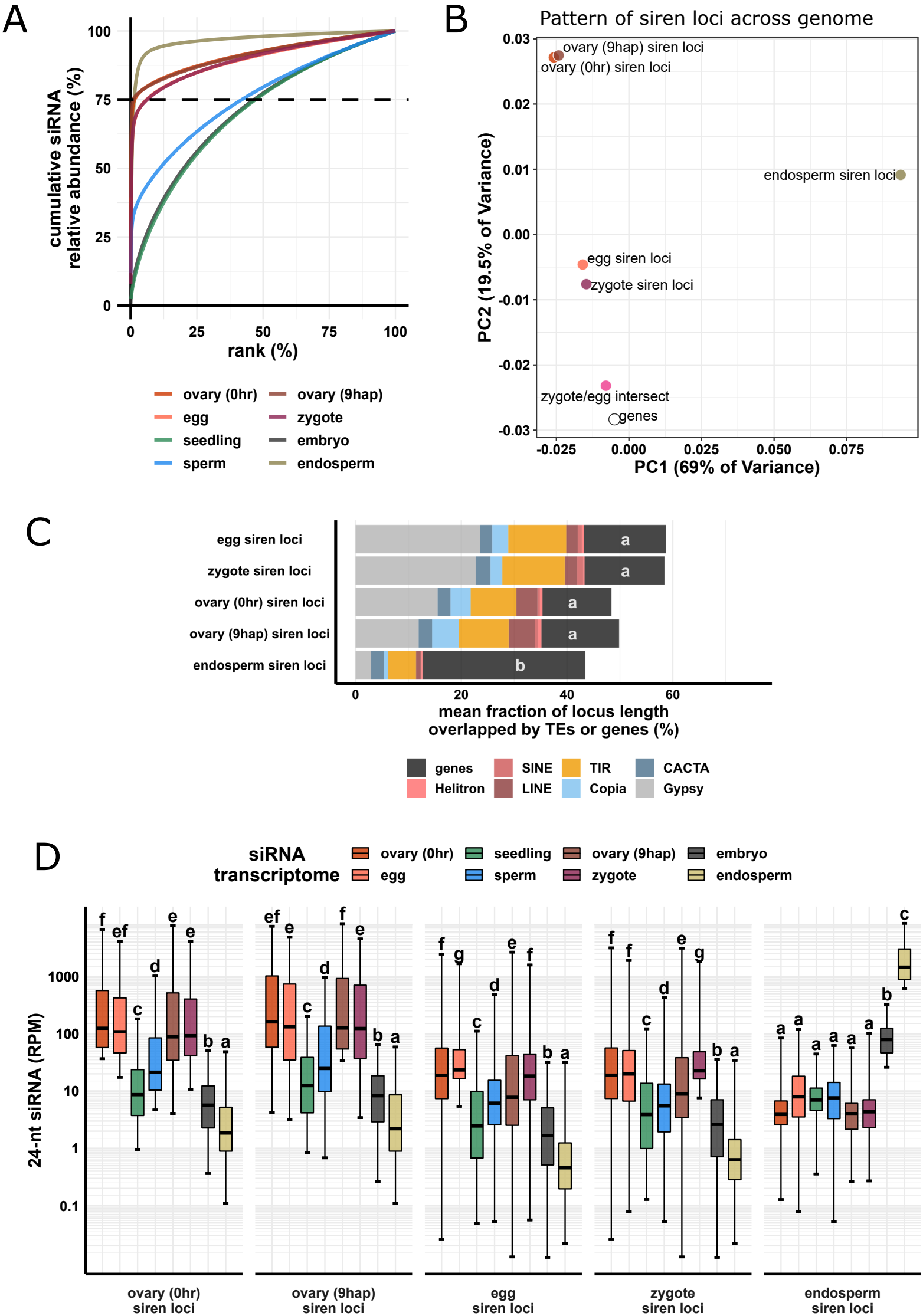

A
cumulative siRNA relative abundance (%)
rank (%)
ovary (0hr)
ovary (9hap)
egg
zygote
seedling
embryo
sperm
endosperm
B
Pattern of siren loci across genome
ovary (9hap) siren loci
ovary (0hr) siren loci
endosperm siren loci
egg siren loci
zygote siren loci
zygote/egg intersect
genes
PC2 (19.5% of Variance)
PC1 (69% of Variance)
C
egg siren loci
zygote siren loci
ovary (0hr) siren loci
ovary (9hap) siren loci
endosperm siren loci
mean fraction of locus length overlapped by TEs or genes (%)
genes
SINE
TIR
CACTA
Helitron
LINE
Copia
Gypsy
D
siRNA transcriptome
ovary (0hr)
seedling
ovary (9hap)
embryo
egg
sperm
zygote
endosperm
24-nt siRNA (RPM)
ovary (0hr) siren loci
ovary (9hap) siren loci
egg siren loci
zygote siren loci
endosperm siren loci

Fig 5: zygote siren loci are similar to siren loci detected in ovary and egg cell, but dissimilar to siren loci detected in endosperm.

(A) x-axis is the rank order of siRNA loci. siRNA loci with highest siRNA abundances are ranked first. y-axis is cumulative relative abundance of siRNA in all siRNA loci. Axis values are scaled between 0 and 100%. 1% of siRNA loci accounted for 75% of siRNA reads in all siRNA loci in endosperm and ovary. 6% of siRNA loci accounted for 75% of siRNA reads in all siRNA loci in egg and zygote.

(B) Principal component plot siren loci distribution across the genome. Distributions are evaluated at 50-kb resolution across the genome. Each data point is the distribution of a siren loci category.

(C) Stacked bar plots showing mean fraction of locus length overlapped by TEs or genes. Letter grouping ($\alpha = 0.05$) is based generalized linear model with logit link followed by Tukey tests.

(D) Boxplots showing 24-nt siRNA relative abundances across siren classes across cell types. Middle lines are median. Boxes span interquartile range. y-axis values are relative to per million total 24-nt siRNA in each siRNA transcriptome. Whiskers span 2.5^{th} and 97.5^{th} percentiles. Letter groupings ($\alpha = 0.05$) are based on linear model with log(RPM + 1) followed by Tukey tests.

Embryo and endosperm data from Rodrigues et al., (2013). Seedling, gametes and pre-fertilization ovary data from Li et al., (2020).