Title
Genome-wide Redistribution of siRNAs in Rice Gametes and Zygotes

Author
Li, Chenxin

Publication Date
2021

Peer reviewed|Thesis/dissertation

Genome-wide Redistribution of siRNAs in Rice Gametes and Zygotes

By

CHENXIN LI
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Plant Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Venkatesan Sundaresan, Chair

Luca Comai

Stacey Harmer

Committee in Charge

2021

Table of Contents

Abstract

Genome-wide Redistribution of siRNAs in Rice Gametes and Zygotes

Gametes and zygote constitute a critical stage in the life cycle for all sexual organisms, including plants. For angiosperms, the seed is the basic reproductive unit, which is produced by double fertilization. The male gametophyte, pollen, contains two sperm cells. One of the sperm cells fuses with the egg cell to produce the zygote, which will develop into the embryo and the next sporophytic generation. The other sperm cell fuses with the diploid central cell to produce the endosperm, a nutritive tissue that supports the growth of the embryo or germinating seedling. Seeds ─ either the endosperm (e.g., rice, maize, wheat, and other cereals) or the embryo (e.g., soybean, peanut, and other pulses) ─ directly or indirectly account for most of the calorie human consume, and seeds are the product of double fertilization. Thus, understanding the biology of gametes and zygote has broad applications in biotechnology and agriculture. In this dissertation I detail an aspect of the biology of rice gametes and zygote: the small RNA transcriptome and its implications regarding epigenome and plant development.

In the first chapter I introduce 1) zygotic genome activation in plant, a process during which the fertilized egg cell transition from the gametic cell fate to the embryonic and totipotent cell fate; and 2) epigenomic reprogramming that was predicted or reported before and after fertilization in angiosperms.

In the second chapter, I present detail protocols for isolating rice gametes, which were used to generate data presented in later chapters of this dissertation.

In the third chapter, I present results of a study characterizing the small RNA transcriptomes and DNA methylomes of rice gametes. The results indicate that a genome-wide

redistribution of 24-nt siRNAs has occurred in rice gametes, which was unexpected and not predictable from our canonical understanding of siRNA functions. However, in both gametes, the patterns of CHH methylation, a strong indicator of RNA-directed DNA methylation, remain similar to each other and to vegetative tissues. These findings are suggestive to unexplored roles of gamete small RNAs.

In the fourth chapter, I present results of a study characterizing the small RNA transcriptome of rice zygotes. The results indicate that wide-spread redistribution of siRNAs occurred in zygote, and newly detected siRNA loci in zygote have a distribution similar to that of canonical siRNA loci detected in embryo and seedling. In addition, zygote siRNA loci, but not egg siRNA loci, were associated with hypermethylation in mature embryo. These findings suggest that resetting of the gametic epigenome towards the canonical vegetative pattern is initiated in the zygote, setting the stage for RdDM activity during embryogenesis before the first embryonic division.

Lastly, I briefly summarize key findings of my dissertation and discuss possible future experiments in the final chapter of this dissertation.

Chapter 1

Introduction

1.1 Zygotic genome activation in plants

Gametes and zygote constitute a critical stage in the life cycle for all sexual organisms. Plant gametogenesis occurs in the haploid gametophyte generation [for reviews, see Twell (2011); Yang, Shi, and Chen (2010)]. In angiosperms, the mature female gametophyte typically contains seven cells and four cell types: one egg cell, two synergid cells, three antipodal cells, and one diploid central cell. The male gametophyte typically contains three cells and two cell types: two sperm cells engulfed in a larger pollen vegetative cell. In angiosperms, upon fertilization, one sperm cell fuses with the egg cell to form a diploid zygote, which give rise to the next sporophytic generation through embryogenesis; the other sperm cell fuses with the central cell to form the endosperm, a nutritive tissue for the developing embryo (Lord and Russell 2002). Understanding the biology of plant gametes and zygotes has agricultural and biotechnological applications, such as haploid induction, synthetic apomixis in crops and plant regeneration from tissue culture.

Maternal to zygotic transition occurs after fertilization. In animals, the maternal to zygotic transition has been well studied. In metazoans, the early embryo is largely transcriptionally silent, and early development depends on maternally pre-deposited RNA and proteins. In addition, the animal zygotic transition is characterized by miRNA guided maternal mRNA degradation and zygotic genome activation (ZGA) (Tadros and Lipshitz 2009). Plant zygotic transition appears to be different from animal (Armenta-Medina and Gillmor 2019). In rice zygotes, publications from our lab showed that almost five hundred genes are upregulated by

the completion of karyogamy. Before the first zygote division, about two thousand genes are up- or down-regulated in the zygote. Furthermore, almost two hundred transcripts undetected in the egg cell are detected in the zygotes (Anderson et al. 2013; 2017). Similar observations were made in maize and *Arabidopsis*, where the zygote has started to produce new transcripts that were undetected in the egg cell (Chen et al. 2017; Zhao et al. 2019). In addition, zygotic transcription in the zygote was required for early embryogenesis (Zhao et al. 2019; Kao and Nodine 2019). These observations suggest that in angiosperms 1) the zygote transcriptome is highly dynamic and 2) the zygotic genome activation (ZGA) occurs in the unicellular zygote.

Although the timing of zygotic genome activation was found to be early, the parental contribution to the zygote transcriptome has been debated. Some studies found maternal dominance in early *Arabidopsis* embryo (Autran et al. 2011; Del Toro-De Leo´n 2014), whereas others found equal-parental contribution (Nodine and Bartel 2012). These conflicting results have been attributed to RNA contamination from maternal tissue (Schon and Nodine 2017), or effects of hybridization between different accessions of *Arabidopsis* (Baroux C, Autran D, Raissig MT, Grimanelli D 2013). However, these studies used four- to eight-cell embryos and not unicellular zygote (**Fig. 1**).

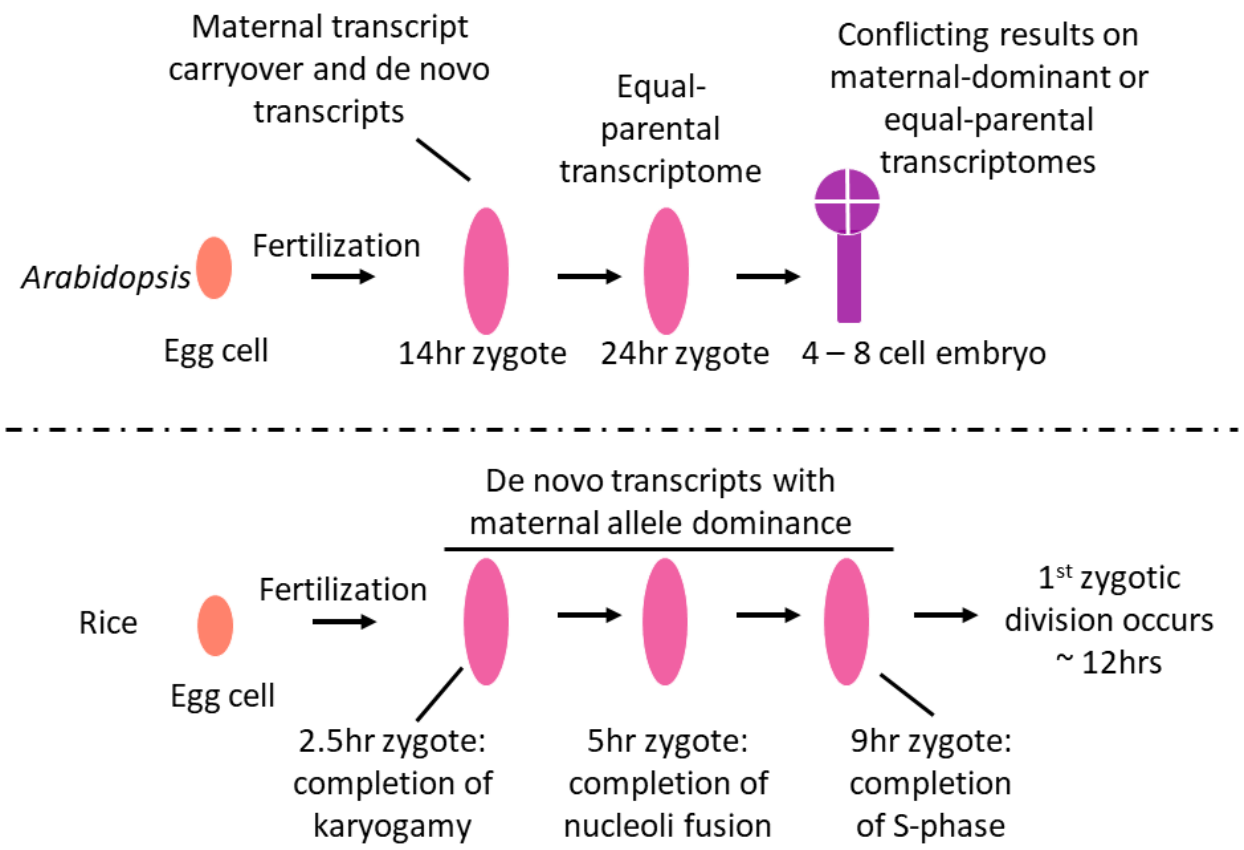


Fig 1: Schematics of zygotic genome activation in *Arabidopsis* and rice. Data: *Arabidopsis* zygote (Zhao et al. 2019); *Arabidopsis* 4 – 8 cell embryos (Autran et al. 2011; Del Toro-De Leo´n 2014; Nodine and Bartel 2012); rice zygote (Anderson et al. 2017).

To answer this question, our lab has selected rice as the model for the following reasons. Rice has a very rapid pollination to fertilization time. Gamete fusion occurs in 30 min after pollination (HAP), and the first zygotic division happens by 12 HAP (Ding et al. 2009). Because rice flowers are well separated and have single ovules, it is possible to hand pollinate and precisely stage *each* zygote. By analysis of SNPs in RNA-seq data from hybrid zygotes, we found a global maternal allele bias for the zygote transcriptome, but also genes with biparental

expression, as well as few genes with a paternal allele bias (Anderson et al. 2017). The maternal bias was likely due to transcript carryover from the egg cell. However, most de novo expressed genes were also maternally biased, suggesting transcript carryover could not be the sole explanation. A recent study in *Arabidopsis* showed that the early zygote exhibited maternal bias, likely due to transcript carryover, while in later stage zygotes, equal-parental contribution was found (Zhao et al. 2019). Such differences may in part be explained by the different timings of the first zygotic division, as the first zygotic division happens ~12 HAP in rice (Ding et al. 2009), but ~24 HAP in *Arabidopsis*. A longer period between pollination and the first zygotic division may provide more time to clear maternal transcripts in *Arabidopsis*, as well as more time for chromatin reprogramming for paternal genome activation, and thus an equal-parental transcriptome in later stage *Arabidopsis* zygotes.

Interestingly, some of the paternally biased genes in rice zygote play important roles in establishing the zygotic identity and promote embryogenesis. For example, the genes *OsBBM1* and *OsWOX9a* exhibited paternal bias. *OsBBM1* homologs have been shown to induce somatic embryos in *Arabidopsis* and *Brassica* (Boutilier et al. 2002). In addition, we found that ectopic expression of *OsBBM1* in the egg cell led to haploid, asexual embryos (Khanday et al. 2018). In *Arabidopsis*, *OsWOX9a* homologs *WOX8/9* are expressed in the zygote and play important roles in embryo polarity (Haecker et al. 2004; Ueda et al. 2017). Further, overexpression of *BBM1* and *WOX9a* related genes promotes somatic embryogenesis, transformation efficiency, and regeneration (Lowe et al. 2016). It is clear that *OsBBM1* (and perhaps *OsWOX9a*) is a plant pluripotency factor, analogous to pluripotency factors in animals (Takahashi and Yamanaka 2006). Taken together, the paternal genome has a minor but significant contribution to the zygote transcriptome.

1.2 Epigenetic reprograming during plant reproduction

Along with dynamic changes of gene expression, reprogramming of the epigenome during gametogenesis and zygotic genome activation has been speculated. In rice and maize, the egg cell is ~10 times larger than the sperm in diameter, and thus ~1000 times larger than sperm cell in volume (Kranz, Bautor, and Lörz 1991; Anderson et al. 2013). For example, rice egg cell is ~50 μm in diameter, whereas rice sperm cells are ~5 μm in diameter [observations made in (Anderson et al. 2013; Li et al. 2019)]. It has been observed that the chromatin of maize egg cell is diffused (Scholten, Lörz, and Kranz 2002). Many other sex-specific changes in chromatin were reported in *Arabidopsis* (Wang and Köhler 2017). In contrast, the sperm cell chromatin undergoes global condensation, which might be related to the deposition of yet unknown nuclear proteins [paralleling protamine deposition in animals (Kimmins and Sassone-corsi 2005)], or yet uncharacterized post-translational modifications of chromatin [paralleling chromatin condensation during cell division]. Meanwhile, a male-germline specific histone H3 variant MGH3/HTR10 (H3.10) is deposited in the sperm cell and its immediate precursor generative cell (Okada et al. 2005; Borg and Berger 2015). It has been observed that H3.1, the major form of H3 protein, is removed during sperm cell development, which is followed by the deposition of H3.10 (Borg et al. 2020). H3.10 is resistant to methylation at lysine 27, which primes the activation of key genes for sperm differentiation and embryogenesis (Borg et al. 2020). For example, proposed pluripotency factors, such as BBM and WOX8/9, were found to be marked by H3 methylated at lysine 4, an activating mark (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, companion cells of gametes experience global chromatin changes as well. Heterochromatin is decondensed in the central cell (Pillot et al. 2010). A similar phenomenon occurs in the pollen vegetative nucleus (Schoft et al. 2009; Mérai et al. 2014; P. Hsieh et al. 2016). Relaxation of heterochromatin pollen vegetative cell was shown to produce short interfering RNA (siRNA) that traffic into the gametes to reinforce silencing in the gametes (Slotkin et al. 2009; Calarco et al. 2011; Martínez et al. 2016; Martinez and Kohler 2017; Park et al. 2016; 2016). Similarly, it has been speculated that siRNAs traffic from the central cell to the egg cell (T.-F. Hsieh et al. 2009; Ibarra et al. 2012), although evidence for which was weaker. In addition, it was also proposed that siRNAs traffic from the endosperm into the embryo during seed development (T.-F. Hsieh et al. 2009; Martinez and Kohler 2017).

