bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Resetting of 24-nt siRNA landscape is initiated in the unicellular zygote 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 stage of the life cycle for all sexually reproducing organisms. It is produced by fusion of two differentiated cells, egg and sperm, that in plants have radically different siRNA transcriptomes from each other and from multicellular embryos. Here we examined the small RNA transcriptome of unicellular rice zygotes. We find that the overall 24-nt siRNA landscape in the zygote resembles that of the unfertilized egg cell, consistent with maternal carry-over. A large fraction (~75%) of the siRNAs in the zygote arise from a small proportion (~2%) of siRNA loci, which corresponded to similar loci in the egg cell and ovary. However, these highly expressing loci were distinct from endosperm siren loci that were detected in later stage endosperm. miRNA abundances changed rapidly after fertilization, resulting in a miRNA profile distinct from the egg cell. Notably, the *de novo* 24-nt siRNAs expressed by the zygote had characteristics of canonical siRNAs, such as proximity to genes and tendency to overlap TIR DNA transposable elements, and resembled seedling siRNA loci, indicating a return to the canonical siRNA profile typical of vegetative tissues. Taken together, our results suggest that resetting of the gametic epigenome towards the canonical vegetative profile is initiated before the first embryonic division.

Introduction

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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). However, MZT in flowering plants proceeds differently [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 (Zhao et al. 2019; Kao and Nodine 2019). These observations suggest that in angiosperms, unlike animals, zygotes are transcriptionally active, and plant ZGA occurs in the zygote.

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 (Kranz, Bautor, and Lörz 1991; Anderson et al. 2013), and its chromatin is diffused (Scholten, Lörz, and Kranz 2002). In contrast, the sperm cell chromatin undergoes global condensation, paralleling protamine deposition in animal sperm cells (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 (Okada et al. 2005; Borg and Berger 2015), 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 proposed to 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; Martinez and Köhler 2017).

In addition to 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 DRM2. DRM2 leads to methylation at all sequence contexts, but methylation at 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 (Erdmann et al. 2017; Borges et al. 2018; Martinez 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 essential for normal plant reproduction.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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, Knowles, and Solter 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, Knowles, and Solter 2014; Saitou, Kagiwada, and Kurimoto 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 (Schoft et al. 2009; Slotkin, et al. 2009; Calarco et al. 2012; Ibarra et al. 2012; Li et al. 2020; Grover et al. 2020), 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, which are enriched around genes, marking the ends of TEs and euchromatin-heterochromatin boundaries (Gent et al. 2013; Li et al. 2015); we refer to such an siRNA profile as the canonical siRNA profile. 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 genome-wide, 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 siRNAs during plant reproduction, we sequenced the small RNA transcriptome of rice zygotes and investigated changes in the small RNA transcriptome soon after fertilization. We inferred that the siRNA transcriptome initiates resetting towards the canonical profile before the first cell division, along with zygotic genome activation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

We collected unicellular rice zygotes 8 hours after pollination (HAP), which corresponds to the completion of S-phase, just prior to the first zygotic division (Ding et al. 2009; Anderson et al. 2017). We produced small RNA transcriptomes from 6 replicates, with ~50 zygotes in each replicate. For our analyses, we also included small RNA transcriptome data from rice gametes, ovary and seedlings (Li et al. 2020). 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 [miRBase v22, (Kozomara, Birgaoanu, and Griffiths-Jones 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.

The global siRNA pattern in zygote is determined by siRNA transcript carryover from the egg cell

We previously found that genome-wide redistribution of siRNAs occurred in gametes (Li et al. 2020). In canonical vegetative tissues, such as seedling shoot, 24-nt siRNAs were produced from gene-rich, euchromatic regions, which correspond to mCHH islands around genes (Gent et al. 2013; Li et al. 2015). In contrast, the sperm cell has a complementary pattern, in which 24-nt siRNA were spread out across wide pericentromeric regions instead. The egg cell and ovary had a different pattern, where 24-nt siRNAs were concentrated at discrete loci (**Fig 1A**, top five tracks). We found that in a whole-genome view, the zygote had a very similar pattern to the egg cell (**Fig 1A**, bottom track). This pattern was reproducible across all zygote replicates (**Fig S1A**). In addition, the 24-nt siRNA expression pattern of zygote clustered with the egg cell, rather than with sperm cell or the average of the gametes (**Fig S1B**). Further, genome-wide 24-nt siRNA relative abundances of zygote and egg cell were highly correlated (r = 0.94, P = 0, Pearson's correlation on egg vs. zygote tracks in **Fig 1A**). The siRNA landscape in zygote can be explained by siRNA 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 siRNAs 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. 2013; 2017). Further, the highest expressed miRNAs in sperm were also very much downregulated in the zygote (**Fig S8A**), again consistent with the prediction that sperm small RNAs were diluted by the egg cell cytoplasm, and that the overall zygote siRNA landscape is determined by transcript carryover from the egg cell.

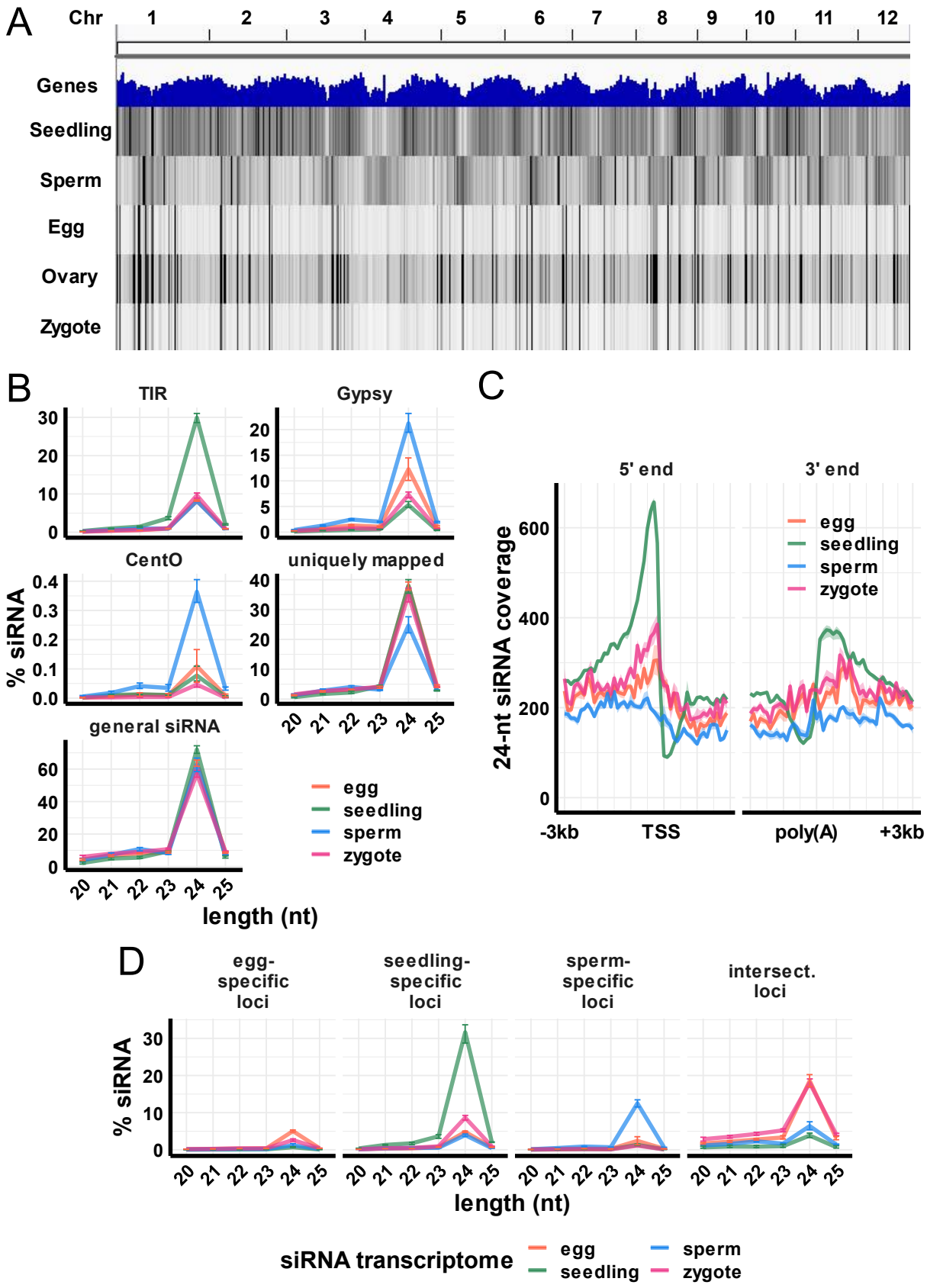
To take a deeper look at the siRNA composition in zygote, we looked at the length profile of siRNAs and compared with our recent data from other cell and tissue types (Li et al. 2020). We found that in zygotes, as in all other tissues, 24-nt siRNAs were the most predominant (**Fig 1B**). This is consistent with the observation that in angiosperms, the most abundant siRNAs are 24-nt siRNA that are involved in RNA-directed DNA methylation [reviewed in (Cuerda-Gil and Slotkin 2016)]. Zygote trended well with egg, except with a slight reduction of *Gypsy* (LTR retrotransposon) reads and CentO (centromeric tandem repeat) reads. Like the egg cell, the zygote has a low abundance of siRNAs overlapping terminal inverted repeat (TIR) transposons (PIF/Harbinger, Tc1/Mariner, Mutator, or hAT superfamilies). These patterns showed that overall, the zygote siRNA pattern is similar to that of the egg cell, supporting that overall siRNAs in zygote were dominated by maternal transcripts carryover from the egg cell.

Next, we looked at distribution of 24-nt siRNAs relative to genes. We produced metagene siRNA coverage plots for seedling, gametes, and zygote (**Fig 1C**). Seedling had a strong peak about 700-bp 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, Qin, and Wessler 2013), and such a peak was absent in gametes. Zygote was similar to egg; however, zygote had a slight increase in 24-nt siRNA coverage upstream of TSS. This observation indicated that although overall zygote siRNA pattern is like the egg cell, the zygote might be starting to return to the canonical siRNA pattern. We previously identified representative genomic regions as egg-specific siRNA loci, seedling-specific siRNA loci, sperm-specific siRNA loci and three-way intersection loci (Li et al. 2020). When we quantified the relative abundances of zygote reads in each category, a couple of observations were apparent. 1) Zygote had very little siRNAs in sperm-specific loci (**Fig 1D**). This is consistent with the prediction that the sperm siRNAs were diluted by the egg cell cytoplasm. 2) More interestingly,

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

zygote had gained siRNAs in seedling-specific loci. This observation is another indication that the recovery of canonical siRNA pattern may be initiated in the zygote.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



188

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig 1 Overall siRNA landscape in zygote is similar to the egg cell and not the sperm cell

(A) Whole genome 24-nt siRNA heat maps. Top track: gene density; second to sixth tracks: 24-nt siRNAs from seedling shoot, sperm cell, egg cell, ovary, and zygote. Chr: chromosomes.