In addition to chromatin reprogramming, there is also evidence for changes in DNA methylation during plant reproduction [for review, see Gehring (2019)]. In angiosperms, de novo DNA methylation can be established via RNA-directed DNA methylation [RdDM, for review, see Cuerda-gil, and Slotkin (2016)]. Briefly, RdDM starts with transcription by RNA polymerase IV (Pol IV). Pol IV is associated with an RNA-dependent RNA polymerase RDR2, which produces a dsRNA from the Pol IV transcript. This dsRNA is processed by DCL3 into 24-nt siRNA, which is then loaded onto an argonaute protein AGO4. This siRNA-AGO complex base pairs with the nascent transcript of another RNA polymerase, Pol V, and use it as a scaffold to recruit a DNA methyltransferase, DRM2. DMR2 leads to DNA cytosine methylation at all sequence context, but methylation at the CHH context (mCHH), where H is not G, is a strong indicator of RdDM in rice and maize (Tan et al. 2016, 2018; Gent et al. 2013), but not in

Arabidopsis (Zemach et al. 2013), as CG and CHG methylation are primarily maintained by MET1 and CMT3, respectively (reviewed by Law and Jacobsen 2010). In Arabidopsis, mCHH methylation is primary catalyzed by CMT2 instead (Zemach et al. 2013). Methylated DNA is recognized by chromatin remodelers, which in turn deposit repressive histone marks. Repressive histone marks such as H3 dimethylated at lysine 9 (H3K9me2) recruit Pol IV, and thus RdDM self-reinforces. DNA methylome for *Arabidopsis* egg cell is not yet available, but data are available for *Arabidopsis* sperm cell and pollen vegetative cell. In *Arabidopsis*, the sperm cell has reduced mCHH relative to the pollen vegetative cell (Calarco et al. 2011; P. Hsieh et al. 2016; Walker et al. 2018). In both *Arabidopsis* and rice sperm cells, mCHH and mCHG levels at specific loci coincided with reduced mCG at corresponding loci in the pollen vegetative nucleus, supporting the hypothesis that heterochromatin decondensation in the pollen vegetative cell facilitate production of siRNAs which are trafficked into the sperm cell to direct DNA methylation (Park et al. 2016; Kim et al. 2019).

It is interesting to note that maternal mutants of RdDM components appeared to affect the embryo (Autran et al. 2011; Grover et al. 2018), whereas paternal mutants of RdDM components affect the endosperm 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), which was referred to as the triploid-block. Given the parent-of-origin dependent phenotypes, it is likely that siRNAs exert regulatory functions in gametes. Furthermore, gametic siRNAs may be transmitted into the embryo and endosperm and play important roles during seed development. When this project was initiated, there were no available small RNA transcriptomes from plant egg cells and zygotes, and none from plant sperm cells other than *Arabidopsis*. We set out to sequence small RNA transcriptomes of rice egg cell, sperm cell and zygote. We also included

ovary and seedling shoot as reference tissues. Here are some questions that we had hoped to answer with the small RNA transcriptome data that we gathered:

1) Is there evidence for epigenomic reprogramming in gametes and zygotes?
2) What are the small RNA landscapes in rice gametes? How do they differ from each other and how do they differ from general vegetative tissues, such as seedling?
3) Does miRNA and siRNA regulate gene expression in gametes?
4) Does siRNA direct DNA methylation in gametes?
5) Does rice zygote produce de novo miRNA and siRNA during zygotic genome activation?
6) How does the small RNA transcriptome of zygote differ from those of the gametes?
7) Does siRNA and DNA methylation regulate asynchronous transcriptional activities of parental genomes?

Small RNA transcriptomes for rice gametes and zygotes, together with their mRNA transcriptomes and methylomes, could be used to explore the possibility that small RNAs including miRNAs may be involved in regulating gene expression and chromatin conformation in these cells. Since small RNA are involved in transcriptional silencing through RdDM and chromatin remodeling, we could use them to probe heterochromatin in gametes and zygotes. Since siRNA production is influenced by histone modifications, and that siRNA can direct or maintain DNA methylation, as well as histone modifications, we argue that siRNA is the output and indicator of the epigenome. As low-input chromatin profiling experiments (such as ChIP-seq) is currently unfeasible for egg cells and zygotes due to technical limitations, small RNA offers a unique opportunity to reveal chromatin landscapes in plant gametes and zygote. Lastly, we could use small RNA transcriptome data to understand the asynchronous transcriptional activity of

parental genomes in the zygote. In summary, small RNA transcriptome data for rice gametes and zygote may bridge knowledge gaps in zygotic genome activation, role of RdDM in gametes and zygotes, and epigenomic reprogramming during gametogenesis and zygote genome activation.

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 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." *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.

Baroux C, Autran D, Raissig MT, Grimanelli D, Grossniklaus U. 2013. "Parental Contributions to the Transcriptome of Early Plant Ce," 72–74. https://doi.org/10.1016/j.gde.2013.01.006.

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.

Boutilier, Kim, Remko Offringa, Vijay K Sharma, Henk Kieft, Thérèse Ouellet, Lemin Zhang, Jiro Hattori, et al. 2002. "Ectopic Expression of BABY BOOM Triggers a Conversion from Vegetative to Embryonic Growth." *Plant Cell* 14 (August): 1737–49. https://doi.org/10.1105/tpc.001941.tissue.

Calarco, Joseph P, Filipe Borges, Mark T A Donoghue, Frédéric Van Ex, Pauline E Jullien, Telma Lopes, Rui Gardner, et al. 2011. "Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA." *Cell*, no. Mmc. 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.” *Plant Cell* 29 (September): 2106–25. https://doi.org/10.1105/tpc.17.00099.

Ding, J, J Shen, W Li, and H Wang. 2009. “Cytological Observation of Double Fertilization and Its Duration in Oryza Sativa.” *Chinese Bulletin of Botany* 44 (3880041): 473–83. https://doi.org/10.3969/j.issn.1674-3466.2009.04.009.

Dna, Non-canonical Rna-directed, Diego Cuerda-gil, and R Keith Slotkin. 2016. “Non-Canonical RNA-Directed DNA Methylation,” no. November. https://doi.org/10.1038/NPLANTS.2016.163.

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?,” 0–3. https://doi.org/10.1111/nph.15856.

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, 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*, 575–82. https://doi.org/10.1111/tpj.13910.

Haecker, Achim, Rita Groß-hardt, Bernd Geiges, Ananda Sarkar, Holger Breuninger, Marita Herrmann, and Thomas Laux. 2004. “Expression Dynamics of WOX Genes Mark Cell Fate Decisions during Early Embryonic Patterning in Arabidopsis Thaliana.” https://doi.org/10.1242/dev.00963.

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.

Khanday, Imtiyaz, Debra Skinner, Bing Yang, Raphael Mercier, and Venkatesan Sundaresan. 2018. “A Male-Expressed Rice Embryogenic Trigger Redirect for Asexual Propagation through Seeds.” *Nature* 19. https://doi.org/10.1038/s41586-018-0785-8.

Kim, M Yvonne, Akemi Ono, Stefan Scholten, Tetsu Kinoshita, Daniel Zilberman, and Takashi Okamoto. 2019. “DNA Demethylation by ROS1a in Rice Vegetative Cells Promotes Methylation in Sperm,” 1–6. 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.

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.

Law, Julie A, and Steven E Jacobsen. 2010. “Establishing , Maintaining and Modifying DNA Methylation Patterns in Plants and Animals.” *Nature Reviews Genetics* 11 (3): 204–20. https://doi.org/10.1038/nrg2719.

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.

Lord, Elizabeth M, and Scott D Russell. 2002. “THE MECHANISMS OF POLLINATION AND FERTILIZATION IN PLANTS.” *Annu. Rev. Cell Dev. Biol.*, no. 18: 81–105. https://doi.org/10.1146/annurev.cellbio.18.012502.083438.

Lowe, Keith, Emily Wu, Ning Wang, George Hoerster, Craig Hastings, Myeong-je Cho, Chris Scelonge, et al. 2016. “Morphogenic Regulators Baby Boom and Wuschel Improve Monocot Transformation.” *Plant Cell* 28 (September): 1998–2015. https://doi.org/10.1105/tpc.16.00124.

Martinez, German, and Claudia Kohler. 2017. “Role of Small RNAs in Epigenetic Reprogramming during Plant Sexual Reproduction.” *Current Opinion in Plant Biology*, no. Figure 1. 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.

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.

Nodine, Michael D, and David P Bartel. 2012. “Maternal and Paternal Genomes Contribute Equally to the Transcriptome of Early Plant Embryos.” *Nature* 482 (7383): 94–97. https://doi.org/10.1038/nature10756.

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.” *PNAS*. 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.

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.

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.

Schon, Michael A., and Michael D. Nodine. 2017. “Widespread Contamination of Arabidopsis Embryo and Endosperm Transcriptome Data Sets.” *The Plant Cell* 29 (4): 608–17. https://doi.org/10.1105/tpc.16.00845.

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.

Tadros, Wael, and Howard D Lipshitz. 2009. “The Maternal-to-Zygotic Transition : A Play in Two Acts” 3042: 3033–42. https://doi.org/10.1242/dev.033183.

Takahashi, Kazutoshi, and Shinya Yamanaka. 2006. “Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors.” *Cell* 2: 663–76. https://doi.org/10.1016/j.cell.2006.07.024.

Tan, Fan, Yue Lu, Wei Jiang, Ruoyu Zhang, Yu Zhao, and Dao-xiu Zhou. 2018. “DDM1 Represses Noncoding RNA Expression and RNA-Directed DNA Methylation in Heterochromatin” 177 (July): 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 Speci Fi c Features of Rice DNA Methylation Pathways.” *Plant Physiology* 171 (July): 2041–54. https://doi.org/10.1104/pp.16.00393.

Toro-De Leo´n, G. et al. Del. 2014. “Non-Equivalent Contributions of Maternal and Paternal Genomes to Early Plant Embryogenesis,” 2–6. https://doi.org/10.1038/nature13620.

Twell, David. 2011. “Male Gametogenesis and Germline Specification in Flowering Plants,” 149–60. https://doi.org/10.1007/s00497-010-0157-5.

Ueda, Minako, Ernst Aichinger, Wen Gong, Edwin Groot, Inge Verstraeten, Lam Dai Vu, Ive De Smet, Tetsuya Higashiyama, Masaaki Umeda, and Thomas Laux. 2017. “Transcriptional Integration of Paternal and Maternal Factors in the Arabidopsis Zygote.” *Genes and Development*, 617–27. https://doi.org/10.1101/gad.292409.116.Sakakibara.

Walker, James, Hongbo Gao, Jingyi Zhang, Billy Aldridge, Martin Vickers, and James D Higgins. 2018. “Sexual-Lineage-Specific DNA Methylation Regulates Meiosis in Arabidopsis.” *Nature Genetics* 50 (January). https://doi.org/10.1038/s41588-017-0008-5.

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.

Yang, Wei-cai, Dong-qiao Shi, and Yan-hong Chen. 2010. “Female Gametophyte Development in Flowering Plants.” https://doi.org/10.1146/annurev-arplant-042809-112203.

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.

Zhao, Peng, Xuemei Zhou, Kun Shen, Zhenzhen Liu, Tianhe Cheng, Danni Liu, Yanbing Cheng, and Xiongbo Peng. 2019. “Article Two-Step Maternal-to-Zygotic Transition with Two-Phase Parental Genome Contributions.” *Developmental Cell*, 1–12. https://doi.org/10.1016/j.devcel.2019.04.016.

Chapter 2

Step-by-step protocols for rice gamete isolation

Chenxin Li[1,4], Hengping Xu[2,4], Scott D. Russell[2,5*], Venkatesan Sundaresan[1,3,5,6*]

[1]Department of Plant Biology, University of California, Davis, CA 95616, USA

[2]Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, USA

[3]Department of Plant Sciences, University of California, Davis, CA 95616, USA

[4]These authors contributed equally

[5]Senior author

[6]Lead contact

*Correspondence: srussell@ou.edu (S.D.R.), sundar@ucdavis.edu (V.S.)

2.1 Abstract

Characterization of the transcriptome and other -omics studies of flowering plant gametes are challenging as a consequence of the small sizes and relative inaccessibility of these cells. Collecting such poorly represented cells is also complicated by potential contamination from surrounding sporophytic, adjacent gametophytic tissues and difficulties in extracting high quality intact cells. Here we present detailed, step-by-step procedures for collecting intact, unfixed rice (*Oryza sativa*) egg cells and sperm cells without enzymatic treatments. In addition, we also present a general workflow for assessing sample purity by RT-PCR, using primers specific for marker genes preferentially expressed in surrounding cells and tissues. These protocols should facilitate future studies of genome-scale characterization of gametes in this important model crop.

Author contributions

CL conducted experiments regarding egg cell isolation (Fig 2 and Fig S1B) and molecular diagnosis of samples (Fig 4 and Fig S2). CL also summarized the procedures (Fig 1). Sperm cell isolation (depicted in Fig 3) was conducted by HX. Micrograph depicting an egg cell and a synergid cell side-by-side (Fig S1A) was provide by SDR. All figures except Fig 3 were assembled by CL. CL and HX drafted the manuscript.

This chapter is published in *Plant Reproduction* (2019): https://doi.org/10.1007/s00497-019-00363-y

Plant Reproduction
https://doi.org/10.1007/s00497-019-00363-y

METHODS PAPER

Step-by-step protocols for rice gamete isolation

Chenxin Li[1] · Hengping Xu[2] · Scott D. Russell[2] · Venkatesan Sundaresan[1,3]

Received: 25 October 2018 / Accepted: 28 January 2019

Abstract

Key message **A detailed, step-by-step protocol for isolation of rice gametes for transcriptional profiling, with a general workflow that includes controls for RNA contamination from surrounding cells and tissues is presented.**

Abstract Characterization of the transcriptome and other -omics studies of flowering plant gametes are challenging as a consequence of the small sizes and relative inaccessibility of these cells. Collecting such poorly represented cells is also complicated by potential contamination from surrounding sporophytic, adjacent gametophytic tissues and difficulties in extracting high-quality intact cells. Here we present detailed, step-by-step procedures for collecting intact, unfixed rice (*Oryza sativa*) egg cells and sperm cells without enzymatic treatments. In addition, we also present a general workflow for assessing sample purity by RT-PCR, using primers specific for marker genes preferentially expressed in surrounding cells and tissues. These protocols should facilitate future studies of genome-scale characterization of gametes in this important model crop.

Keywords *Oryza sativa* · Egg cell · Sperm cell · Gametes · Transcriptomes

Introduction

Molecular studies of flowering plant gametes have been impeded by the relatively small size and inaccessibility of these cells. The female gamete or egg cell is embedded within layers of maternal tissue, whereas the two male gametes or sperm cells are engulfed by the pollen vegetative cell. Despite recent advancements in sequencing technology, the quality of transcriptome data of plant gametes, embryos and endosperm has been debated due to potential contamination from surrounding sporophytic tissues (Schon and Nodine 2017). Most of these studies have used Arabidopsis for its many advantages as a plant model, but it presents additional challenges due to the small size of the floral organs and reproductive cells. Rice (*Oryza sativa*) is a particularly appropriate model as it is one of the most important food crops in the world, in addition to being the subject of extensive genomic characterization (Kawahara et al. 2013), and has served as a model cereal for genetic and molecular studies of gene regulation and epigenetics.

Isolation of rice egg cells for global expression studies has been previously published (Abiko et al. 2013; Anderson et al. 2013, 2017; Ohnishi et al. 2011; Zhang et al. 1999). Gene expression has also been profiled in rice sperm cells (Anderson et al. 2013; Okamoto 2017; Russell et al.

A contribution to the special issue 'Cellular Omics Methods in Plant Reproduction Research'.

Communicated by Thomas Dresselhaus.

Chenxin Li and Hengping Xu have contributed equally to this work.

Scott D. Russell and Venkatesan Sundaresan: Senior authors.

Venkatesan Sundaresan: Lead contact.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00497-019-00363-y) contains supplementary material, which is available to authorized users.

✉ Scott D. Russell
srussell@ou.edu

✉ Venkatesan Sundaresan
sundar@ucdavis.edu

[1] Department of Plant Biology, University of California, Davis, CA 95616, USA

[2] Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, USA

[3] Department of Plant Sciences, University of California, Davis, CA 95616, USA

Published online: 12 February 2019

2012, 2017), and a detailed but involved sperm cell isolation method has been published recently (Russell et al. 2017). Here we present a detailed, step-by-step protocol for the isolation of rice gametes, that can be easily adapted by researchers new to the field, in which we also present a general workflow to sufficiently reduce RNA contamination from surrounding cells to be considered negligible. Egg cells are isolated by manual dissection, each captured individually, without fixation or enzymatic treatments. Additionally, we present an improved sperm cell isolation method, also in the same protocol format as for the egg cell. The previous sperm cell isolation protocol (Gou et al. 2011) begins with isolated late pre-anthesis anthers, which are limited by the number of anthers separated. However, based on mass isolations of maize and tobacco sperm cells (Xu and Tsao 1997; Xu et al. 2002), we found that harvesting spikelets was less time-consuming than hand selecting anthers. The current protocol eliminates the tedious manual collection of large amounts of anthers, significantly shortens pollen collection time, and thus facilitates collecting greater volumes and improves repeatability. Sperm cells are then purified by a series of Percoll density gradient centrifuge steps. During isolation of each of the respective gametes, we incorporated extra rinse steps, which in theory should reduce RNA contamination from surrounding cells. The degree of contamination can be subsequently assessed by RT-PCR amplification of marker genes preferentially expressed in nearby cells.

Materials

List I: materials for egg cell isolation

1. Flowering rice plants;
2. 0.3 M mannitol solution: add 27.3 g of mannitol into ~ 200 mL distilled deionized water (ddH_2O). Fill up final volume to 500 mL. Autoclave;
3. Scissors (to remove florets from panicle);
4. 50-mL test tubes;
5. Weighing boats or petri dishes;
6. Dissecting microscope and light source for initial dissection and cleaning;
7. Inverted phase contrast microscope for final collection;
8. Fine tweezers, needles (Covidien 8881250305 Monoject Hypo Needle, 25G × 1″);
9. Thin razor blades (Merkur double edge razor blades);
10. Acupuncture needles or insect pins;
11. Transfer pipettes (Thermo Scientific Samco Standard Disposable Transfer Pipettes, 13-711-9D);
12. Microcapillaries to collect egg cells (microcapillaries are purchased from World Precision Instruments, Inc., Sarasota, Florida USA TW120F-4, pulled to a fine tip using a PUL-1 micropipette (WPI, Inc.) Micropipettes are screened using a dissecting microscope and cleaved to an appropriate hole diameter to allow egg cells and zygotes to be collected with a minimum of damage and contamination).

List II: materials for sperm isolation using blender method

Items 1 and 8 from List I.

1. Scissors (to cut rice stem at the lower part);
2. Bucket (2~3 gallons);
3. 300 mL 45% sucrose: dissolve 135 g sucrose into ~ 250 mL ddH_2O, fill the final volume with ddH_2O up to 300 mL and store at 4 °C or on ice;
4. 100 mL 15% sucrose: dissolve 15 g sucrose into ~ 80 mL ddH_2O, fill the final volume with ddH_2O up to 100 mL and then filter (0.2 μm) and store at 4 °C or on ice;
5. 25 mL 40% Percoll solution in 15% sucrose: dissolve 3.75 g sucrose in 10 mL ddH_2O, add 10 mL Percoll, fill the final volume up to 2 5 mL with ddH_2O and store at 4 °C or on ice;
6. 25 mL 15% Percoll solution in 15% sucrose: dissolve 3.75 g sucrose in 10 mL ddH_2O, add 3.75 mL Percoll, fill the final volume up to 25 mL with ddH_2O and store at 4 °C or on ice;
7. Centrifuge tubes, sterile plastic, 50 mL (Corning 430290) and 15 mL (Corning 430052);
8. Three-milliliters syringe with needle (B–D, 309579) bent in right angle at ~8 mm from tip (for preparing the discontinuous Percoll solution and collecting sperm-rich portion);
9. Four or more isolation tubes: in 15 mL centrifuges, layer 2.5 mL 15% Percoll in 15% sucrose above 2.5 mL 40% Percoll in 15% sucrose;
10. 200-mL flasks;
11. 100-mL beakers;
12. Funnels (2 fl. oz);
13. Electric blender (Hamilton Beach);
14. Three pieces of nylon mesh, 80–100 cm^2 each, with mesh sizes of 100, 30 and 10 μm, respectively;
15. Centrifuge with swinging bucket capable of temperature control and speeding up to 4000 × g;
16. Light microscope.

List III: materials for RNA extraction and quality control

1. Nuclease free water;
2. 100% ethanol;
3. Ambion RNAqueous-Micro Total RNA Isolation Kit (AM1931);

4. Qiagen RNase free DNase (79254);
5. Bench top centrifuge capable of spinning 12,000×*g* or higher;
6. PCR thermal cycler;
7. Agilent Bioanalyzer 2100 and RNA Pico kit (5067-1513);
8. NuGEN Ovation RNA-seq System V2 (7102);
9. Qiagen MinElute Reaction Cleanup Kit (28204);
10. ND 1000 Nanodrop Spectrophotometer (Thermo Fisher Scientific);
11. PCR and agarose gel electrophoresis supplies.

Rice plant growth condition

Rice (*Kitaake* variety) seeds are surface-sterilized (70% bleach for 10 min followed by washes with autoclaved water) and germinated in ddH2O in petri dishes wrapped with parafilm for 2 weeks in growth chamber (28 °C, 500 µmol m^{-2} s^{-1} light for 14.5 h and 25 °C, dark for 9.5 h, with constant 80% relative humidity). Two weeks after sowing, seedlings are then transplanted to greenhouse. The temperature of greenhouse is 28 °C for daytime (7:00 am to 9:30 pm) and 25 °C for nighttime, with 14.5 h light per day. Plants are irrigated with deionized water twice a week and supplied with nutrient water every other week (Fig. 1).

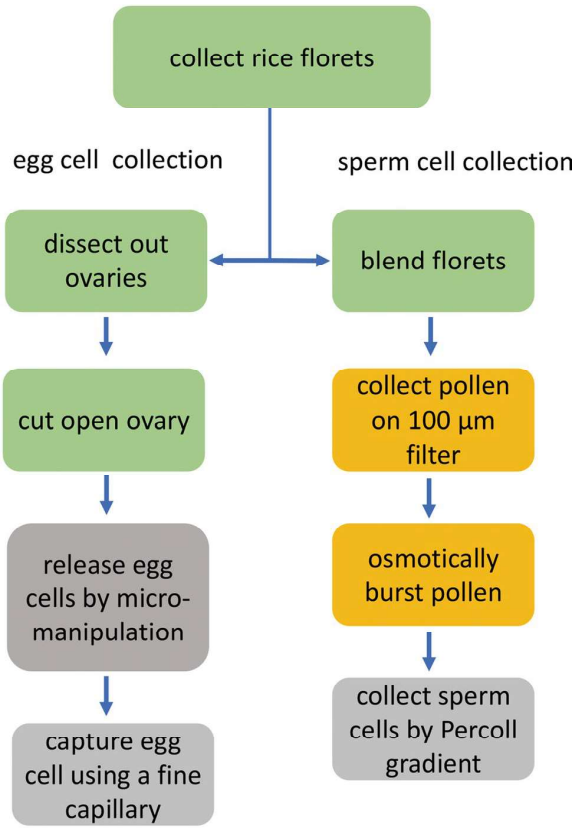


Fig. 1 Flowchart showing an overview of the gamete isolation procedures. The colors of text boxes are indicative of the color of the sample at specific steps

Methods

Method I: isolation of rice egg cells

1. Collect rice florets into 0.3 M mannitol in a 50-mL test tube in the morning before anthesis. Rice florets (*Kitaake* variety) usually enter anthesis around 10:00 am under our growth conditions.[1]
2. To open the floret, a floret is held by two pairs of tweezers with the opening of the floret facing down. Use one pair of tweezers to hold the palea, while the other pair holds the lemma (Fig. 2a). Gently separate these until the lemma falls off. Gently remove the stamens at the base using a pair of tweezers (Fig. 2b). Tearing apart the palea along its symmetric axis should give access to the base of stamen. After all six stamens are removed, use the tip of the tweezers to very gently remove the ovary from the base of the floret (Fig. 2c). Care should be taken not to damage the ovary. After the ovaries are removed, they are allowed to float on 0.3 M mannitol in a weighing boat or petri dish until further dissection. Note: take care to control evaporation or molarity of mannitol may excessively increase. Dissect ovaries out from all the florets before proceeding to the next steps.[2]
3. Rinse a microscope slide with deionized water while rubbing it with an ungloved thumb. Pipette 6 µL of 0.3 M mannitol onto the slide, and the liquid should assume a small convex droplet and not spread out.
4. Under a dissecting stereomicroscope, examine the stigma of the ovary to ensure no pollen has landed. Mount the ovary into the mannitol droplet. Remove any remaining stamens using a needle. Use a thin razor blade to make a transverse cut through the middle region of the ovary (Fig. 2d). A clean cut should be made taking care not to squeeze or crush the ovary. Sharp blades have a better chance in producing clean cuts and therefore

[1] Care should be taken to identify mature florets that are staged correctly. In mature florets, stamens should occupy most of the floret prior to anthesis.

[2] For RT-PCR, we used MyTaq Red Mix (Bioline BIO-25043). All primers listed in Supplementary Table 1 performed using an annealing temperature of 59 °C.