(B) siRNA abundance by length and category. y-axis values are relative to all siRNAs. Error bars are 95% confidence intervals based on biological replicates.

(C) Metagene coverage plot for zygote 24-nt siRNAs. Coverage is measured over 100-bp intervals and normalized per 1000 total siRNAs. Ribbons are 95% confidence intervals for each tissue type. Vertical grid lines indicate 500-bp intervals. TSS transcription start site, poly(A) polyadenylation site.

(D) Zygote siRNA abundance at the 24-nt siRNA loci defined in Li et al (2020). "Intersect." refers to intersection loci, those identified as 24-nt siRNA loci in egg cell, sperm cell and seedling shoot. y-axis values are % of siRNAs of each length relative to the total number of siRNAs that mapped to the genome. Error bars are 95% confidence intervals based on biological replicates of each tissue.

Except zygotes, all data are from (Li et al. 2020).

<u>A small number of siRNA loci in zygote expressed abundant siRNA, which correspond to ovary siren loci</u>

It has been previously reported that rice developing endosperm (7-8 days after pollination) 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 more accurately define edges of variably-sized siRNA loci here we used Shortstack (Axtell 2013) rather than simply dividing the genome into 100-bp bins as previously (Li et al. 2020). In ovary, ~2% (n = 818) of the siRNA loci accounted for 75% of the siRNA, and the numbers were similar in egg cell and zygote (**Fig 2A**). These 818 loci found in ovary we termed ovary siren loci. Here we use the term siren loci in a general sense to describe loci with highly abundant siRNAs, independently of their presence in endosperm, because the abundant siRNA loci in rice ovaries and eggs showed little correlation with the siren loci in endosperm, at least at the specific later endosperm stage examined (Li et al. 2020; Rodrigues et al. 2013).

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

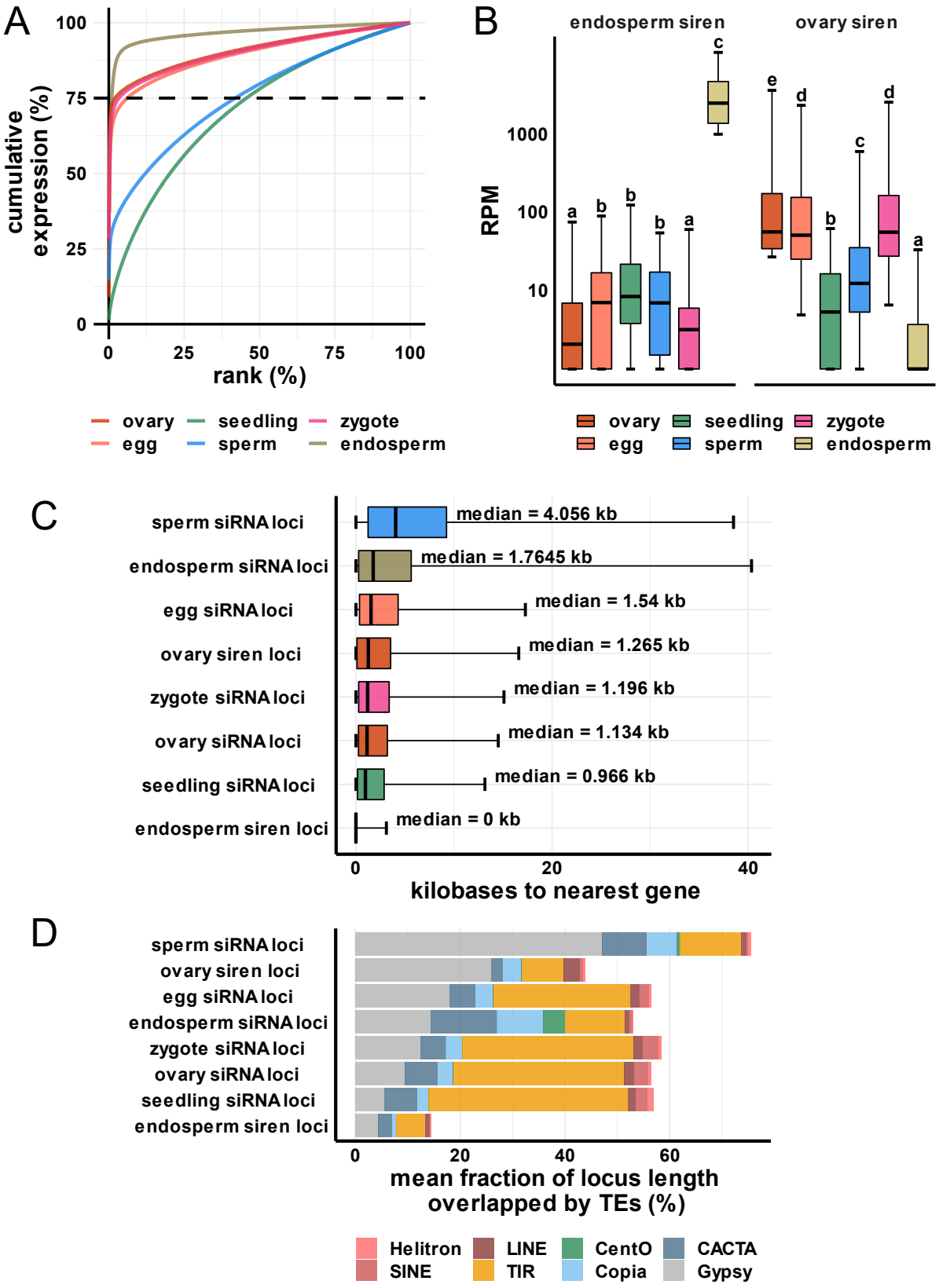
Reanalysis of the endosperm siRNAs revealed an even more extreme distribution, where ~1% (n = 122) of the siRNA loci accounted for 75% of the siRNA expression. For subsequent analysis, we defined these 122 loci as endosperm siren loci. In contrast, in seedling and sperm cell, ~45% of the siRNA loci accounted for 75% of the siRNA expression (**Fig 2A**). Siren loci were found on all 12 chromosomes and had no clear relationship relative to gene density in terms of their genomic distributions (**Fig S2A**). Similar to what was reported in *Brassica* (Grover et al. 2020), siren loci were on average longer than non-siren siRNA loci (**Fig S2B**). Median lengths for endosperm siren loci and ovary siren loci are 6.4-kb and 1.4-kb, respectively, whereas median lengths for other siRNA locus categories were well below 1-kb (**Fig S2B**). However, even after adjusting for length, siren loci were still ~10-fold higher expressed per kilobase than non-siren loci in endosperm and ovary (**Fig S2C**).

We reanalyzed publicly available DNA methylome datasets, including those from our previous study, to examine the methylation status at ovary siren loci (Tan et al. 2016; 2018; Li et al. 2020). mCG and mCHG were higher at ovary siren loci relative to non-ovary-siren loci across all tissues examined, even in *ddm1* and *drm2* leaves (**Fig S3A – B**, $P < 10^{-5}$, generalized linear model with logit link followed by Tukey tests). Our previous analysis of ovary small RNA loci, which also included thousands of small RNA loci with lower 24-nt siRNA abundance, revealed that most egg/ovary-specific small RNA loci have low levels of mCHH (Li et al. 2020). In contrast, in this set of 818 ovary siren loci, mCHH was elevated compared to non-ovary-siren loci in ovary and egg cell (**Fig S3C**, $P < 10^{-5}$, generalized linear model with logit link followed by Tukey tests). Palea and lemma of rice florets and mature embryo also had elevated mCHH at ovary siren loci (**Fig S3C**). mCHH levels were not as strongly elevated, if at all, at ovary siren loci in the rest of the tissues (**Fig S3C**). DNA methylation at all contexts was also consistently higher around and on genes nearest or overlapping ovary siren loci in ovary and egg cell (**Fig S4**). To test whether ovary siren loci were associated with histone modifications indicative of heterochromatin, we reanalyzed publicly available ChIP-seq datasets (**Fig S5**) (Tan et al. 2016; Lu et al. 2018; Liu et al. 2016; Zhang et al. 2012; Lu et al. 2015; Maher et al. 2018; Lu et al. 2020; Zahraeifard et al. 2018; Zhang et al. 2017). We clustered siRNA locus categories based on their chromatin profiles (**Fig S6**). We found that in vegetative tissues, such as seedling shoot, on which most of the published datasets were based, ovary siren loci had an intermediate chromatin profile between euchromatin and heterochromatin (**Fig S5** and **Fig S6**).

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In *Brassica rapa* and *Arabidopsis*, siren loci detected in ovules matched those detected in endosperm (Grover et al. 2020). Our 8 hour zygote siRNAs did not match the 7-8 day endosperm siRNAs in rice embryo (Rodrigues et al. 2013). We found that the 7-8 day endosperm siren loci had low siRNA expression in other tissues, whereas ovary siren loci had high expression in egg cell and zygote, but low expression in seedling, sperm cell and endosperm (**Fig 2B**, linear model with log(RPM) followed by Tukey tests under a $P = 0.05$ cutoff), consistent with the levels of mCHH at ovary siren loci in these tissues (**Fig S3C**). In addition, most endosperm siren loci overlapped genes (median distance to nearest genes = 0-bp), whereas ovary siren loci had an intermediate distance from their adjacent genes (median distance to nearest genes = 1.54-kb, **Fig 2C**). Unlike other siRNA loci, siren loci were less likely to overlap annotated transposable elements. On average, ~15% of the length of endosperm siren loci overlapped annotated transposable elements, and ~45% for ovary siren loci (**Fig 2D**). The majority of ovary siren loci that did overlap transposon loci overlapped *Gypsy* retrotransposons. On average, ~25% of the length of ovary siren loci overlapped *Gypsy* retrotransposons (**Fig 2D**). The subset of loci that did overlap annotated *Gypsy* elements had on average ~3-fold fewer siRNAs per kilobase than the those that did not (**Fig S3D**, $P = 6.4e\text{-}42$, linear model with log(RPKM) followed by Tukey test). In addition, ovary siren loci are unlikely to match an unannotated high-copy transposable element, as siRNAs from ovary siren loci were less repetitive than siRNAs from ovary siRNA loci as a whole (**Fig S3E**). Taken together, these results showed that a small number of siRNA loci that were not associated with transposable elements accounted for the majority of siRNAs in zygote, and siRNAs from these loci were also highly expressed in the egg cell and ovary, and where they coincided with high mCHH.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



273

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig 2 Loci with abundant siRNA expression in zygote corresponded to ovary siren loci rather than endosperm siren loci

(A) Cumulative expression plotted against the ranks of 24-nt-dominant loci. Expression at each locus is determined in reads per million (RPM) after merging replicates, and cumulative expression is plotted against percentage rank, with highest and lowest expressing loci scaled to 0 and 100%, respectively. The 75% cutoff is selected based on the turning points of ovary, egg cell and zygote curves.