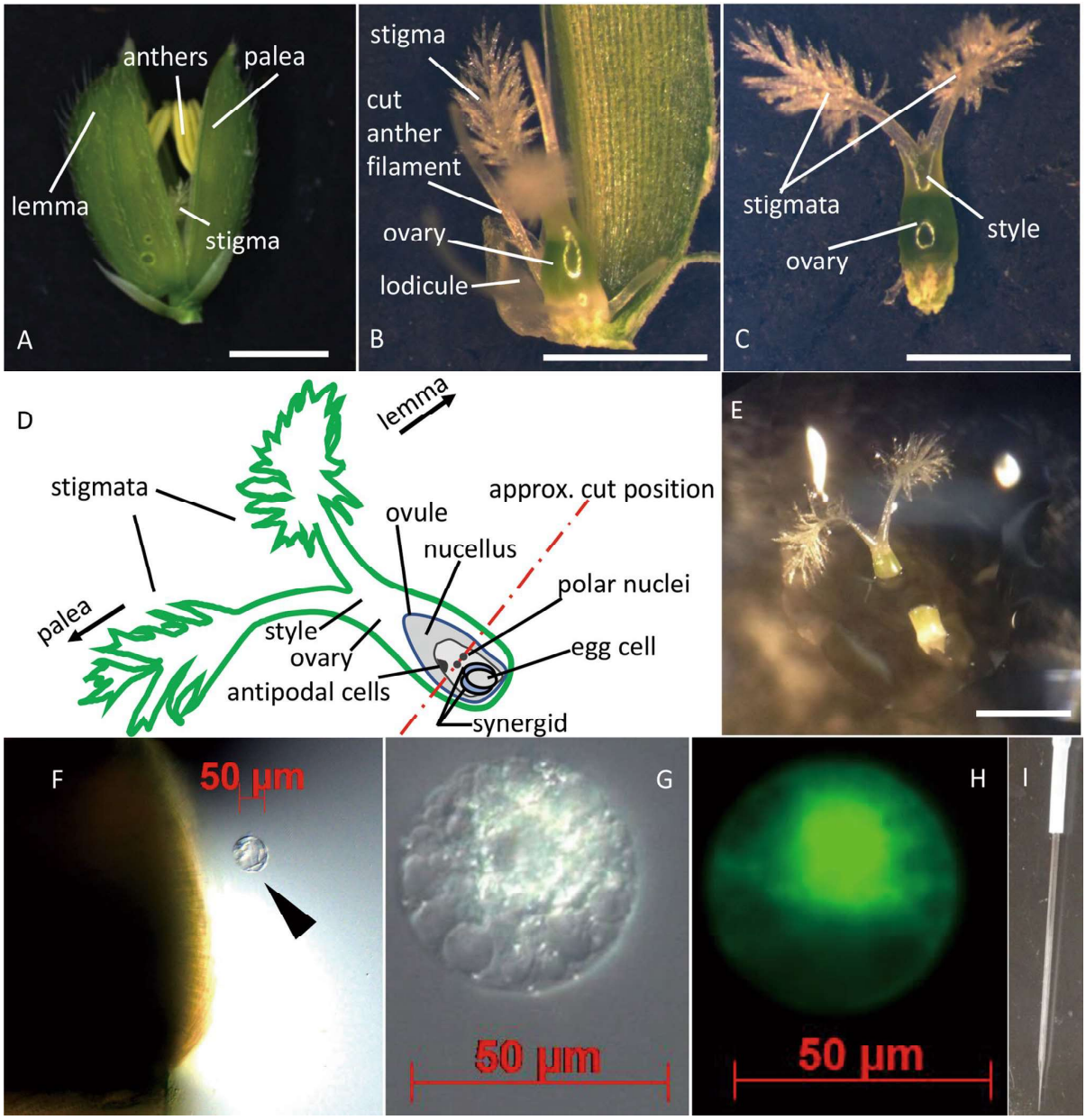


Fig. 2 Key steps of rice egg cell isolation **a**: a rice floret after the bracts are pulled open; **b**: floret after lemma and anthers are removed; **c**: dissected rice ovary; **d**: scheme of inner structure of rice ovary used to indicate the incision line and the positions of embryo sac cells; **e**: cut ovary; **f**: egg cell (indicated by the black arrowhead) has been released from ovary; **g**: differential interference contrast image of isolated egg cell; **h**: FDA staining of isolated egg cell; **i**: fine capillary used to collect egg cells connected to the tip of a transfer pipette. Bar = 5 mm for **a–d**; 1 mm for **e**, **f**, **h**; 50 µm for **i**

each side of the razor blade should be used only once in a given position to avoid dulling of the blade (Fig. 2e).

5. To reduce the risk of RNA contamination from floral organs, once the ovary is cut, pipette a new 6 µL droplet of 0.3 M mannitol onto the slide and move the basal part of ovary so as to rinse the ovary in the new droplet.
6. Using an inverted phase contrast microscope (10× objective lens), use an acupuncture needle or insect pin to

very gently exert pressure on the ovary with the goal of relaxing tension in the cell wall. In around 3 min, the egg cell may be released from the ovary (Fig. 2f). The rice egg cell is about 50 µm wide, having many large vacuoles. There is only one egg cell per floret. Pay attention in picking the egg cell based on the cell morphology. Egg cells may not always be captured because: (1) the ovary has been damaged and the egg cell has been ruptured or (2) staging of the flower is not correct and the egg cell has not yet developed.

7. Once the egg cell floats out of the incision in the ovary, it is captured by a fine capillary connected to a transfer pipette (Fig. 2i), under the 40× object lens (Fig. 2g). To reduce the risk of RNA contamination, pipette the egg-containing liquid into a new 0.3 M mannitol droplet to rinse the egg cell.
8. The isolated egg cell can be stained with fluorescent diacetate (FDA) to examine its viability (Fig. 2h). Alternatively, for transcriptome or other omics studies, the egg cell is then immediately transferred into a 1.5-mL centrifuge tube and frozen in liquid nitrogen. The captured cells should always be kept frozen in liquid nitrogen or dry ice and maintained at −80 °C freezer temperatures until RNA extraction. Around 35 egg cells are adequate as a biological replicate that will provide sufficient RNA for reverse transcription.

Synergid cells are sometimes also released at step 6. However, the two cell types can easily be distinguished (Supplementary Fig. 1). The egg cell appears clearer and more transparent, with larger vacuoles. In contrast, the synergid, which is slightly smaller in size, appears more opaque with many small vesicles. This same protocol can also be adapted for rice zygote isolation if the starting material is pollinated, time-staged florets (Anderson et al. 2017).

Depending on the skill of the experimenter, our protocol processes 10 ovaries (post-step 2) in about 40 min per person, with a 30~50% frequency in capturing intact egg cells. In a 2-h period, 10–15 egg cells can be collected by one person. The success rate increases as the experimenter becomes more experienced with the protocol. The non-enzymatic isolation method from Uchiumi et al. (2007) can collect 30–40 egg cells per day. This protocol has the potential to achieve similar yields if sufficient rice plants at the appropriate reproductive stage are available. The previous method from Zhang et al. (2010) could collect 3–5 egg cells from 30 ovaries in an hour. Although fewer ovaries are processed in an hour, this protocol is more efficient than that described in Zhang et al. (2010).

Method II: sperm isolation using blender method

1. Select ~ 50 rice panicles with mature florets around 10 am (for variety *Kitaake*) before anthesis. Cut panicles at their base and place them into the bucket with ~ 200 mL water to keep hydrated.
2. Collect all mature and almost mature florets (1 day before flowering) into the blender (Hamilton Beach, with single speed setting), containing ~150 mL 45% sucrose (Fig. 3a).
3. Blend tissues for 3 times to release pollen grains from anthers: 30 s each time.
4. Separate pollen-containing liquid into a flask from the blended mixture using 100 µm nylon mesh on a funnel (Fig. 3b).
5. To remove floral organ cytoplasm from pollen, pass the 100 µm filtrate (solution containing pollen) through a 30-µm nylon mesh (Fig. 3c). Pollen will be trapped on the mesh.
6. Carefully transfer pollen from 30-µm mesh into a beaker (Fig. 3d) by rinsing the mesh with 45% sucrose using a transfer pipette. Then add an additional ~ 50 mL 45% sucrose into the beaker and swirl for ~ 1 min.
7. Repeat Steps 5 and 6 once or more until the filtered sucrose solution is clear and colorless (Fig. 3e). Intact and pure pollen grains are on the 30-µm mesh (Fig. 3f, g).

   For sperm cell isolation, following procedures are based on the protocol developed by Russell et al. (2017) with modifications.
8. To burst pollen for sperm isolation, quickly rinse 1–2 mL of the pollen into a 50-mL tube in ~20 mL 15% sucrose (Fig. 3h), seal the cap and rotate horizontally and slowly at room temperature for ~ 25 min.
9. Filter the mixture through a 30-µm mesh to separate sperm cells from sperm-depleted pollen (Fig. 3i).
10. Equally split the filtrate into two aliquots. Carefully layer each aliquot using a syringe with a 90° tip-bent needle on top of a 15% Percoll solution in an isolation tube.
11. Centrifuge at 4 °C, 4000×*g* for 45 min with slower acceleration and deceleration. For example, for a Thermo Scientific Heraeus Multifuge X3R Centrifuge, set acceleration to 8 (with 9 being fastest) and deceleration at 9 (with 10 being fastest); the purpose is to prevent the interface of discontinuous Percoll solutions from being disturbed by rapid acceleration or deceleration. The sperm-rich portion at the interface of 40/15% Percoll will be visible after centrifuging (Fig. 3j, Russell et al. 2017 with permission of the publisher).
12. Remove supernatant by pipetting until 0.5–1 cm above the interface and collect up to ~ 0.5 mL sperm-rich

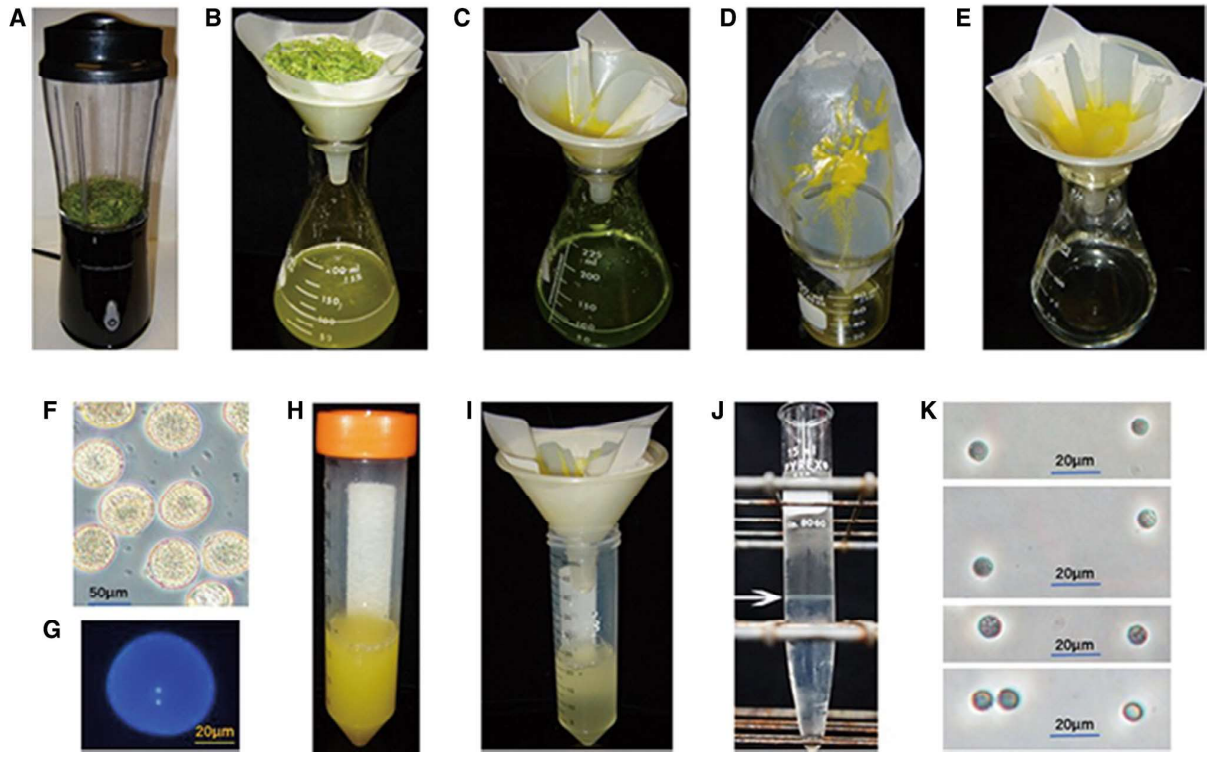


Fig. 3 Key steps for rice sperm cell isolation using blender method. **a**: rice florets collected into the blender containing 45% sucrose solution; **b**: pollen-containing solution in flask separated from the blended mixture through 100-μm mesh; **c**: pass the filtrate of step B through 30-μm mesh; **d**: pollen on the mesh from step C transferred to a clean beaker for washing in 45% sucrose; **e**: pollen washed twice; **f**: intact and almost clean pollen grains viewed under microscope note that minor debris can be observed, likely arising from individual burst pollen grains; **g**: DAPI staining to show the two sperm cells within the pollen grain; **h**: pollen incubated with 15% sucrose to release sperm cells; **i**: sperm cell-containing filtrate in a 50-mL tube; **j**: Sperm cell-rich layer indicated by arrow at the interface of 40/15% Percoll (Russell et al. 2017, with permission of the publisher); **k**: rice sperm cells under microscope from different isolations to show repeatability of this method

layer using the syringe with the tip-bent needle from each aliquot.