(B) Box plot showing siRNA expression at endosperm siren loci and ovary siren loci. RPM: reads per million. Center line is median; box spans interquartile range; whiskers span 2.5th to 97.5th percentiles. y-axis is log10 transformed. Letter groupings are based on linear model with log(RPM) followed by Tukey tests under a $P = 0.05$ cutoff.

(C) Box plot showing distance from siRNA loci to respective nearest or overlapping genes. Center line is median; box spans interquartile range; whiskers span 2.5th to 97.5th percentiles.

(D) Stacked bar plots showing mean fraction of locus length overlapping different transposable element superfamilies.

Endosperm data sources (Rodrigues et al. 2013). Ovary, egg, sperm and seedling data sources: (Li et al. 2020).

<u>Zygote experienced major change in miRNA expression and produced *de novo* miRNAs</u>

Our previous study found that the rice zygote expressed *de novo* mRNA transcripts before the first zygotic division (Anderson et al. 2017), indicating that plant zygotic genome activation (ZGA) occurs in the zygote. *De novo* mRNA expression was also observed in maize and *Arabidopsis* zygotes (Chen et al. 2017; Zhao et al. 2019). However, whether the zygote has started to produce miRNAs, which are known regulators of gene expression, has not been explored in plants. It was found that the global mRNA expression of zygote clustered with the egg cell, rather than with sperm cell or the average of the gametes (Anderson et al. 2013; 2017), and thus we hypothesized that the global expression of zygote miRNAs will be similar to that of the egg cell. However, that was not the case. We performed a principal component analysis (PCA) on miRNA expression pattern (values for each miRNA listed in **Supplemental Dataset 1**). We found that the first PC axis explained ~55% of variance in our dataset, corresponding to the global differences between seedling and gametes, whereas the second PC axis explained

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

~14% of the variances, corresponding to the global differences between egg cell and sperm cell. Contrary to our prediction that the zygote would cluster with the egg cell, the global expression of zygote miRNAs was about mid-way in between the egg and seedling (**Fig 3A**). This observation can be explained by a major change in expression of miRNAs between the egg cell and the zygote. For example, miR159a.1/b, family members of miR159, which were the most abundant miRNAs in the egg cell and ovary, accounted for ~70% of all miRNA reads in these tissues (Li et al. 2020); in contrast, in the zygote, the relative abundances of miR159a.1/b decreased to ~34% of all miRNA reads in the zygote, which is closer to ~25% in seedling and sperm cell, albeit still abundant in the zygote (**Fig 3B**). The high expression of miR159a.1/b raised concerns on whether they dominated the principal component analysis. This was not the case. A principal component analysis excluding expression values for miRNA159a.1/b returned the same results (**Fig S7A**). Moreover, miRNAs higher expressed in the zygote contributed in displacing zygote libraries to the left along PC1, and miRNAs higher expressed in the egg contributed in displacing egg libraries to the right along PC1 (**Fig S7B**, r = -0.62, P = 3e-34, Pearson's correlation). Taken together, these results suggested that miRNA expression is dynamic in unicellular zygotes.

We also looked at *de novo* expressed miRNAs, which were defined as miRNAs that were lowly expressed in the egg cell (< 1 read per million siRNA reads) but upregulated in the zygote **(Fig. 3C)**. Since top sperm-enriched miRNAs were all downregulated by orders of magnitude in the zygote (**Fig S8A**), we concluded that miRNA carryover from the sperm cell is very limited. Thus, we defined *de novo* miRNAs as zygote-expressed miRNAs that were lowly expressed in egg cell, instead of lowly expressed in both gametes. We detected 18 miRNAs representing 11 miRNA families (from 314 expressed miRNAs in our dataset) that met this criterion (**Fig 3C**). These miRNAs were detected as differentially expressed between zygote and egg under a log2FC > 1, FDR < 0.05 cutoff, and they all had near undetectable expression in all six replicates of egg cell. All 18 were also lowly expressed in sperm, further confirming sperm small RNA contribution was limited. These *de novo* expressed miRNAs showed that the zygote has started to express new miRNAs before the first zygotic division, much like the case of mRNAs (Anderson et al. 2017) during plant ZGA.

Upregulation of *de novo* miRNAs in the zygote would predict the downregulation of their mRNA targets. Using published degradome data (Zhou et al. 2010) and 5' RACE validated

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

targets (Liu et al. 2009) for rice miRNAs, together with published mRNA transcriptomes from rice gametes and zygotes (Anderson et al. 2013; 2017), we examined the expression of predicted targets of miRNAs (**Fig S8B**). We found that targets of miR393 family members, *auxin F-box 2* (*AFB2*, Os04g32460) and *TIR1* (Os05g05800), were downregulated in the zygote (**Fig S8B**). In addition, a target of miR171, *Hairy Meristem 3* (*HAM3*, Os04g46860), a GRAS family transcription factor was also downregulated in the zygote. In *Arabidopsis*, triple mutant of *ham3* and its paralogs *ham1 ham2* exhibited abnormal floral meristem initiation (Engstrom et al. 2011). Two other *HAM*-like genes (Os02g44360, Os02g44370), both detected as miR171 targets by a published degradome (Zhou et al. 2010), were also downregulated in the zygote. Other predicted miRNA targets were not downregulated in the zygote; these targets were either lowly expressed in the egg cell to begin with, or had variances in expression levels that were too high to confidently call downregulation (**Fig S8B**). Lastly, we previously showed that RNA contamination from surrounding sporophytic tissue in the egg cell was minimal (Li et al. 2020). Using similar analyses, we detected five miRNAs that were highly expressed in pre-anthesis ovary but nearly undetectable in all six replicates of zygote, suggesting that RNA contamination from surrounding sporophytic tissue during zygote isolation is negligible (**Fig S8C**).

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

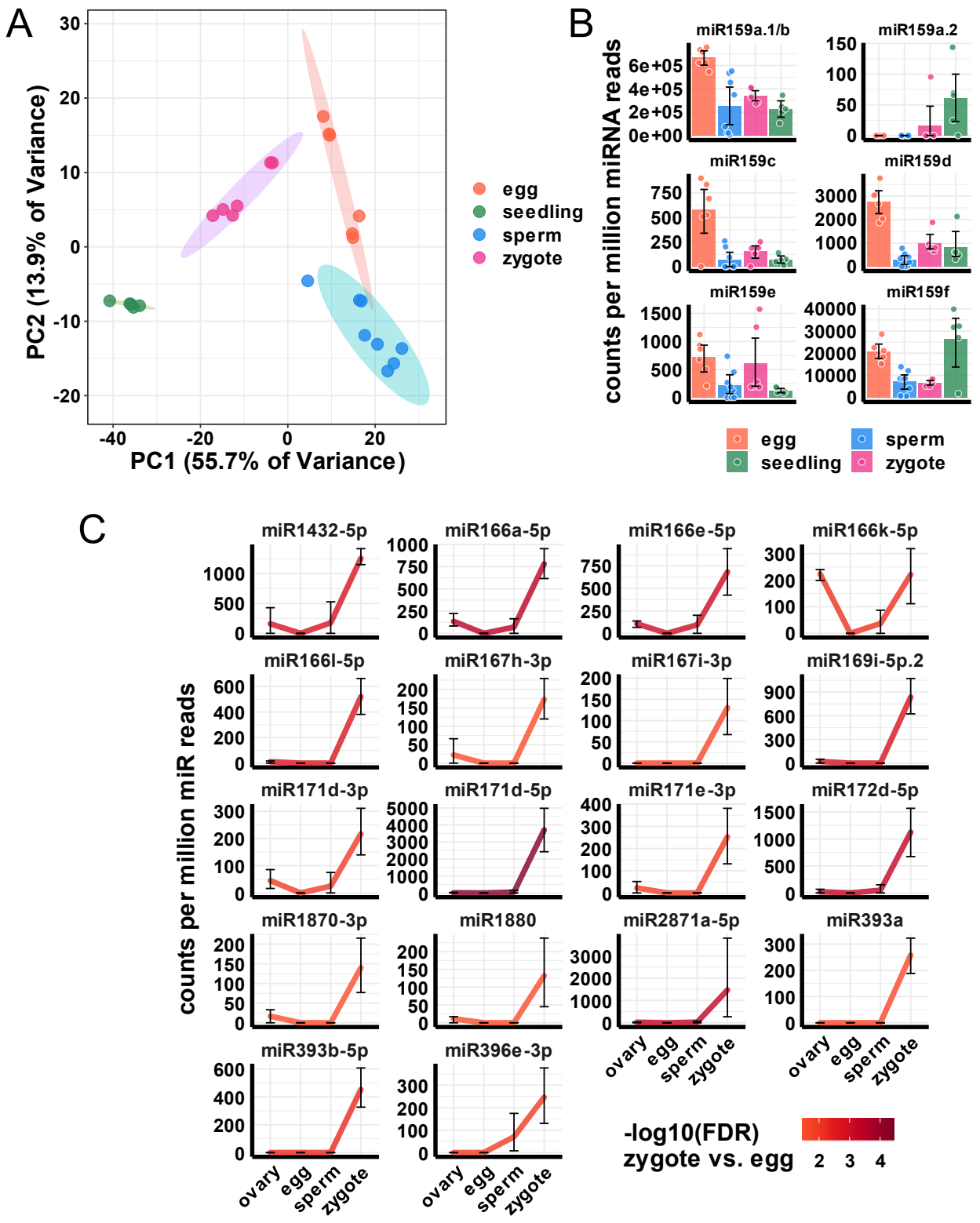


Fig 3 miRNA expression pattern in zygote

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(A) Principal component analysis (PCA) of miRNA expression. Ellipses are 90% confidence intervals for each tissue type.
(B) miR159 family expression. y-axis values are relative to total number of miRNA reads in each tissue. Error bars are 95% confidence intervals based on biological replicates of each tissue.
(C) *De novo* miRNA expression in zygote. *De novo* miRNAs are defined as miRNAs with less than 1 reads per million small RNA reads in the egg cell and significantly upregulated in the zygote. Significant upregulation is determined by log2FC > 1 between zygote and egg and FDR (false discovery rate) < 0.05 cutoffs. y-axis values are relative to total number of miRNA reads in each tissue. Color of lines is based on -log10 values of FDR between zygote and egg. Error bars are 95% confidence intervals based on biological replicates of each tissue.
Except zygotes, all data are from (Li et al. 2020).

<u>Zygote *de novo* siRNA loci had characteristics of canonical vegetative siRNA loci</u>

Since genome-wide reprogramming of siRNA loci occurs in gametes whereas later-stage (7-8 days after flowering) immature embryos resemble seedlings (Rodrigues et al. 2013, Li et al. 2020), it is expected that the siRNA transcriptome resets to the canonical siRNA profile during embryogenesis. To test whether such a resetting process may be initiated in the zygote, we first determined the *de novo* siRNA loci expression in zygotes. First, we identified siRNA loci from seedling, gametes, and zygote using Shortstack (Axtell 2013) and kept only 24-nt-dominant loci. We then identified zygote siRNA loci that did not overlap with any egg siRNA loci. Since transcript carryover from the sperm cell is minimum (**Fig 1D**, **Fig S8A**), *de novo* siRNA loci were defined as zygote NOT egg siRNA loci, consistent with the standard definition of *de novo* mRNA loci in zygotes (Chen et al. 2017; Anderson et al. 2017; Zhao et al. 2019).

The genomic distribution of zygote *de novo* loci mirrored seedling siRNA loci and gene density, unlike zygote/egg or zygote/sperm intersection loci (**Fig 4A**). Zygote NOT gametes loci, defined as zygote siRNA loci that did not overlap either egg siRNA loci or sperm siRNA loci, can be viewed as a more stringent subset of zygote *de novo* loci, and these showed the same trend as zygote *de novo* loci. Zygote NOT gametes loci and zygote *de novo* loci were largely equivalent, as zygote NOT gametes loci constituted 80% of zygote *de novo* loci (**Fig 4B**, 14993/18467 loci). A feature of the canonical siRNA pattern is that 24-nt siRNAs are enriched near genes, as exemplified by the distribution of seedling loci closely matching that of genes (**Fig**

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

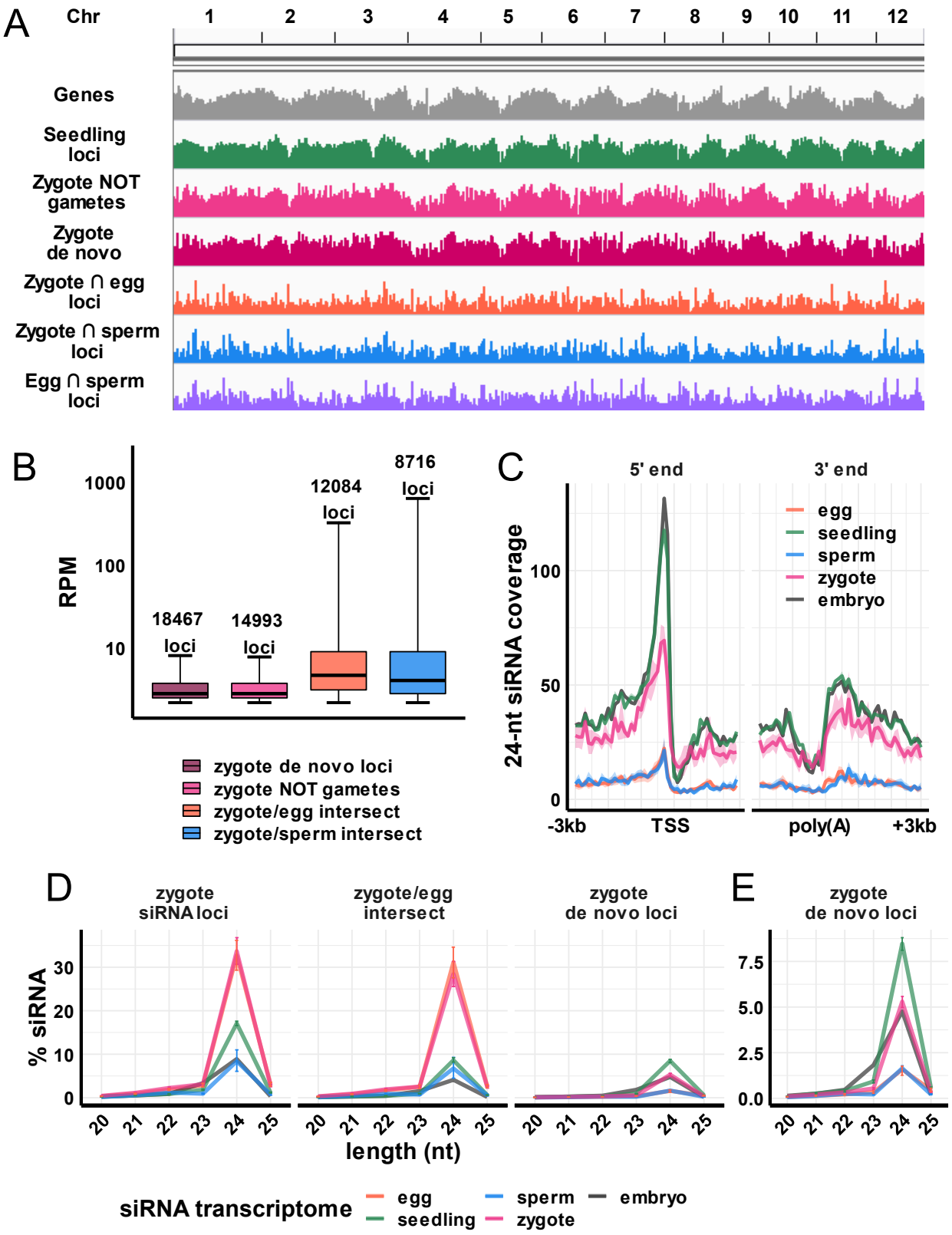
4A). We found that 84% of the zygote *de novo* loci overlapped seedling siRNA loci (**Fig S9A**). In fact, zygote/seedling intersection loci were enriched among zygote *de novo* loci (**Fig S9B**, *P* = 1.3e-34, Fisher Exact test). Hierarchical clustering also showed that the genomic distribution of zygote *de novo* loci was more closely related to that of genes or seedling siRNA loci than zygote/egg or zygote/sperm intersection loci (**Fig S9C**). We found that 4% of zygote *de novo* loci overlapped sperm siRNA loci and not seedling siRNA loci. However, these loci were unlikely the results of siRNA carryover from sperm cell for the following reasons. First, there was a lack of correlation between expression levels in sperm and zygote (**Fig S9D**). Second, carryover would predict these are among the highest expressed loci in sperm, which was not the case, as the vast majority of them (98%) were not among the most highly expressed siRNA loci in sperm (**Fig S9D**). Although spread across all 12 rice chromosomes, the siRNA abundances at zygote *de novo* loci were overall low (**Fig 4B**), only accounting for ~14% of siRNA reads in zygote. This is consistent with the data that overall, the zygote siRNA landscape is similar to that of the egg cell. Nonetheless, the zygote has started to produce siRNAs that were undetected in the egg cell before the first zygotic division, in addition to producing *de novo* miRNAs. Further, since the distribution of zygote *de novo* loci coincided with seedling loci and gene density, siRNA expression in zygote had started to return to the stable canonical siRNA profile, that is maintained in post-germination development.

We produced metagene plots for siRNA reads overlapping zygote *de novo* loci. Consistent with the genome-wide distribution, in the zygote, there was a noticeable peak upstream of TSS, resembling the profile seen in seedling (**Fig 4C**). Immature embryo (Rodrigues et al. 2013), like seedling, also had a prominent peak upstream of TSS, suggesting that during embryogenesis, the siRNA landscape redistributes from the gametic profile back to the canonical profile, and such a process has begun in the zygote. We also quantified the relative abundances of siRNAs at total zygote siRNA loci, zygote/egg intersection loci, and zygote *de novo* loci. First, we found that zygote and egg had near identical relative abundances across total zygote loci and zygote/egg intersection loci (**Fig 4D**), which was expected given siRNA carryover from the egg cell (**Fig 1A**). Second, we found that zygote had low siRNA abundance at zygote *de novo* loci, which was again expected given that zygote *de novo* loci were low expressors of siRNAs (**Fig 4B**). However, we also found that zygote *de novo* loci were not zygote specific. Rather, seedling and immature embryo also had siRNAs arising from zygote *de novo* loci, and

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

seedling had an even higher relative abundance of siRNAs at zygote *de novo* loci than the zygote itself (**Fig 4D – E**). Lastly, we detected 12% of the zygote *de novo* loci to be zygote specific, not overlapping gamete siRNA loci or seedling siRNA loci (**Fig S9A**). This small subset of loci accounted for only less than 1% of siRNAs in zygote, and even fewer in embryo (**Fig S9E**), and thus they may have represented siRNAs expressed transiently in the zygote.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



421

422 **Fig 4 Characteristics of zygote *de novo* loci**

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(A) Whole genome distribution of siRNA loci. Top track: gene density; second to seventh tracks: seedling siRNA loci, zygote NOT gametes loci, zygote *de novo* loci, zygote/egg siRNA loci intersection, zygote/sperm siRNA loci intersection, and egg/sperm siRNA loci intersection. Zygote NOT gametes loci are zygote siRNA loci that do not overlap any egg cell siRNA loci or sperm cell siRNA loci. Zygote *de novo* loci are zygote siRNA loci that do not overlap any egg cell siRNA loci. Height of bars represent number of loci in non-overlapping 50-kb windows normalized by total number of loci for each category. Chr: chromosomes.

(B) Box plot showing expression of siRNA at zygote NOT gametes loci, zygote *de novo* loci, zygote/egg siRNA loci intersection, and zygote/sperm siRNA loci intersection. RPM: reads per million. Center line is median; box spans interquartile range; whiskers span 2.5th to 97.5th percentiles. y-axis is log10 transformed.