13. Dilute sperm-rich fraction with 4 volumes of 15% sucrose in a new tube and filter with 10-μm mesh to prevent the potential aggregation of sperm cells (aggregated sperms, along with impurity, may form a new layer on the surface of 15% Percoll and thus reduce the yield of sperm cells).
14. Layer the filtrate on the top of 15% Percoll in a new isolation tube, and centrifuge at 4 °C, 3000×*g* for 25 min.
15. For higher purity (but lower sperm cell yield), repeat Steps 13 and 14 once or more.
16. Collect 0.2–0.5 mL from the interface into a new tube, add 3 mL of 15% sucrose, and centrifuge at 4 °C, 1000×*g* for 10 min; sperm cells are collected at the tube bottom.
17. Remove most supernatant by pipetting but leave 0.1–0.2 mL from the bottom. Add 1 mL 15% sucrose and centrifuge again (4 °C, 1000×*g*, 10 min).
18. Slowly remove the supernatant. Leave 30–50 μL, the sperm-rich portion, at the bottom.
19. Use 1 μl for microscopy (Fig. 3k); save the rest in a new Eppendorf tube at −80 °C; for RNA isolation, use DEPC treated Eppendorf tube and freeze it in liquid nitrogen, then store at −80 °C until use.

Method III: RNA extraction and quality control

RNA extraction

Efficient total RNA isolation from low input materials, such as egg cell and sperm cell, can be achieved using the Ambion RNaqueous Micro Total RNA kit. We also perform

an on-column DNase treatment using Qiagen DNase. Our protocol, which also recovers RNA species < 200 nt that would be useful for miRNA and siRNA profiling, is included here:

- Make 20 µL DNase solution per sample.
  5 µL DNase I
  15 µL RDD buffer.
  20 µL Total.
- Pre-warm DNase solution at 37 °C.
- Bring wash buffers 1 and 2/3 to room temperature.
- Add 200 µL lysis solution to cells and immediately vortex until sample is completely thawed. Do not thaw samples prior to adding lysis solution.
- Add 250 µL of 100% ethanol, vortex briefly and centrifuge briefly.
- Load lysate/ethanol mixture (up to 150 µL) on to column, spin at maximum speed for 10 s.
- Load remaining lysate, spin at maximum speed for 10 s.
- Add 180 µL Wash Solution I and spin at maximum speed for 10 s.
- Add 20 µL DNase solution to the center of the filter.
- Incubate at 37 °C for 30 min.
- Pre-warm elution buffer at 75 °C, 25 µL per sample.
- Wash with 180 µL Wash Solution I, spin at maximum speed for 10 s.
- Wash with 180 µL Wash Solution 2/3, spin at maximum speed for 10 s.
- Repeat with a second 180 µL aliquot of Wash Solution 2/3, spin at maximum speed for 10 s.
- Empty flow through, spin at maximum speed for 1 min to dry.
- Elute 2× with 7 µL heated elution buffer into a new tube, incubate 1 min and then spin at maximum speed for 30 s.
- Pipette 1.5 µL into 0.2-mL tubes to be quality checked using RNA 6000 Pico Kit on the Bioanalyzer, using the Eukaryotic Total RNA Pico program. Typically, a high-quality RNA sample should return a Bioanalyzer trace with two rRNA peaks in addition to the marker peak, whereas the higher molecular weight peak should be twice as intense as the lower molecular weight peak (Supplementary Fig. 2).
- Store RNA at −80 °C.

cDNA synthesis

Due the extreme difficulties of collecting rice gametes, the number of cells in each biological replicate is low and thus the RNA concentration of each sample is also low. We adapted NuGEN Ovation RNA-seq System V2 for reverse transcription to produce ~1 µg range cDNA from total RNA input as low as 1 ng. After cDNA synthesis, the cDNA products are purified using Qiagen MinElute Reaction Cleanup Kit and quantified on a Nanodrop spectrophotometer. The RNA input can be variable across samples; however, equal amount of cDNA should be used for RT-PCR for each sample. One microgram of cDNA sample can be sheared or fragmented for subsequent RNA-seq library construction.

Alternatively, RNA samples can be quantified using a Qubit fluorometric instrument (Thermo Fisher Scientific) by RNA HS assays (Q32852). Equal amounts of RNA across samples can then be used in reverse transcription reactions using iSript Select cDNA Synthesis Kit (BIO-RAD 1708896) or similar. After cDNA synthesis, 2 µL of cDNA product can be used in a 20 µL PCR.

PCR and gel electrophoresis

To address the issue of possible contamination from sporophytic tissue, we surveyed the literature for marker genes expressed preferentially in floral organs but very low in gametes (Anderson et al. 2013). A list of primer sequences can be found in Supplementary Table 1. For each gene, 25 PCR cycles were performed, and equal volumes of PCR products were run on a 2% agarose gel, at 120 V for 30 min.

To validate the identity of the egg cells collected by the protocol, we used ECA-like 1/2 as egg cell marker genes (Ohnishi et al. 2011). Strong signals were obtained with PCR products for ECA-like 1/2 transcripts, confirming their egg cell identity (Fig. 4a). To assess contamination from ovary, we used as a marker the MADS box gene OsMADS16, which is preferentially expressed in the ovary, and has low expression in the egg cell (Anderson et al. 2013; Xiao et al. 2003). We observed in the PCR products, a strong signal for ovary tissue and a weak signal for the egg cells, indicating that there is minimum contamination from the sporophytic ovary (Fig. 4a). For RNA-seq and other -omics experiments, we suggest performing control collections and control libraries. For example, for each egg cell collected, ~1 µL of cell-free, tissue-free mannitol should be collected into a separate tube as a control sample. Control samples should be included in RNA extraction and subsequent library construction. Control samples with no library products will suggest the cognate cell collections are free of RNA contamination from surrounding tissues.

To verify that this protocol does collect sperm cells, we used MGH3 as the sperm marker gene (Anderson et al. 2013; Okada et al. 2005). Strong signals for PCR products for MGH3 transcript in sperm samples indicated that we were successful in collecting sperm cells (Fig. 4b). Since mature rice pollen grains are trinuclear (Fig. 3g), mature anthers yield sperm cells rather than generative cells.

To assess RNA contamination from the pollen vegetative cell, we used LAT52 as the marker gene (Cook and Thilmony 2012). A sperm cell preparation collected using our previously published method (Russell et al. 2017) was

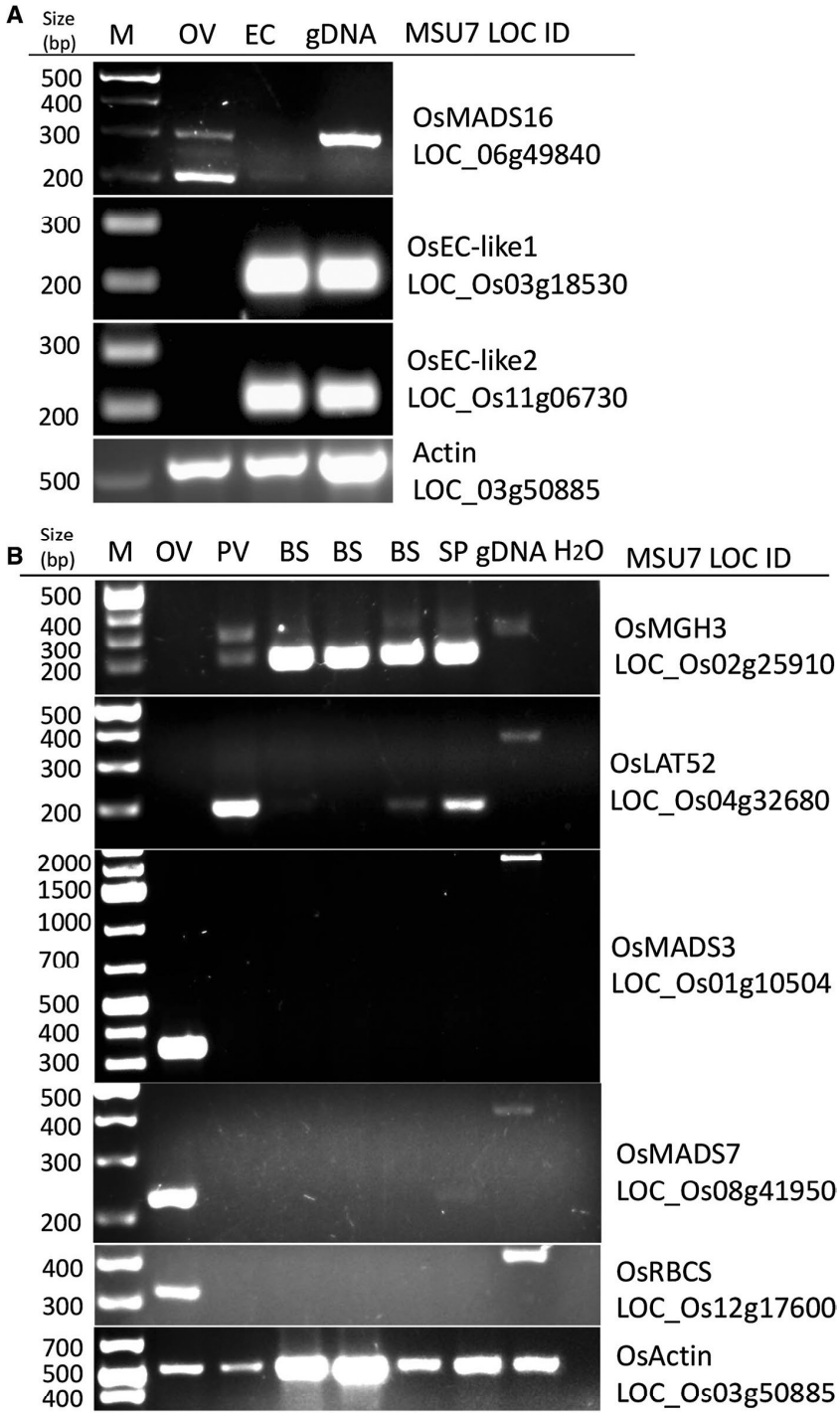


Fig. 4 Confirmatory semiquantitative RT-PCR experiments for **a**: egg cell and ovary marker genes; **b**: sperm cell marker genes. *M* DNA size marker; *OV* ovary, *EC* egg cell, *PV* pollen vegetative cells, *BS* sperm cells collected by blender method (this study), *SP* sperm cells collected using Russell et al. 2017 method; *gDNA* genomic DNA, *MSU7 LOC ID* locus ID number from MSU Rice Genome Annotation Project, Release 7 (http://rice.plantbiology.msu.edu)

also included as a control sample. Sperm samples collected with this protocol appear to perform better than our previously published protocol (Russell et al. 2017) in terms of pollen vegetative cell contamination, since the bands for LAT52 are much fainter in the current protocol (Lanes BS vs. SP in Fig. 4b).

To assess RNA contamination from floral organs, we used a set of marker genes for floral organs. Since whole florets

were collected into the blender, an ovary sample was used as a positive control. MADS3 is preferentially expressed in bracts and stamen (Kyozuka et al. 2002). MADS7 is preferentially expressed in stamen and tapetum (Lu et al. 2006). These MADS genes all have detectable expression in ovary (Fig. 4b). In addition, we also included Rubisco small unit (RBCS) as a marker gene for green tissues (Kawahara et al. 2013). RT-PCR showed that while each gene amplified a strong band for ovary samples, transcripts of these marker genes are undetectable in sperm samples collected by blender method, suggesting that sperm samples collected by this method are free of floral organ and sporophytic tissue contamination.

Author contributions statement All authors were involved in the concepts and design of the experiments. CL and HX conducted experiments and analyzed data. CL and HX wrote the manuscript, with input from SR and VS. All authors read and approved the manuscript.

Acknowledgements We thank Imtiyaz Khanday, Jonathan Gent, Sarah Anderson and Daniel Jones for helpful advice for optimizing the experimental methods. We thank Debra Skinner for assistance in artwork. This research was funded by the National Science Foundation (Award No. IOS-1547760) and the USDA Agricultural Experiment Station (Project No. CA-D-XXX-6973-H).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Abiko M, Furuta K, Yamauchi Y, Fujita C, Taoka M, Isobe T, Okamoto T (2013) Identification of proteins enriched in rice egg or sperm cells by single-cell proteomics. PLoS ONE. https://doi.org/10.1371/journal.pone.0069578

Anderson SN, Johnson CS, Jones DS, Conrad LJ, Gou X, Russell SD, 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(5):729–741. https://doi.org/10.1111/tpj.12336

Anderson SN, Johnson CS, Chesnut J, Jones DS, Khanday I, Woodhouse M, Li C, Conrad LJ, Russell SD, Sundaresan V (2017) The zygotic transition is initiated in unicellular plant zygotes with asymmetric activation of parental genomes. Dev Cell 43(3):349. e4–358.e4. https://doi.org/10.1016/j.devcel.2017.10.005

Cook M, Thilmony R (2012) The OsGEX2 gene promoter confers sperm cell expression in transgenic rice. Plant Mol Biol Rep 30(5):1138–1148. https://doi.org/10.1007/s11105-012-0429-3

Gou X, Wang S, Chen F (2011) Isolation and cytological observation of viable sperm cells of rice. Acta Botanica Sinica 41(6):669–671. https://doi.org/10.1360/zd-2013-43-6-1064

Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, Childs KL, Davidson RM, Lin H, Quesada-Ocampo L, Vaillancourt B, Sakai H, Lee SS, Kim J, Numa H, Itoh T, Buell CR, Matsumoto T (2013) Improvement of the *Oryza sativa* nipponbare reference genome using next generation sequence and optical map data. Rice 6(1):3–10. https://doi.org/10.1186/1939-8433-6-4

Kyozuka J, Shimamoto K, Ag OR (2002) Ectopic expression of OsMADS3, a rice ortholog of AGAMOUS, caused a homeotic transformation of lodicules to stamens in transgenic rice plants. Plant Cell Physiol 43(1):130–135