(C) Metagene coverage plot of 24-nt siRNA reads that overlap zygote *de novo* loci across all genes in the genome. Coverage is measured over 100-bp intervals and normalized per 1000 total siRNAs. Vertical grid lines indicate 500-bp intervals. Ribbons are 95% confidence intervals for each tissue type. TSS transcription start site, poly(A) polyadenylation site.

(D) siRNA relative abundances at all zygote siRNA loci, zygote/egg siRNA loci intersection, and zygote *de novo* loci. y-axis values are % of siRNAs of each length relative to the total number of siRNAs that mapped to the genome. Error bars are 95% confidence intervals based on biological replicates of each tissue.

(E) Enlarged plot of zygote *de novo* loci panel from (D).

Embryo data sources: (Rodrigues et al. 2013). Egg, sperm and seedling data sources: (Li et al. 2020).

We further tested the hypothesis that zygote *de novo* loci are indeed similar to canonical siRNA loci by comparing the distance to the nearest genes and proportion of loci overlapping different transposable element superfamilies. We found that like seedling siRNA loci, which had a median distance of 966-bp to their nearest genes, zygote *de novo* loci had a median distance of just 1.11-kb (**Fig 5A**), whereas egg siRNA loci and sperm siRNA loci were on average much further away from their nearest genes, with median distances of 1.54-kb and 4.06-kb, respectively. In addition, canonical siRNA loci tend to overlap with TIR DNA transposons rather than *Gypsy* retrotransposons, since TIR transposons are enriched near genes (Han, Qin, and

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Wessler 2013), while *Gypsy* retrotransposons are enriched in gene-poor pericentromeric regions (Kawahara et al. 2013). As expected, 10% of the zygote *de novo* loci overlapped a *Gypsy*, and 42% overlapped a TIR element, consistent with a shift towards seedling siRNA loci (8.8% and 63%, respectively, **Fig 5B**). In contrast, 19% and 49% of the egg and sperm siRNA loci overlapped with a *Gypsy* element, consistent with farther distances to their nearest genes. These overlaps were defined by overlapping a TIR or *Gypsy* element with at least 1-bp, which was very stringent. In a supporting analysis, we used a relaxed cutoff, defining overlap as overlapping a TIR or *Gypsy* element with at least 33% of the length of the locus, which produced the same pattern (**Fig S10**). Taken together, both lines of evidence support that the reestablishment of canonical siRNA loci has initiated in the zygote.

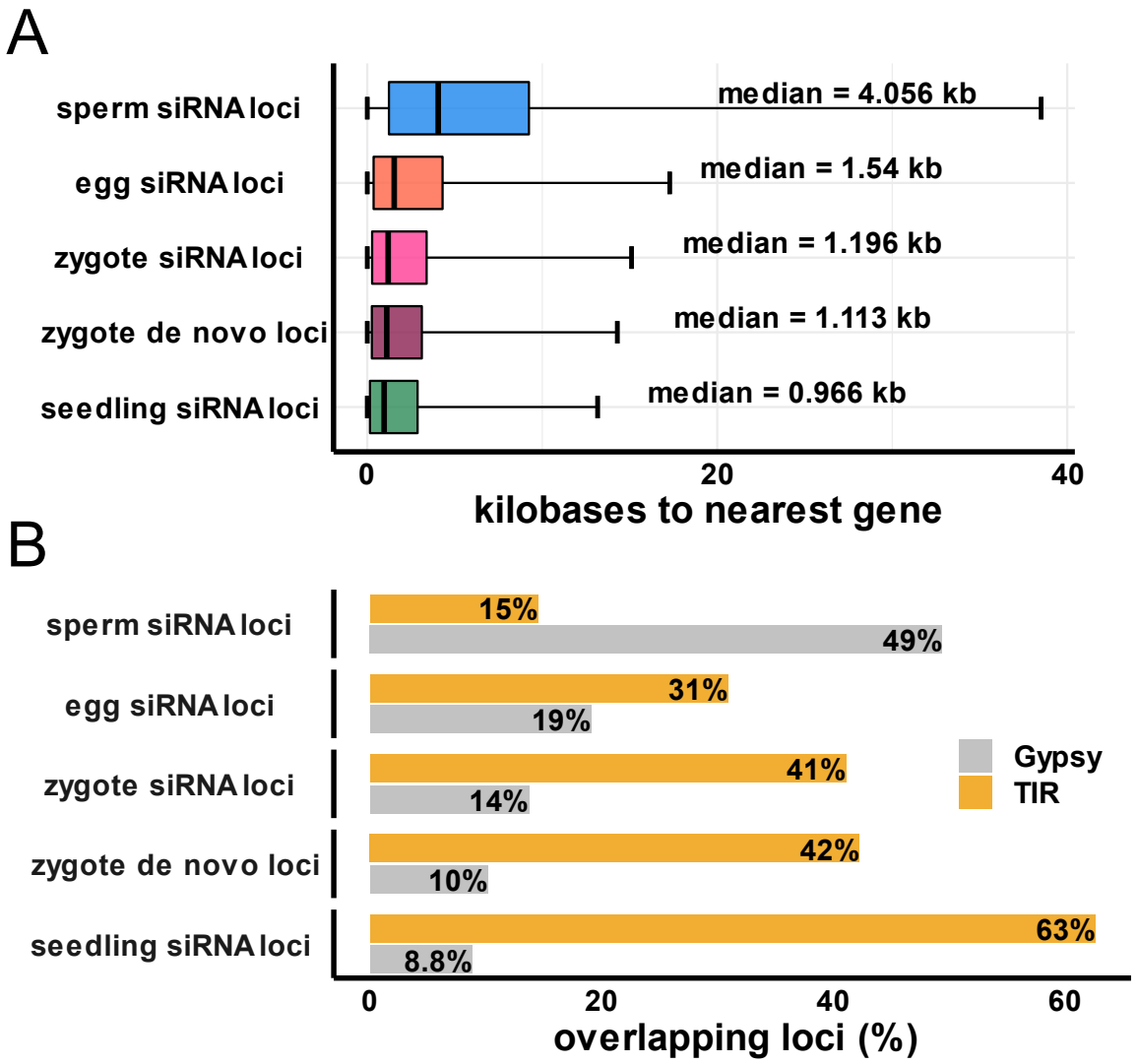


Fig 5 Zygote *de novo* loci are similar to seedling siRNA loci

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(A) Box plot showing distance from siRNA loci to respective nearest or overlapping genes. Center line is median; box spans interquartile range; whiskers span 2.5th to 97.5th percentiles. (B) Bar plot showing percentage of siRNA loci overlapping a TIR transposon or *Gypsy* retrotransposon. Overlap is defined by overlapping a TIR or Gypsy element with at least 1-bp. Except zygotes, all data are from (Li et al. 2020).

Discussion

We report here the first small RNA transcriptome of plant unicellular zygotes, which provides insights into epigenome reprogramming during plant reproduction. Overall, the siRNA transcriptome of zygote is similar to that of the egg cell **(Fig 1A – C, Fig S1B**, summarized in **Fig 6**), which can be explained by transcript carryover from the egg cell and dilution of sperm cell siRNAs, since the egg cell is ~1000-fold larger than the sperm cell, and that 24-nt siRNAs are predominantly found in the cytoplasm (Ye et al. 2012). This is consistent with the fact that transcript carryover from the sperm cell is limited, as the zygote had few siRNAs overlapping sperm-specific siRNA loci (**Fig 1D**), and the most highly expressed miRNA in the sperm cell were orders of magnitude downregulated in the zygote (**Fig S8A**).

Similar to *Brassica rapa* and *Arabidopsis* ovules (Grover et al. 2020), a small number of loci accounted for most of the siRNA in zygote (**Fig 2A**), 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 2B – D**). It should be noted, however, that endosperm nuclei divide rapidly, massive changes in cellular organization occur over a short time, and endosperm DNA undergoes active DNA demethylation [reviewed in (Gehring and Satyaki 2017)]. It is quite possible that the central cell and earlier stages of endosperm development have an siRNA transcriptome more like that of the zygote. It has been proposed that the embryo receives siRNAs from the endosperm (Martinez 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

development (Grover et al. 2018; 2020). Likewise, it is possible that siren siRNA in the egg cell and zygotes are produced in the ovary tissue instead. As ovary siren loci have higher mCHH in ovary and the egg cell (**Fig S3C**), siren siRNA may serve biological functions in these tissues. Since these siren siRNA persisted in the zygote, mCHH may also be higher at both maternal and paternal alleles in the zygote, which may suggest a functional significance of these siren siRNA post-fertilization.

Unlike the siRNA transcriptome, which was globally similar to the egg cell, there was a global change in miRNAs associated with the zygotic transition (**Fig 3A**), which included the downregulation of miRNAs, as well as the expression of *de novo* miRNAs (**Fig 3B – C**, **Fig S7**). Eighteen miRNAs that were undetected in the egg cell were significantly upregulated in the zygote (**Fig 3C**), implying that before the first embryonic division the expression of miRNAs as well as mRNAs (Anderson et al. 2017) from the zygotic genome is a feature of zygotic genome activation in plants. The predicted targets of these miRNAs include auxin receptors (*AFB2* and *TIR1*) that we speculate might be involved in the fine tuning of auxin signaling, and meristem related factors (*HAM* transcription factors), whose downregulation may be involved in establishing zygotic identity (**Fig S8B**). However, it remains to be investigated whether these differentially expressed miRNAs and their targets have significant functions in ZGA and embryogenesis.

The observation that the distribution of *de novo* siRNA loci in the zygote mirrored canonical siRNA loci (**Fig 4A**, **Fig S9A – C**) suggests that the transition to the canonical siRNA transcriptome is initiated in the zygote (summarized in **Fig 6**). This is further supported by a few lines of evidence. 1) zygote siRNAs from zygote *de novo* loci, like seedling and embryo siRNAs, had a noticeable peak upstream of TSS (**Fig 4C**), and such a peak was minimal in gametes. 2) Zygote *de novo* loci had similar characteristics as canonical siRNA loci, such as proximity to genes and tendency to overlap TIR rather than *Gypsy* transposable elements (**Fig 5**). 3) The properties of zygote de novo loci are not an artefact of how these loci were defined, as zygote overall had more gene-proximal siRNAs (**Fig 1C**) and more siRNAs overlapping seedling-specific siRNA loci (**Fig 1D**). Zygote siRNA loci overall were closer to genes (**Fig 5A**) and tended to overlap more TIR DNA transposons but fewer *Gypsy* retrotransposons (**Fig 5B**) when compared to the egg cell. Lastly, 4) Most zygote de novo loci were not zygote specific. Seedling had even higher siRNA abundance at zygote de *novo* loci (**Fig 4D – E**). All the above results

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

suggest that the resetting of siRNA transcriptome is initiated in the zygote and is completed by the time seedling emerges (**Fig 6**).

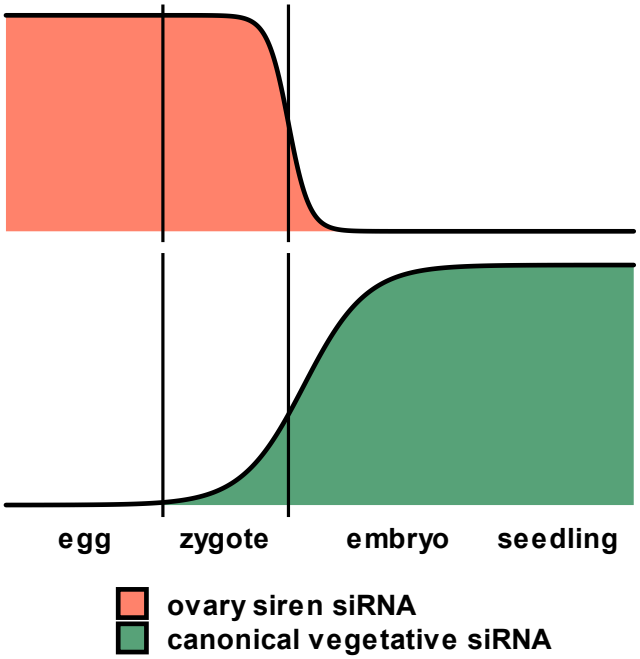


Fig 6 Schematic summary of developmental progression of the zygote siRNA transcriptome
Schematic (non-quantitative) representation of 24-nt siRNA abundance in egg cell, zygote, embryo, and seedling. Height of curves represent abundances of siRNA. Sperm cell siRNA loci are not included due to their limited siRNA contribution in the zygote.

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 unicellular rice zygote, we inferred that the resetting to canonical siRNA transcriptome is initiated along with the zygotic transition, soon after fertilization, which is consistent with previous observations in *Arabidopsis* that the resetting of H3 variants also occurs in the zygote before the first cell division (Ingouff et al. 2007; 2010). Resetting of the siRNA transcriptome after fertilization would predict a lack of

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transgenerational siRNA inheritance in plants. However, genome-wide reprogramming of siRNAs does not preclude the persistence and inheritance of selected siRNA molecules. Therefore, transgenerational siRNA inheritance in plants is not excluded by our data. 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 epigenome, 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 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 (**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. All libraries were amplified at 24 cycles. The library product was size selected using PippinHT (Sage Science) 3% agarose gel cassettes.

Small RNA sequencing analysis

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Analyses were based on the Os-Nipponbare-Reference-IRGSP-1.0 reference genome (Kawahara et al. 2013). 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, Robinson, and Mesirov 2013). For better visualization of midrange values, heatmap intensity was maxed out at 1.25X coverage (per 10 million 24-nt siRNAs).

<u>miRNA expression analysis</u>

To measure miRNA expression, the BEDTools coverage tool was used to count the number of 20-25nt reads that overlapped at least 90% of their length with annotated miRNA positions (**Supplemental Dataset 1)**. R package EdgeR was used to analyze miRNA expression (McCarthy, Chen, and Smyth 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

ChIP-seq data analyses

Chromatin Immunoprecipitation (ChIP) data were obtained from the NCBI Sequence Reads Archive, accession numbers listed in **Supplemental Table 2**. Reads were aligned to the Os-Nipponbare-Reference-IRGSP-1.0 reference genome using Bowtie2 (Langmead and Salzberg 2012) default parameters, except the --trim-to parameter was set to 101 for all paired-end reads (from Tan et al 2016, Liu et al 2018, Lu et al 2020), to 36 for a subset of single-end reads (from Zhang et al 2020), and to 49 for the other single-end reads (Lu et al 2015, Zahraeifard et al 2018, Maher et al 2018). Read coverage was calculated for each set of loci using the BEDTools coverage tool. 2000 loci were randomly subsampled from each set to reduce memory requirements, except for the ovary siren loci that only had 818 total without subsampling. Principal component analysis was done using the log-transformed values of median of each locus category for each ChIP-seq library.

Definition of siren loci and zygote *de novo* loci

Small RNA loci were identified from the initial 20-25nt total small RNA alignment bam files using Shortstack (Axtell 2013) after merging replicates using default parameters. For each tissue type (ovary, egg cell, sperm cell, zygote, seedling shoot and endosperm), siRNA loci were defined as RPM > 2 (default), 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, ovary 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 2A**). Zygote *de novo* loci were identified by the BEDTools intersect tool (Quinlan and Hall 2010). Zygote *de novo* loci were defined as zygote NOT egg loci (zygote siRNA loci that did not overlap any egg cell siRNA loci). To visualize the genomic distribution of siRNA loci, bed files of loci were imported into IGV and visualized across the whole genome.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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, Sarah N., Cameron S. Johnson, Joshua Chesnut, Daniel S. Jones, Imtiyaz Khanday, Margaret Woodhouse, Chenxin Li, Liza J. Conrad, Scott D. Russell, and Venkatesan Sundaresan. 2017. “The Zygotic Transition Is Initiated in Unicellular Plant Zygotes with

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Asymmetric Activation of Parental Genomes." *Developmental Cell* 43 (3): 349-358.e4. https://doi.org/10.1016/j.devcel.2017.10.005.

Anderson, Sarah N., Cameron S. Johnson, Daniel S. Jones, Liza J. Conrad, Xiaoping Gou, Scott D. Russell, and Venkatesan Sundaresan. 2013. "Transcriptomes of Isolated Oryza Sativa Gametes Characterized by Deep Sequencing: Evidence for Distinct Sex-Dependent Chromatin and Epigenetic States before Fertilization." *The Plant Journal* 76 (5): 729–41. https://doi.org/10.1111/tpj.12336.

Armenta-Medina, Alma, and C. Stewart Gillmor. 2019. "Genetic, Molecular and Parent-of-Origin Regulation of Early Embryogenesis in Flowering Plants." In *Current Topics in Developmental Biology*, 131:497–543. Elsevier. https://doi.org/10.1016/bs.ctdb.2018.11.008.

Autran, Daphné, Célia Baroux, Michael T. Raissig, Thomas Lenormand, Michael Wittig, Stefan Grob, Andrea Steimer, et al. 2011. "Maternal Epigenetic Pathways Control Parental Contributions to Arabidopsis Early Embryogenesis." *Cell* 145 (5): 707–19. 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 (6): 740–51. https://doi.org/10.1261/rna.035279.112.

Borg, Michael, and Frédéric Berger. 2015. "Chromatin Remodelling during Male Gametophyte Development." *The Plant Journal* 83 (1): 177–88. https://doi.org/10.1111/tpj.12856.

Borg, Michael, Yannick Jacob, Daichi Susaki, Chantal LeBlanc, Daniel Buendía, Elin Axelsson, Tomokazu Kawashima, et al. 2020. "Targeted Reprogramming of H3K27me3 Resets Epigenetic Memory in Plant Paternal Chromatin." *Nature Cell Biology* 22 (6): 621–29. https://doi.org/10.1038/s41556-020-0515-y.

Borges, Filipe, Jean-sébastien Parent, Frédéric Van Ex, Philip Wolff, German Martínez, Claudia Köhler, and Robert A Martienssen. 2018. "Transposon-Derived Small RNAs Triggered by MiR845 Mediate Genome Dosage Response in Arabidopsis." *Nature Genetics* 50 (February). https://doi.org/10.1038/s41588-017-0032-5.

Calarco, Joseph P., Filipe Borges, Mark T.A. Donoghue, Frédéric Van Ex, Pauline E. Jullien, Telma Lopes, Rui Gardner, et al. 2012. "Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA." *Cell* 151 (1): 194–205. https://doi.org/10.1016/j.cell.2012.09.001.

Chen, Junyi, Nicholas Strieder, Nadia G. Krohn, Philipp Cyprys, Stefanie Sprunck, Julia C. Engelmann, and Thomas Dresselhaus. 2017. "Zygotic Genome Activation Occurs Shortly after Fertilization in Maize." *The Plant Cell* 29 (9): 2106–25. https://doi.org/10.1105/tpc.17.00099.

Cuerda-Gil, Diego, and R. Keith Slotkin. 2016. "Non-Canonical RNA-Directed DNA Methylation." *Nature Plants* 2 (11): 16163. https://doi.org/10.1038/nplants.2016.163.

Ding, Jianting, Jiaheng Shen, Wei Li, and Hong Yang. 2009. "Cytological Observation of Double Fertilization and Its Duration in Oryza Sativa." *Chi n Bull Bot*, no. 44: 473–83.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Diego Cuerda-Gil, and R Keith Slotkin. 2016. "Non-Canonical RNA-Directed DNA Methylation," no. November. https://doi.org/10.1038/NPLANTS.2016.163.

Dorweiler, Jane E, Charles C Carey, Kenneth M Kubo, Jay B Hollick, Jerry L Kermicle, and Vicki L Chandler. 2000. "Mediator of Paramutation1 Is Required for Establishment and Maintenance of Paramutation at Multiple Maize Loci," 19.

Engstrom, Eric M., Carl M. Andersen, Juliann Gumulak-Smith, John Hu, Evguenia Orlova, Rosangela Sozzani, and John L. Bowman. 2011. "Arabidopsis Homologs of the *Petunia HAIRY MERISTEM* Gene Are Required for Maintenance of Shoot and Root Indeterminacy." *Plant Physiology* 155 (2): 735–50. https://doi.org/10.1104/pp.110.168757.

Erdmann, Robert M, Prasad R V Satyaki, Maja Klosinska, Mary Gehring, Robert M Erdmann, Prasad R V Satyaki, Maja Klosinska, and Mary Gehring. 2017. "A Small RNA Pathway Mediates Allelic Dosage in A Small RNA Pathway Mediates Allelic Dosage in Endosperm." *CellReports* 21 (12): 3364–72. https://doi.org/10.1016/j.celrep.2017.11.078.

Gehring, Mary. 2019. "Epigenetic Dynamics during Flowering Plant Reproduction: Evidence for Reprogramming?" *New Phytologist* 224 (1): 91–96. https://doi.org/10.1111/nph.15856.

Gehring, Mary, and P.R. Satyaki. 2017. "Endosperm and Imprinting, Inextricably Linked." *Plant Physiology* 173 (1): 143–54. https://doi.org/10.1104/pp.16.01353.