Lu XC, Gong HQ, Huang ML, Bai SL, He YB, Mao X, Geng Z, Li SG, Wei L, Yuwen JS, Xu ZH, Bai SN (2006) Molecular analysis of early rice stamen development using organ-specific gene expression profiling. Plant Mol Biol 61(6):845–861. https://doi.org/10.1007/s11103-006-0054-3

Ohnishi T, Takanashi H, Mogi M, Takahashi H, Kikuchi S, Yano K, Okamoto T, Fujita M, Kurata N, Tsutsumi N (2011) Distinct gene expression profiles in egg and synergid cells of rice as revealed by cell type-specific microarrays. Plant Physiol 155(2):881–891. https://doi.org/10.1104/pp.110.167502

Okada T, Endo M, Singh MB, Bhalla PL (2005) Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3. Plant J 44(4):557–568. https://doi.org/10.1111/j.1365-313X.2005.02554.x

Okamoto T (2017) Analysis of proteins enriched in rice gamete. In: Schmidt A (ed) Plant germline development: methods and protocols. Springer, New York, pp 251–263. https://doi.org/10.1007/978-1-4939-7286-9_20

Russell SD, Gou XP, Wong CE, Wang X, Yuan T, Wei XP, Bhalla PL, Singh MB (2012) Genomic profiling of rice sperm cell transcripts reveals conserved and distinct elements in the flowering plant male germ lineage. New Phytol 195(3):560–573. https://doi.org/10.1111/j.1469-8137.2012.04199.x

Russell SD, Jones DS, Anderson S, Wang X, Sundaresan V, Gou X (2017) Isolation of rice sperm cells for transcriptional profiling. In: Schmidt A (ed) Plant germline development: methods and protocols. Springer, New York, pp 211–219. https://doi.org/10.1007/978-1-4939-7286-9_17

Schon MA, Nodine MD (2017) Widespread contamination of Arabidopsis embryo and endosperm transcriptome data sets. Plant Cell 29(4):608–617. https://doi.org/10.1105/tpc.16.00845

Uchiumi T, Uemura I, Okamoto T (2007) Establishment of an in vitro fertilization system in rice (*Oryza sativa* L.). Planta 226(3):581–589. https://doi.org/10.1007/s00425-007-0506-2

Xiao H, Wang Y, Liu D, Wang W, Li X, Zhao X, Xu J, Zhai W, Zhu L (2003) Functional analysis of the rice AP3 homologue OsMADS16 by RNA interference. Plant Mol Biol 52(5):957–966. https://doi.org/10.1023/A:1025401611354

Xu HP, Tsao TH (1997) Detection and immunolocalization of glycoproteins of the plasma membrane of maize sperm cells. Protoplasma 198(3–4):125–129. https://doi.org/10.1007/BF01287560

Xu H, Weterings K, Vriezen W, Feron R, Xue Y, Derksen J, Mariani C (2002) Isolation and characterization of male-germ-cell transcripts in *Nicotiana tabacum*. Sex Plant Reprod 14(6):339–346. https://doi.org/10.1007/s00497-002-0128-6

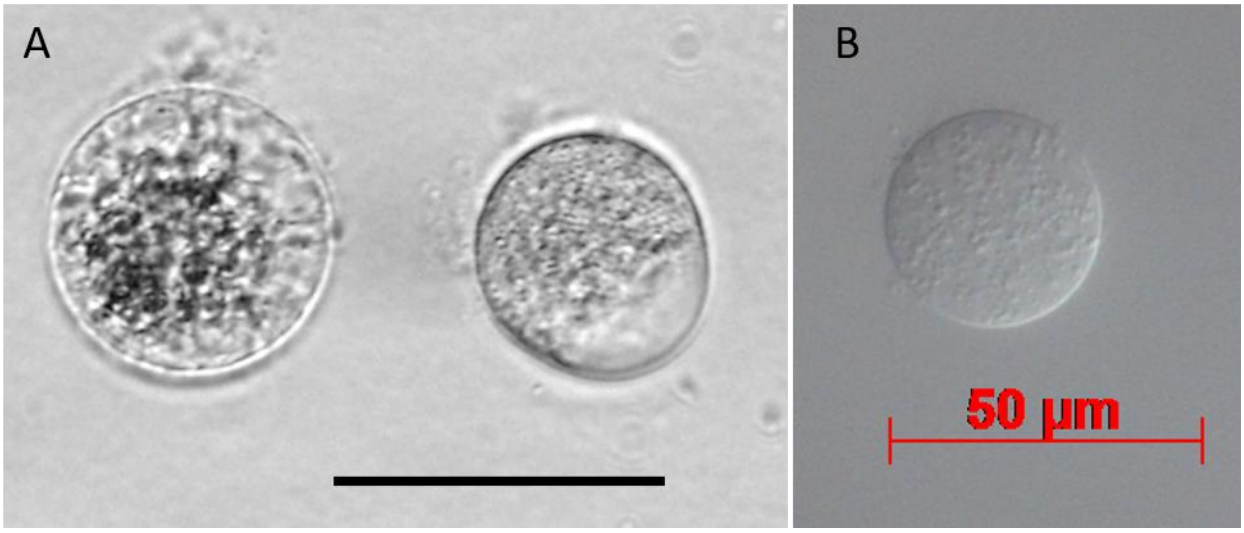
Zhang J, Dong WH, Galli A, Potrykus I (1999) Regeneration of fertile plants from isolated zygotes of rice (*Oryza sativa*). Plant Cell Rep 19(2):128–132. https://doi.org/10.1007/s002990050722

Zhang YN, Wei DM, He EM, Miao S, Tian HQ, Russell SD (2010) Isolation of male and female gametes of rice. Crop Sci 50(6):2457-2463. https://doi.org/10.2135/cropsci2010.02.0066

Supplementary Materials for Step-by-Step Protocols for Rice Gamete Isolation

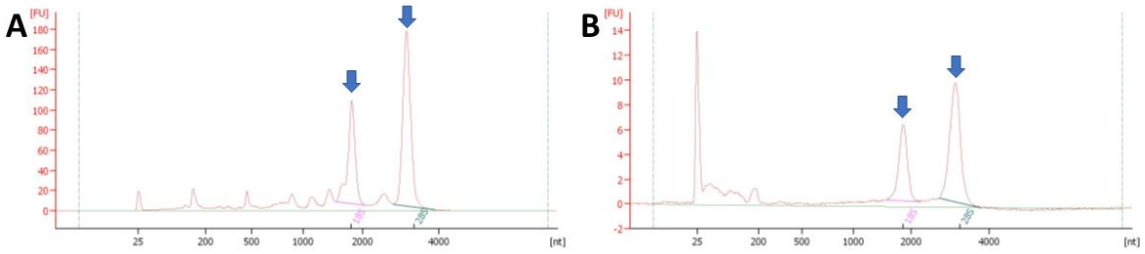
Chenxin Li[1,4], Hengping Xu[2,4], Scott D. Russell[2,5*], Venkatesan Sundaresan[1,3,5,6*]
[1]Department of Plant Biology, University of California, Davis, CA 95616, USA
[2]Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, USA
[3]Department of Plant Sciences, University of California, Davis, CA 95616, USA
[4]These authors contributes equally
[5]Senior author
[6]Lead contact
*Correspondence: srussell@ou.edu (S.D.R.), sundar@ucdavis.edu (V.S.)

Supplementary Fig. 1



Supplementary Fig. 1 Comparison of egg and synergid morphology after isolation. **A**: Left: egg cell, right synergid; **B**: An isolated synergid. Bar = 50 µm.

Supplementary Fig. 2



Supplementary Fig. 2 Representative Bioanalyzer traces, with arrows pointing to the rRNA peaks. **A**: egg cells; **B**: sperm cells.

Supplementary Table 1: Primer sequences and amplicon lengths.

oligo	sequence 5'-3'	cDNA product size/bp	gDNA product size/bp
Actin F	GAAGGATATGCTCTCCCCCATG	524	524
Actin R	GGATCCTCCAATCCAGACACTG		
lat52 F	CAAGGCCGTGTCTACTGTGA	192	378
lat52 R	GATGTCCTCCTCATGGCTGT		
MADS3F1	GACACCTCCAACTCTGGC	341	1965
MADS3R1	GCTGCCCCCATCATGTTC		
MADS7F	CTGGAGGAAAGCAACCATGT	220	418
MADS7R	ATGGGGGCATGTAGGTGTT		
MADS16F	GAGATCAAGCGGATCGAGAA	198	299
MADS16R	CTGGTAGCGGTCAAAGATCC		
MGH3F	ACGGAGCTGCTGATAAGGAA	197	296
MGH3R	CGTCTTTGGACATGATGGTG		
RBCSF	GCAGCAGTAGCTGAGCTTGA	318	421
RBCSR	GTGTCCCACCCGTAGTCG		
EC-like1F	AGCAGTGCTGGGAGGTTCT	200	200
EC-like1R	GCAGTAGCCCTTGAGCATGT		
EC-like2F	CGCCGTCCTCCTACTACTTG	209	209
EC-like2R	CCGTTGACGAAGAAGAGCAC		

Chapter 3

Genome-wide redistribution of 24-nt siRNAs in rice gametes

Chenxin Li*[1], Hengping Xu*[2], Fang-Fang Fu[3], Scott D. Russell[†2], Venkatesan Sundaresan[†1,4], and Jonathan I. Gent[3†]

[1]Department of Plant Biology, University of California, Davis, California 95616, USA

[2]Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, USA

[3]Department of Plant Biology, University of Georgia, Athens, GA 30602, USA

[4]Department of Plant Sciences, University of California, Davis, California 95616, USA

[†]To whom correspondence should be addressed: gent@uga.edu (J.I.G), sundar@ucdavis.edu (V.S.), and srussell@ou.edu (S.D.R.)

*****Equal contributors

3.1 Abstract

Gametes constitute a critical stage of the plant life cycle during which the genome undergoes reprogramming in preparation for embryogenesis. Here we examined genome-wide distributions of small RNAs in the sperm and egg cells of rice. We found that 24-nt siRNAs, which are a hallmark of RNA-directed DNA methylation (RdDM) in plants, were depleted from heterochromatin boundaries in both gametes relative to vegetative tissues, reminiscent of siRNA patterns in DDM1-type nucleosome remodeler mutants. In sperm cells, 24-nt siRNAs were spread across heterochromatic regions, while in egg cells, 24-nt siRNAs were concentrated at a smaller number of heterochromatic loci throughout the genome, especially at loci which also produced siRNAs in other tissues. In both gametes, patterns of CHH methylation, typically a

strong indicator of RdDM, were similar to vegetative tissues, although lower in magnitude. These findings indicate that the small RNA transcriptome undergoes large-scale redistribution in both male and female gametes, which is not correlated with recruitment of DNA methyltransferases in gametes and suggestive of unexplored regulatory activities of gamete small RNAs.

Author contributions

CL isolated egg cells for small RNA experiments. HX isolated sperm cells for small RNA and DNA methylation experiments, who also collected two additional samples of egg cell for DNA methylation experiments. CL collected seedling and ovary for small RNA experiments. CL collected ovary and floral bracts for DNA methylation experiments. CL produced *drm2* knockout lines using CRISPR-Cas9 gene editing technology. FFF dissected embryo and endosperm from wildtype and *drm2* rice. CL produced all small RNA libraries. FFF produced all bisulfite-seq libraries. JIG and CL drafted the outline of figures and determined which analyses to include in the figures. JIG performed all command line analyses (from raw reads to data tables). CL performed all downstream analyses in R (from data tables to figures and statistics), producing figures 1 through 4 and all supplemental figures. All figures, expect Fig 5, were assembled by CL. The graphical summary (Fig 5) was produced by JIG. CL and JIG drafted the manuscript.

This chapter is published in *Genome Research* (2020). Supplemental tables and datasets can be found with the online version of this article: https://genome.cshlp.org/content/30/2/173

Genome-wide redistribution of 24-nt siRNAs in rice gametes

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

[1]*Department of Plant Biology, University of California, Davis, California 95616, USA;* [2]*Department of Microbiology and Plant Biology, University of Oklahoma, Norman, Oklahoma 73019, USA;* [3]*Department of Plant Biology, University of Georgia, Athens, Georgia 30602, USA;* [4]*Department of Plant Sciences, University of California, Davis, California 95616, USA*

Gametes constitute a critical stage of the plant life cycle during which the genome undergoes reprogramming in preparation for embryogenesis. Here, we examined genome-wide distributions of small RNAs in the sperm and egg cells of rice. We found that 24-nt siRNAs, which are a hallmark of RNA-directed DNA methylation (RdDM) in plants, were depleted from heterochromatin boundaries in both gametes relative to vegetative tissues, reminiscent of siRNA patterns in DDM1-type nucleosome remodeler mutants. In sperm cells, 24-nt siRNAs were spread across heterochromatic regions, while in egg cells, 24-nt siRNAs were concentrated at a smaller number of heterochromatic loci throughout the genome, especially at loci which also produced siRNAs in other tissues. In both gametes, patterns of CHH methylation, typically a strong indicator of RdDM, were similar to vegetative tissues, although lower in magnitude. These findings indicate that the small RNA transcriptome undergoes large-scale redistribution in both male and female gametes, which is not correlated with recruitment of DNA methyltransferases in gametes and suggestive of unexplored regulatory activities of gamete small RNAs.

[Supplemental material is available for this article.]