Gent, J. I., N. A. Ellis, L. Guo, A. E. Harkess, Y. Yao, X. Zhang, and R. K. Dawe. 2013. "CHH Islands: De Novo DNA Methylation in near-Gene Chromatin Regulation in Maize." *Genome Research* 23 (4): 628–37. https://doi.org/10.1101/gr.146985.112.

Gent, Jonathan I, Nathanael A Ellis, Lin Guo, Alex E Harkess, Yingyin Yao, Xiaoyu Zhang, and R Kelly Dawe. 2013. "CHH Islands : De Novo DNA Methylation in near-Gene Chromatin Regulation in Maize," 628–37. https://doi.org/10.1101/gr.146985.112.as.

Grover, Jeffrey W., Diane Burgess, Timmy Kendall, Abdul Baten, Suresh Pokhrel, Graham J. King, Blake C. Meyers, Michael Freeling, and Rebecca A. Mosher. 2020. "Abundant Expression of Maternal SiRNAs Is a Conserved Feature of Seed Development." *Proceedings of the National Academy of Sciences*, June, 202001332. https://doi.org/10.1073/pnas.2001332117.

Grover, Jeffrey W., Timmy Kendall, Abdul Baten, Diane Burgess, Michael Freeling, Graham J. King, and Rebecca A. Mosher. 2018. "Maternal Components of RNA -directed DNA Methylation Are Required for Seed Development in *Brassica Rapa*." *The Plant Journal* 94 (4): 575–82. https://doi.org/10.1111/tpj.13910.

Han, Yujun, Shanshan Qin, and Susan R Wessler. 2013. "Comparison of Class 2 Transposable Elements at Superfamily Resolution Reveals Conserved and Distinct Features in Cereal Grass Genomes." *BMC Genomics* 14 (1): 71. https://doi.org/10.1186/1471-2164-14-71.

Houri-Zeevi, Leah, and Oded Rechavi. 2017. "A Matter of Time: Small RNAs Regulate the Duration of Epigenetic Inheritance." *Trends in Genetics* 33 (1): 46–57. https://doi.org/10.1016/j.tig.2016.11.001.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hsieh, Ping-hung, Shengbo He, Toby Buttress, Hongbo Gao, Matthew Couchman, and Robert L Fischer. 2016. "Arabidopsis Male Sexual Lineage Exhibits More Robust Maintenance of CG Methylation than Somatic Tissues." *PNAS* 113 (52). https://doi.org/10.1073/pnas.1619074114.

Hsieh, T.-F., C. A. Ibarra, P. Silva, A. Zemach, L. Eshed-Williams, R. L. Fischer, and D. Zilberman. 2009. "Genome-Wide Demethylation of Arabidopsis Endosperm." *Science* 324 (5933): 1451–54. https://doi.org/10.1126/science.1172417.

Ibarra, C. A., X. Feng, V. K. Schoft, T.-F. Hsieh, R. Uzawa, J. A. Rodrigues, A. Zemach, et al. 2012. "Active DNA Demethylation in Plant Companion Cells Reinforces Transposon Methylation in Gametes." *Science* 337 (6100): 1360–64. https://doi.org/10.1126/science.1224839.

Ingouff, Mathieu, Yuki Hamamura, Mathieu Gourgues, and Tetsuya Higashiyama. 2007. "Distinct Dynamics of HISTONE3 Variants between the Two Fertilization Products in Plants." *Current Biology*, 1032–37. https://doi.org/10.1016/j.cub.2007.05.019.

Ingouff, Mathieu, Svenja Rademacher, Sarah Holec, Nie Xin, Anne Readshaw, and Shi Hui Foo. 2010. "Report Zygotic Resetting of the HISTONE 3 Variant Repertoire Participates in Epigenetic Reprogramming in Arabidopsis." *Current Biology*, 2137–43. https://doi.org/10.1016/j.cub.2010.11.012.

Kao, Ping, and Michael D. Nodine. 2019. "Transcriptional Activation of Arabidopsis Zygotes Is Required for Initial Cell Divisions." *Scientific Reports* 9 (1): 17159. https://doi.org/10.1038/s41598-019-53704-2.

Kawahara, Yoshihiro, Melissa de la Bastide, John P Hamilton, Hiroyuki Kanamori, W Richard McCombie, Shu Ouyang, David C Schwartz, et al. 2013. "Improvement of the Oryza Sativa Nipponbare Reference Genome Using next Generation Sequence and Optical Map Data." *Rice* 6 (1): 4. https://doi.org/10.1186/1939-8433-6-4.

Kim, M. Yvonne, Akemi Ono, Stefan Scholten, Tetsu Kinoshita, Daniel Zilberman, Takashi Okamoto, and Robert L. Fischer. 2019. "DNA Demethylation by ROS1a in Rice Vegetative Cells Promotes Methylation in Sperm." *Proceedings of the National Academy of Sciences* 116 (19): 9652–57. https://doi.org/10.1073/pnas.1821435116.

Kimmins, Sarah, and Paolo Sassone-corsi. 2005. "Chromatin Remodelling and Epigenetic Features of Germ Cells." *Nature*, no. Box 1: 583–89.

Kozomara, Ana, Maria Birgaoanu, and Sam Griffiths-Jones. 2019. "MiRBase: From MicroRNA Sequences to Function." *Nucleic Acids Research* 47 (D1): D155–62. https://doi.org/10.1093/nar/gky1141.

Kranz, E., J. Bautor, and H. Lörz. 1991. "In Vitro Fertilization of Single, Isolated Gametes of Maize Mediated by Electrofusion." *Sexual Plant Reproduction* 4 (1). https://doi.org/10.1007/BF00194565.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Langmead, Ben, and Steven L Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Li, Chenxin, Hengping Xu, Fang-Fang Fu, Scott D. Russell, Venkatesan Sundaresan, and Jonathan I. Gent. 2020. "Genome-Wide Redistribution of 24-Nt SiRNAs in Rice Gametes." *Genome Research* 30 (2): 173–84. https://doi.org/10.1101/gr.253674.119.

Li, Chenxin, Hengping Xu, Scott D. Russell, and Venkatesan Sundaresan. 2019. "Step-by-Step Protocols for Rice Gamete Isolation." *Plant Reproduction* 32 (1): 5–13. https://doi.org/10.1007/s00497-019-00363-y.

Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform." *Bioinformatics* 25 (14): 1754–60. https://doi.org/10.1093/bioinformatics/btp324.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Li, Qing, Jonathan I. Gent, Greg Zynda, Jawon Song, Irina Makarevitch, Cory D. Hirsch, Candice N. Hirsch, et al. 2015. "RNA-Directed DNA Methylation Enforces Boundaries between Heterochromatin and Euchromatin in the Maize Genome." *Proceedings of the National Academy of Sciences* 112 (47): 14728–33. https://doi.org/10.1073/pnas.1514680112.

Liu, Bing, Gang Wei, Jinlei Shi, Jing Jin, Ting Shen, Ting Ni, Wen-Hui Shen, Yu Yu, and Aiwu Dong. 2016. "SET DOMAIN GROUP 708, a Histone H3 Lysine 36-Specific Methyltransferase, Controls Flowering Time in Rice (*Oryza Sativa*)." *New Phytologist* 210 (2): 577–88. https://doi.org/10.1111/nph.13768.

Liu, Qing, Yu-Chan Zhang, Cong-Ying Wang, Yu-Chun Luo, Qiao-Juan Huang, Shao-Yu Chen, Hui Zhou, Liang-Hu Qu, and Yue-Qin Chen. 2009. "Expression Analysis of Phytohormone-Regulated MicroRNAs in Rice, Implying Their Regulation Roles in Plant Hormone Signaling." *FEBS Letters* 583 (4): 723–28. https://doi.org/10.1016/j.febslet.2009.01.020.

Lord, Elizabeth M., and Scott D. Russell. 2002. "The Mechanisms of Pollination and Fertilization in Plants." *Annual Review of Cell and Developmental Biology* 18 (1): 81–105. https://doi.org/10.1146/annurev.cellbio.18.012502.083438.

Lu, Li, Xiangsong Chen, Dean Sanders, Shuiming Qian, and Xuehua Zhong. 2015. "High-Resolution Mapping of H4K16 and H3K23 Acetylation Reveals Conserved and Unique Distribution Patterns in *Arabidopsis* and Rice." *Epigenetics* 10 (11): 1044–53. https://doi.org/10.1080/15592294.2015.1104446.

Lu, Yue, Feng Tan, Yu Zhao, Shaoli Zhou, Xiangsong Chen, Yongfeng Hu, and Dao-Xiu Zhou. 2020. "A Chromodomain-Helicase-DNA-Binding Factor Functions in Chromatin Modification and Gene Regulation." *Plant Physiology* 183 (3): 1035–46. https://doi.org/10.1104/pp.20.00453.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Lu, Yue, Qiutao Xu, Yuan Liu, Yue Yu, Zhong-Yi Cheng, Yu Zhao, and Dao-Xiu Zhou. 2018. "Dynamics and Functional Interplay of Histone Lysine Butyrylation, Crotonylation, and Acetylation in Rice under Starvation and Submergence." *Genome Biology* 19 (1): 144. https://doi.org/10.1186/s13059-018-1533-y.

Maher, Kelsey A., Marko Bajic, Kaisa Kajala, Mauricio Reynoso, Germain Pauluzzi, Donnelly A. West, Kristina Zumstein, et al. 2018. "Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules." *The Plant Cell* 30 (1): 15–36. https://doi.org/10.1105/tpc.17.00581.

Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads." *EMBnetjournal* 17: 3.

Martinez, German, and Claudia Köhler. 2017. "Role of Small RNAs in Epigenetic Reprogramming during Plant Sexual Reproduction." *Current Opinion in Plant Biology* 36 (April): 22–28. https://doi.org/10.1016/j.pbi.2016.12.006.

Martínez, Germán, Kaushik Panda, Claudia Köhler, and R Keith Slotkin. 2016. "Silencing in Sperm Cells Is Directed by RNA Movement from the Surrounding Nurse Cell." *Nature Plants* 2 (4): 1–8. https://doi.org/10.1038/nplants.2016.30.

Martinez, German, Philip Wolff, Zhenxing Wang, Jordi Moreno-romero, Juan Santos-gonzález, Lei Liu Conze, Christopher Defraia, R Keith Slotkin, and Claudia Köhler. 2018. "Paternal EasiRNAs Regulate Parental Genome Dosage in Arabidopsis." *Nature Genetics* 50 (February). https://doi.org/10.1038/s41588-017-0033-4.