During sexual reproduction in angiosperms, cells undergo a transition from vegetative to reproductive fates to produce spore mother cells. These spore mother cells then undergo meiosis to produce haploid spores that then undergo mitosis to produce gametes and other cells of the gametophyte. The development of reproductive cells is marked by multiple sex-specific changes in chromatin structure (for review, see Wang and Köhler 2017). For example, histone H1 and the centromere-specific histone variant CENH3 are depleted from the megaspore mother cell and the egg cell (Ingouff et al. 2010, 2017; She et al. 2013). Heterochromatin is decondensed in the central cell, which gives rise to the endosperm (Pillot et al. 2010; Yelagandula et al. 2014; Ingouff et al. 2017). A similar phenomenon occurs in the pollen vegetative nucleus (Schoft et al. 2009; Ingouff et al. 2010; Mérai et al. 2014; Hsieh et al. 2016). Most striking is the compaction of sperm cell chromatin, likely related to deposition of a set of histone variants including a male expressed H3 variant (for review, see Borg and Berger 2015). Specific activity of some types of euchromatic retrotransposons in the male germline in maize may also reflect male-specific chromatin changes (Dooner et al. 2019).

While DNA methylation patterns in general are transmitted stably across generations, there is also evidence for both loss and gain of methylation in specific sequence contexts and cell types in reproduction (for review, see Gehring 2019). RNA-directed DNA methylation (RdDM) is particularly interesting because of its potential to be erased and re-established de novo by siRNAs (for review, see Cuerda-Gil and Slotkin 2016). Methylation in the CHH context (mCHH), where H is A, C, or T, is a strong indicator of RdDM in rice, though not in *Arabidopsis* (Zemach et al. 2013; Stroud et al. 2014; Niederhuth et al. 2016; Tan et al. 2016). Live-cell analysis of transgene-driven expression of fluorescent proteins fused to methyl binding domains in *Arabidopsis* reveals reduced signal for mCHH in the megaspore mother cell, microspore, and in the sperm cell, and for mCG in the egg cell (Ingouff et al. 2017). Whole-genome bisulfite sequencing (WGBS) has been limited by the difficulty of obtaining sufficient quantities of pure reproductive cell types, but methylomes have been produced from sperm and pollen vegetative nuclei of both rice and *Arabidopsis* (Calarco et al. 2012; Ibarra et al. 2012; Hsieh et al. 2016; Kim et al. 2019) and from egg cells in rice (Park et al. 2016). Multiple differences are apparent between cell types by WGBS, including in mCHH. In *Arabidopsis*, the sperm cell has reduced mCHH content relative to the pollen vegetative cell (Calarco et al. 2012; Ibarra et al. 2012; Hsieh et al. 2016; Walker et al. 2018). The majority of mCHH in *Arabidopsis* pollen vegetative nuclei is primarily due to activity of the chromomethyltransferase CMT2 rather than RdDM (Hsieh et al. 2016; Borges et al. 2018). In both *Arabidopsis* and rice sperm, mCHH and mCHG levels at specific loci are correlated with reduced mCG in corresponding loci in the pollen vegetative nucleus, consistent with the hypothesis that chromatin changes in the pollen vegetative nucleus facilitate expression of siRNAs which are then transferred into the sperm cell to direct DNA methylation (Calarco et al. 2012; Ibarra et al. 2012; Kim et al. 2019).

Sequencing small RNAs from *Arabidopsis* whole pollen and isolated sperm indicates that 24-nt siRNAs typical of RNA polymerase IV (Pol IV) activity are the most abundant length, but 21-nt and 22-nt siRNAs increase in abundance in sperm cells relative to

[5]**These authors contributed equally to this work.**
Corresponding authors: gent@uga.edu, sundar@ucdavis.edu, srussell@ou.edu
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.253674.119.

vegetative tissues (Slotkin et al. 2009; Borges et al. 2018). This increase in the ratio of 21- and 22-nt siRNAs to 24-nt siRNAs in *Arabidopsis* pollen is typical of siRNAs in vegetative tissues in mutants with compromised DNA methylation (Creasey et al. 2014; McCue et al. 2015; Corem et al. 2018; Fu et al. 2018; Long et al. 2018; Tan et al. 2018). The *Arabidopsis* pollen vegetative nucleus, however, does not have the reduced DNA methylation characteristic of *ddm1* mutants, nor does the sperm, except in mCHH (Calarco et al. 2012; Ibarra et al. 2012; Hsieh et al. 2016; Walker et al. 2018).

Seeds produced from crosses with one parent defective in RdDM exhibit a variety of phenotypes, including the timing of zygotic genome activation, genomic imprinting, and genome dosage sensitivity (Autran et al. 2011; Erdmann et al. 2017; Borges et al. 2018; Grover et al. 2018; Martinez et al. 2018; Kirkbride et al. 2019; Satyaki and Gehring 2019). There are currently no available small RNA transcriptomes from sperm cells of flowering plants other than *Arabidopsis* and none from flowering plant egg cells. Due to the small size of the sperm cell relative to the egg cell (~1/1000 the volume), the transcriptome of the newly fertilized egg cell is principally determined by the egg cell (Autran et al. 2011; Del Toro-De León et al. 2014; Anderson et al. 2017; Zhao et al. 2019). Given their likely importance both pre- and post-fertilization, we sequenced small RNAs from isolated sperm cells, egg cells, ovaries, and seedling shoots of rice, compared expression patterns of different classes of small RNAs, and investigated the relationship of 24-nt siRNAs to DNA methylation in the gametes.

Results

We prepared and sequenced small RNA Illumina libraries from seedling shoots, isolated sperm cells, isolated egg cells, and whole ovaries (after manual removal of egg cells) of rice variety Kitaake. For each cell or tissue type, we prepared at least three biological replicates. After removing PCR duplicates as described in Methods, we aligned all reads of 20 to 25 nt in length to the MSU7 reference genome assembly (Supplemental Table S1; Kawahara et al. 2013).

Distinct microRNA expression patterns in gametes

We examined microRNA (miRNA) expression across samples (values for each miRNA listed in Supplemental Data Set 1). Such miRNAs are predicted to be less mobile between cells than siRNAs (Grant-Downton et al. 2013), and since mRNA expression is globally different between egg, sperm, and seedling (Russell et al. 2012; Anderson et al. 2013, 2017), we also expected regulators of mRNA expression, such as miRNAs, to be differentially expressed across sample types. We performed principal component analysis (PCA) (Fig. 1A) to examine global differences of miRNA expression across samples. Each tissue formed a cluster on a miRNA PCA plot, with seedling leaf separated from reproductive tissues along the first PC axis, which explained ~40% of variance; and male separated from female along the second PC axis, which explained ~13% of the variance. To discover which individual miRNAs drove the variation between tissues, we performed hierarchical clustering and detected multiple miRNA clusters that distinguished seedling shoots from both gametes, as well as clusters that were enriched in one gamete but not the other (Supplemental Fig. S1A; Supplemental Table S3). Most clusters correlated with either one of the PCs (Supplemental Fig. S1B). Thus, the separation of samples on PCA plots is likely driven by miRNA clusters that have distinct expression patterns across tissues. miR159 family members were the most abundant miRNAs, especially in egg cell and ovary (Fig. 1B). In *Arabidopsis*, miR159 family members repress MYB transcription factors and are expressed in multiple tissues, including the sperm cell, and paternal miR159 regulates nuclear division in endosperm (Allen et al. 2010; Zhao et al. 2018). No maternal function for miR159 has been reported, but two members of the miR159 family—miR159a.1 and miR159b—account for greater than 60% of the total miRNAs in rice egg cell and ovary. In seedling shoot and sperm cell, miR159 accounted for greater than 20% of the total miRNAs.

The manual dissection of ovaries to isolate egg cells raises concern regarding the release of small RNA from the sporophytic ovary into the isolation medium and RNA being carried along with egg cell small RNAs into sequencing libraries. Three lines of evidence, however, indicated RNA contamination from the ovary in egg cells was minimal. First, our egg or zygote isolation method has previously been shown to have little mRNA contamination from the surrounding sporophytic ovule (Anderson et al. 2017; Li et al. 2019). Second, we performed differential expression analysis of the miRNAs and found that eight miRNAs had high levels in all three replicates of ovary but were nearly undetectable in all six replicates of egg cells (Supplemental Fig. S1C). Third, we did mock egg cell isolations in which we dissected ovaries using the same protocol, but we only collected cell-free solution rather than the egg cells. By qPCR quantification, library preparation from the mock samples yielded only about 5% of the number of library molecules as an actual egg cell small RNA library after 20 cycles of PCR amplification (Supplemental Fig. S2). Similarly, our sperm cell isolation method has previously been shown to have little mRNA contamination after isolation from surrounding floral tissues or pollen (Li et al. 2019).

Distinct small RNA composition across gametes and vegetative tissues

Small RNA compositions from gametes were distinct from general vegetative tissues, such as seedling shoot, most noticeably by the reduction of miRNAs relative to total small RNAs (Fig. 1C; Supplemental Fig. S3). In seedling shoots, 5.5% of the small RNA reads mapped to miRNA loci, as compared to 1.8% in egg cell and 0.7% in sperm cell. In addition to miRNAs, the relative abundance of phased secondary siRNAs (phasiRNAs), tRNAs, NOR RNAs, and 5S rRNAs varied across gametes and vegetative tissues. The relative abundance of 21-nt and 24-nt phasiRNAs clearly differentiated gametes. Twenty-four-nucleotide phasiRNAs are expressed in meiotic anthers in diverse angiosperms, including rice (Johnson et al. 2009; Zhai et al. 2015b; Fei et al. 2016; Xia et al. 2019). In female reproductive tissues, phasiRNAs are not well characterized, but the argonaute MEL1 binds 21-nt phasiRNAs and is required for both male and female meiosis in rice (Nonomura et al. 2007; Komiya et al. 2014). While both classes were rare in egg cell relative to seedling shoot, 21-nt phasiRNAs were even rarer in sperm cells (Fig. 1C). In contrast, 24-nt phasiRNAs were greater than sixfold more abundant in sperm cell as compared to ovary, egg cell, and seedling shoot. Comparison with published data (Li et al. 2017), however, revealed far higher relative abundance of both 21-nt and 24-nt phasiRNAs in anthers than in either egg or sperm cell (Supplemental Fig. S3).

The most abundant small RNAs in angiosperm vegetative tissues are 24-nt siRNAs that are involved in RNA-directed DNA methylation and, to a lesser extent, 21- and 22-nt siRNAs (for review, see Cuerda-Gil and Slotkin 2016). We grouped all 20- to

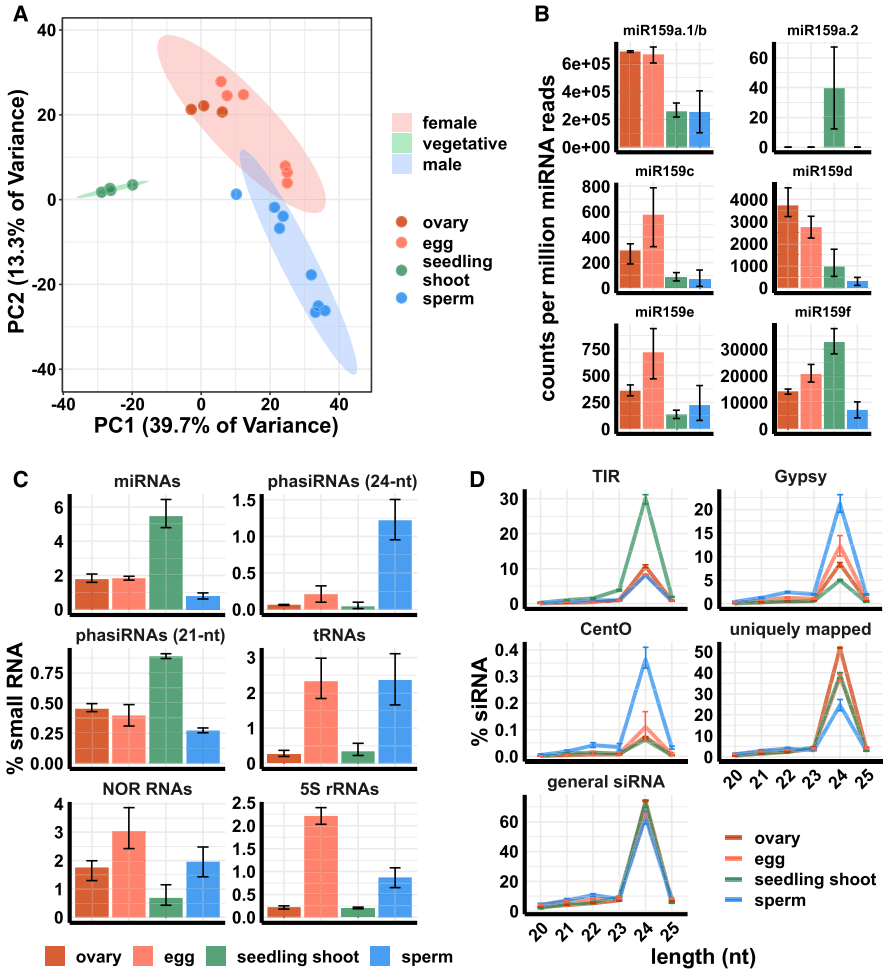


Figure 1. Small RNA composition of ovary, egg cell, sperm cell, and seedling shoot. (*A*) Principle component analysis (PCA) plot of tissues by miRNA expression pattern. (*B*) miRNA159 family expression. *y*-axis values are relative to the total number of miRNA reads in each tissue. Color code is the same as in *A*. (*C*) Small RNA compositions. *y*-axis values are relative to the total number of reads that mapped to the genome. (*D*) siRNA abundance by length and category. "siRNAs" refers to all 20- to 25-nt small RNAs that mapped to the genome but were not included in any category in *C*. "Uniquely mapped" refers to the subset that map with a MAPQ value of at least 20. "*Gypsy*" is the subset that overlaps with an annotated *Gypsy* element, and "TIR" is the subset that overlaps with an annotated DNA transposon of the *Tc1/Mariner*, *PIF/Harbinger*, *Mutator*, or *hAT* superfamilies. "*CentO*" is the subset that overlaps *CentO* tandem repeats. *y*-axis values are the number of siRNAs of each length normalized by the total number of siRNAs. In all panels, error bars are 95% confidence intervals based on biological replicates of each tissue.