Matzke, Marjori A., and Rebecca A. Mosher. 2014. "RNA-Directed DNA Methylation: An Epigenetic Pathway of Increasing Complexity." *Nature Reviews Genetics* 15 (6): 394–408. https://doi.org/10.1038/nrg3683.

McCarthy, Davis J., Yunshun Chen, and Gordon K. Smyth. 2012. "Differential Expression Analysis of Multifactor RNA-Seq Experiments with Respect to Biological Variation." *Nucleic Acids Research* 40 (10): 4288–97. https://doi.org/10.1093/nar/gks042.

Mérai, Zsuzsanna, Nina Chumak, Marcelina García-Aguilar, Tzung-Fu Hsieh, Toshiro Nishimura, Vera K. Schoft, János Bindics, et al. 2014. "The AAA-ATPase Molecular Chaperone Cdc48/P97 Disassembles Sumoylated Centromeres, Decondenses Heterochromatin, and Activates Ribosomal RNA Genes." *Proceedings of the National Academy of Sciences* 111 (45): 16166–71. https://doi.org/10.1073/pnas.1418564111.

Messerschmidt, D. M., B. B. Knowles, and D. Solter. 2014. "DNA Methylation Dynamics during Epigenetic Reprogramming in the Germline and Preimplantation Embryos." *Genes & Development* 28 (8): 812–28. https://doi.org/10.1101/gad.234294.113.

Moritoh, Satoru, Chang-Ho Eun, Akemi Ono, Hisayo Asao, Yosuke Okano, Katsushi Yamaguchi, Zenpei Shimatani, Amane Koizumi, and Rie Terada. 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Development." *The Plant Journal* 71 (1): 85–98. https://doi.org/10.1111/j.1365-313X.2012.04974.x.

Okada, Takashi, Makoto Endo, Mohan B. Singh, and Prem L. Bhalla. 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." *The Plant Journal* 44 (4): 557–68. https://doi.org/10.1111/j.1365-313X.2005.02554.x.

Park, Kyunghyuk, M. Yvonne Kim, Martin Vickers, Jin-Sup Park, Youbong Hyun, Takashi Okamoto, Daniel Zilberman, et al. 2016. "DNA Demethylation Is Initiated in the Central Cells of *Arabidopsis* and Rice." *Proceedings of the National Academy of Sciences* 113 (52): 15138–43. https://doi.org/10.1073/pnas.1619047114.

Pillot, Marion, Célia Baroux, Mario Arteaga Vazquez, Daphne Autran, Olivier Leblanc, Jean Philippe Vielle-calzada, Ueli Grossniklaus, and Daniel Grimanelli. 2010. "Embryo and Endosperm Inherit Distinct Chromatin and Transcriptional States from the Female Gametes in Arabidopsis." *Plant Cell* 22 (February): 307–20. https://doi.org/10.1105/tpc.109.071647.

Quinlan, Aaron R., and Ira M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26 (6): 841–42. https://doi.org/10.1093/bioinformatics/btq033.

Rodrigues, J. A., R. Ruan, T. Nishimura, M. K. Sharma, R. Sharma, P. C. Ronald, R. L. Fischer, and D. Zilberman. 2013. "Imprinted Expression of Genes and Small RNA Is Associated with Localized Hypomethylation of the Maternal Genome in Rice Endosperm." *Proceedings of the National Academy of Sciences* 110 (19): 7934–39. https://doi.org/10.1073/pnas.1306164110.

Saitou, M., S. Kagiwada, and K. Kurimoto. 2012. "Epigenetic Reprogramming in Mouse Pre-Implantation Development and Primordial Germ Cells." *Development* 139 (1): 15–31. https://doi.org/10.1242/dev.050849.

Satyaki, Prasad R V, and Mary Gehring. 2019. "Paternally Acting Canonical RNA-Directed DNA Methylation Pathway Genes Sensitize Arabidopsis Endosperm to Paternal." *Plant Cell* 31 (July): 1563–78. https://doi.org/10.1105/tpc.19.00047.

Schmieder, R., and R. Edwards. 2011. "Quality Control and Preprocessing of Metagenomic Datasets." *Bioinformatics* 27 (6): 863–64. https://doi.org/10.1093/bioinformatics/btr026.

Schoft, Vera K, Nina Chumak, Magdalena Mosiolek, Lucyna Slusarz, Vukoslav Komnenovic, Lynette Brownfield, David Twell, Tetsuji Kakutani, and Hisashi Tamaru. 2009. "Induction of RNA-Directed DNA Methylation upon Decondensation of Constitutive Heterochromatin." *Scientifc Report* 10 (9): 1015–21. https://doi.org/10.1038/embor.2009.152.

Scholten, Stefan, Horst Lörz, and Erhard Kranz. 2002. "Paternal MRNA and Protein Synthesis Coincides with Male Chromatin Decondensation in Maize Zygotes." *The Plant Journal* 32 (2): 221–31. https://doi.org/10.1046/j.1365-313X.2002.01418.x.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Slotkin, R Keith, Matthew Vaughn, Filipe Borges, A Feijo, D Becker, and Robert A Martienssen. 2009. "Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen." *Cell*, 461–72. https://doi.org/10.1016/j.cell.2008.12.038.

Su, Zhenxia, Nannan Wang, Zhimin Hou, Baiyang Li, Dingning Li, Yanhui Liu, Hanyang Cai, Yuan Qin, and Xuemei Chen. 2020. "Regulation of Female Germline Specification via Small RNA Mobility in Arabidopsis." *The Plant Cell*, July, tpc.00126.2020. https://doi.org/10.1105/tpc.20.00126.

Tadros, W., and H. D. Lipshitz. 2009. "The Maternal-to-Zygotic Transition: A Play in Two Acts." *Development* 136 (18): 3033–42. https://doi.org/10.1242/dev.033183.

Tan, Feng, Yue Lu, Wei Jiang, Tian Wu, Ruoyu Zhang, Yu Zhao, and Dao-Xiu Zhou. 2018. "DDM1 Represses Noncoding RNA Expression and RNA-Directed DNA Methylation in Heterochromatin." *Plant Physiology* 177 (3): 1187–97. https://doi.org/10.1104/pp.18.00352.

Tan, Feng, Chao Zhou, Qiangwei Zhou, Shaoli Zhou, Wenjing Yang, Yu Zhao, and Guoliang Li. 2016. "Analysis of Chromatin Regulators Reveals Specific Features of Rice DNA Methylation Pathways." *Plant Physiology* 171 (July): 2041–54. https://doi.org/10.1104/pp.16.00393.

Tang, Walfred W. C., Toshihiro Kobayashi, Naoko Irie, Sabine Dietmann, and M. Azim Surani. 2016. "Specification and Epigenetic Programming of the Human Germ Line." *Nature Reviews Genetics* 17 (10): 585–600. https://doi.org/10.1038/nrg.2016.88.

Thorvaldsdottir, H., J. T. Robinson, and J. P. Mesirov. 2013. "Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration." *Briefings in Bioinformatics* 14 (2): 178–92. https://doi.org/10.1093/bib/bbs017.

Wang, Guifeng, and Claudia Köhler. 2017. "Epigenetic Processes in Flowering Plant Reproduction." *Journal of Experimental Botany*, January, erw486. https://doi.org/10.1093/jxb/erw486.

Wang, Zhenxing, Nicolas Butel, Juan Santos-González, Filipe Borges, Jun Yi, Robert A. Martienssen, German Martinez, and Claudia Köhler. 2020. "Polymerase IV Plays a Crucial Role in Pollen Development in Capsella." *The Plant Cell* 32 (4): 950–66. https://doi.org/10.1105/tpc.19.00938.

Ye, Ruiqiang, Wei Wang, Taichiro Iki, Chang Liu, Yang Wu, Masayuki Ishikawa, Xueping Zhou, and Yijun Qi. 2012. "Cytoplasmic Assembly and Selective Nuclear Import of Arabidopsis ARGONAUTE4/SiRNA Complexes." *Molecular Cell* 46 (6): 859–70. https://doi.org/10.1016/j.molcel.2012.04.013.

Zahraeifard, Sara, Maryam Foroozani, Aliasghar Sepehri, Dong-Ha Oh, Guannan Wang, Venkata Mangu, Bin Chen, Niranjan Baisakh, Maheshi Dassanayake, and Aaron P Smith. 2018. "Rice H2A.Z Negatively Regulates Genes Responsive to Nutrient Starvation but Promotes Expression of Key Housekeeping Genes." *Journal of Experimental Botany* 69 (20): 4907–19. https://doi.org/10.1093/jxb/ery244.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.275958; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Zemach, Assaf, M. Yvonne Kim, Ping-Hung Hsieh, Devin Coleman-Derr, Leor Eshed-Williams, Ka Thao, Stacey L. Harmer, and Daniel Zilberman. 2013. "The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin." *Cell* 153 (1): 193–205. https://doi.org/10.1016/j.cell.2013.02.033.

Zhang, Kang, Wenying Xu, Chunchao Wang, Xin Yi, Wenli Zhang, and Zhen Su. 2017. "Differential Deposition of H2A.Z in Combination with Histone Modifications within Related Genes in *Oryza Sativa* Callus and Seedling." *The Plant Journal* 89 (2): 264–77. https://doi.org/10.1111/tpj.13381.

Zhang, W., Y. Wu, J. C. Schnable, Z. Zeng, M. Freeling, G. E. Crawford, and J. Jiang. 2012. "High-Resolution Mapping of Open Chromatin in the Rice Genome." *Genome Research* 22 (1): 151–62. https://doi.org/10.1101/gr.131342.111.

Zhao, Peng, Xuemei Zhou, Kun Shen, Zhenzhen Liu, Tianhe Cheng, Danni Liu, Yanbing Cheng, Xiongbo Peng, and Meng-xiang Sun. 2019. "Two-Step Maternal-to-Zygotic Transition with Two-Phase Parental Genome Contributions." *Developmental Cell* 49 (6): 882-893.e5. https://doi.org/10.1016/j.devcel.2019.04.016.

Zhou, Ming, Lianfeng Gu, Pingchuan Li, Xianwei Song, Liya Wei, Zhiyu Chen, and Xiaofeng Cao. 2010. "Degradome Sequencing Reveals Endogenous Small RNA Targets in Rice (Oryza Sativa L. Ssp. Indica)." *Frontiers in Biology* 5 (1): 67–90. https://doi.org/10.1007/s11515-010-0007-8.