25-nt small RNAs that did not overlap at least 90% of their length with miRNA, phasiRNA, tRNA, NOR RNA, or 5S rRNA loci into a general siRNA category for further analysis. The length distribution of these siRNAs was similar between tissues, with a strong preference for 24-nt (Fig. 1D; Supplemental Fig. S4). Parsing out siRNAs that overlapped by at least 90% of their length with retrotransposons of the *Gypsy* superfamily and the centromeric tandem repeat *CentO* revealed strong siRNA enrichments for these elements in sperm cell. *CentO* tandem repeats are enriched in centromeres (Nagaki et al. 2003), and *Gypsy* retrotranposons are generally enriched in heterochromatic regions of the genome (Kawahara et al. 2013). siRNAs overlapping with terminal inverted repeat (TIR) DNA transposons of the *PIF/Harbinger*, *Tc1/Mariner*, *Mutator*, or *hAT* superfamilies were enriched in seedling shoot but were depleted in sperm cell, egg cell, and ovary. For brevity, we refer to siRNAs from these four superfamilies of DNA transposons as simply TIR siRNAs. These transposons are enriched near genes (Han et al. 2013).

Distinct genome-wide patterns of 24-nt siRNAs in gametes

Consistent with the high abundance of *CentO* and *Gypsy* siRNAs, a genome-wide view of 24-nt siRNA abundance revealed that 24-nt siRNAs in sperm cells had a complementary pattern to that of seedling shoot. Sperm cell 24-nt siRNAs were enriched in gene-poor, heterochromatic regions, whereas seedling shoot 24-nt siRNAs had the expected pattern for canonical RdDM in vegetative tissues, with enrichment in euchromatic regions that are gene-rich (Fig. 2A; Supplemental Figs. S5–S7). Egg cell 24-nt siRNAs showed a pattern distinct from both sperm cell and seedling shoot, with

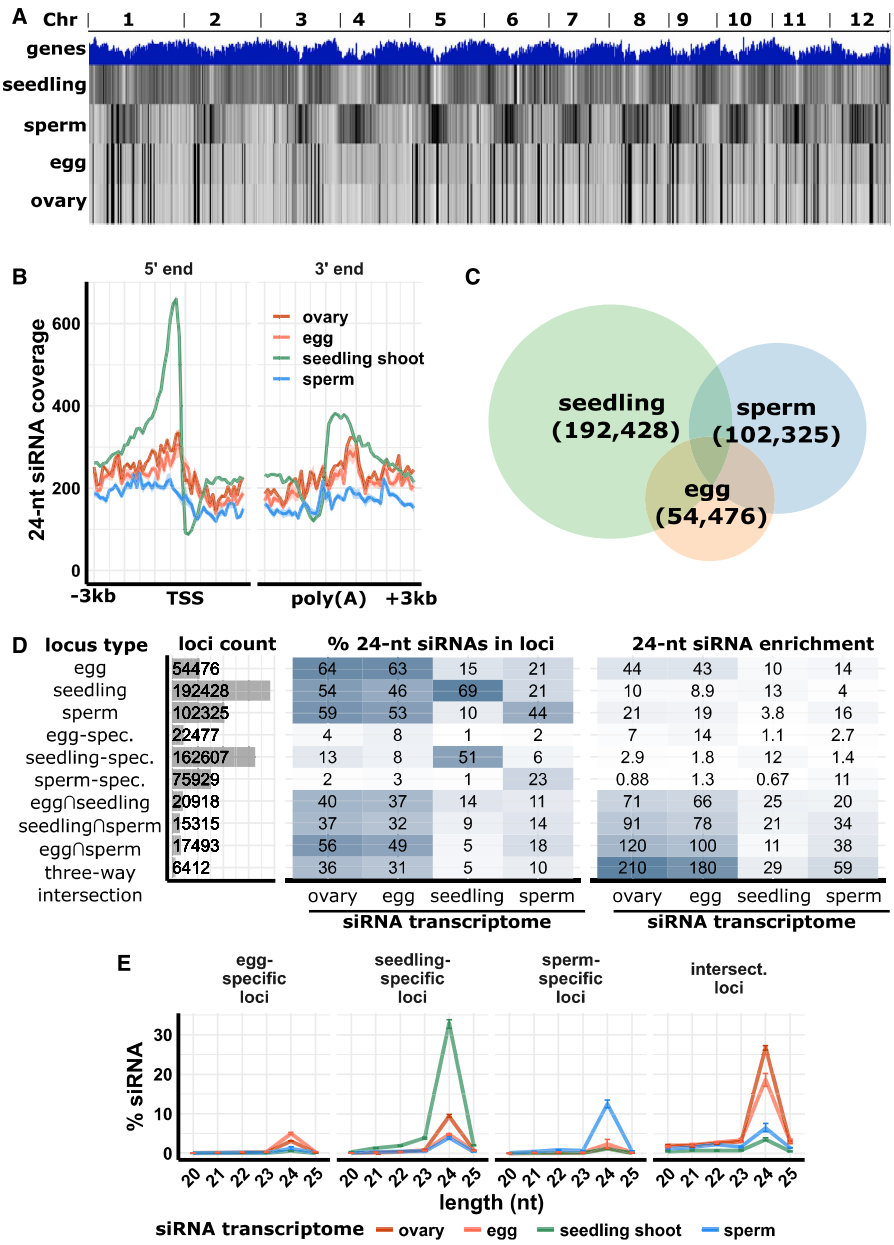


Figure 2. siRNA distributions in ovary, egg cell, sperm cell, and seedling shoot. (*A*) Whole-genome 24-nt siRNA heat maps. *Top* track: gene density; *second* to *fifth* tracks: 24-nt siRNAs from seedling shoot, sperm cell, egg cell, and ovary. (*B*) Metagene coverage plot for 24-nt siRNAs. Coverage is measured over 100-bp intervals and normalized per 1000 total siRNAs. Vertical grid lines indicate 500-bp intervals. (TSS) Transcription start site, (poly[A]) polyadenylation site. (*C*) Venn diagram showing number and overlap of 24-nt siRNA loci. The genome was divided into 100-bp intervals and categorized as 24-nt siRNA loci based on coverage of 24-nt siRNAs from seedling shoot, sperm cell, and egg cell. (*D*) Number of 24-nt siRNA loci and abundance and enrichments of siRNAs from each tissue. Abundance (% 24-nt siRNAs in loci) is the number of 24-nt siRNAs that overlapped with the loci relative to the total number of 24-nt siRNAs from the tissue. Enrichment is the abundance normalized by the number of 24-nt loci. In locus types, "spec." is short for "specific," the set of loci only identified from a single tissue. (*E*) siRNA abundance at 24-nt siRNA loci. "Intersect." refers to intersection loci, those identified as 24-nt siRNA in all three tissues. *y*-axis values are percent 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.

depletion across most of the genome except at a smaller number of loci with no clear relationship to gene density. The distribution of 21- and 22-nt siRNAs also did not have a clear relationship to gene density (Supplemental Fig. S5). Twenty-four-nucleotide siRNAs from ovary showed a nearly identical pattern with egg cell (Fig. 2A). Twenty-one- and 22-nt siRNAs were also similar between

egg cell and ovary but less similar than 24-nt siRNAs (Supplemental Fig. S5). Comparison of 24-nt siRNAs from egg cell with ovaries in earlier stages of development (Li et al. 2017) indicated the egg/ovary pattern was established by the stage of megasporocyte formation (Supplemental Fig. S7). Consistent with the genome-wide analysis, metagene plots of the distribution of 24-nt siRNAs relative to genes revealed a depletion of 24-nt siRNAs both upstream of and downstream from genes in egg cell, sperm cell, and ovary (Fig. 2B), corresponding to where TIR transposons are enriched in the genome, with the exception of the CACTA superfamily (Han et al. 2013).

To identify representative sets of loci associated with siRNAs in each tissue, we divided the genome into nonoverlapping 100-bp intervals and counted the number of mapping siRNAs per two million total siRNAs from each tissue. Any 100-bp interval with at least three mapping 24-nt siRNAs that spanned at least one-third (34 bp) of the 100-bp sequence we categorized as a 24-nt siRNA locus. In seedling shoot, 192,428 loci met these criteria; in sperm, 102,325; and in egg, 54,476 (Fig. 2C). Consistent with the complementary patterns of 24-nt siRNA loci in sperm and seedling shoot in the whole-genome view of siRNA abundance (Fig. 2A), only 15% of the sperm loci were shared with seedling shoot loci (Fig. 2D). In contrast, 38% of the egg loci were shared with seedling shoot. The loci that were shared between all three tissues we called intersection 24-nt siRNA loci, whereas the loci that were unique to one tissue we called either seedling-specific, sperm-specific, or egg-specific. Twenty-four-nucleotide siRNAs from seedling shoot were the most sample-specific of the three in terms of total numbers of siRNAs overlapping seedling-specific loci (69%), whereas 24-nt siRNAs from egg cell were the least sample-specific (8% overlapping egg-specific loci). Instead, the majority of 24-nt siRNAs from egg cells overlapped 24-nt siRNA loci that were shared with one or both other tissues (Fig. 2D,E). Taking into account the number of loci in each category relative to the size of the genome revealed that intersection loci were more strongly enriched for 24-nt siRNAs than sample-specific loci in all three tissues but most strongly in egg cells (Fig. 2D). Twenty-four-nucleotide siRNAs from ovary had a similar pattern as egg cell, whereas other vegetative tissues were similar to seedling shoot (Fig. 2D,E; Supplemental Fig. S8). Counting overlaps between 24-nt siRNA loci and repetitive elements and measuring distances between 24-nt siRNA loci and nearest genes revealed expected trends of seedling-specific 24-nt loci being near genes and overlapping with TIR transposons, typical of canonical RdDM. (Supplemental Fig. S9). Gamete-specific loci were farther from genes and overlapped more with *Gypsy* retrotransposons, whereas intersection loci were similar to seedling-specific loci. Taken together, these data reveal that sperm and egg cells have distributions of siRNAs that are uncharacteristic of canonical RdDM in rice and that are distinct from each other. The sperm cell is complementary to vegetative tissues in its enrichment of siRNAs in large numbers of heterochromatic loci, whereas the egg cell and ovary have different patterns, with large numbers of siRNAs in a small number of loci that tend to be shared in sperm and vegetative tissues.

To test whether the pattern of 24-nt siRNAs in either sperm or egg cell could be explained by loss of 24-nt siRNAs from canonical RdDM loci, we removed all seedling shoot 24-nt siRNAs that overlapped with seedling-specific 24-nt siRNA loci and then examined the distribution of the remaining seedling 24-nt siRNAs in the genome. While this did result in fewer siRNAs near genes, it did not produce the enrichment in heterochromatic regions observed in sperm (Supplemental Fig. S6). Canonical RdDM is indicated by mCHH in addition to 24-nt siRNAs, so we also identified putative canonical RdDM loci independently, using mCHH levels in vegetative tissues from a previous study (Tan et al. 2016). We divided the genome into nonoverlapping 100-bp loci, then identified loci with an average mCHH of at least 5%. Eight percent of the genome, or 315,368 loci, met these criteria, and 47% of seedling shoot 24-nt siRNAs overlapped with these loci. After removing seedling 24-nt siRNAs from mCHH loci, the residual 24-nt siRNAs in seedling shoot were relatively depleted near genes, but again there was no specific enrichment in heterochromatic regions, as was the case with sperm, nor was there a pattern resembling that of egg cells (Supplemental Fig. S6).

The mRNA expression of sperm cell genes has an atypical distribution, with a smaller number of genes accounting for a larger proportion of mRNA in sperm cell than in egg cell (Supplemental Fig. S10A; Anderson et al. 2013). Since abundance of 24-nt siRNAs and mCHH near genes has been positively correlated with gene expression (Li et al. 2015), we wondered whether highly expressed sperm genes might be enriched for flanking 24-nt siRNAs. This was not the case, however. Genes with values of greater than 10 transcripts per million (TPM), which correspond to the top ~25% of mRNA expression level in sperm cell, exhibited a similar lack of flanking siRNAs as the total set of genes (Supplemental Fig. S10B). We also asked whether differences in siRNAs could be explained by differences in mRNA levels of RdDM factors or DDM1 in egg cell and sperm cell. Analysis of published gamete transcriptomes did not reveal any obvious explanation for egg cell or sperm cell siRNA patterns, though they did show differences in mRNA levels for such RdDM factors as RDR2 (Supplemental Fig. S11; Anderson et al. 2013, 2017). At least one *DDM1* homolog was strongly expressed in both gametes; however, in pollen vegetative cell, a single homolog was only moderately expressed. Analysis of 5′ nucleotide preferences of gametic and vegetative 24-nt siRNAs revealed a strong bias toward a 5′ A in seedling shoot and in sperm cell (Supplemental Fig. S12), consistent with known Pol IV activity (Blevins et al. 2015; Zhai et al. 2015a). The 5′ nucleotide biases were also present in ovary and egg cell, but less prominently than in seedling shoot or sperm cell, suggesting that maternal 24-nt siRNAs might also arise from different RNA polymerase activities or processing of siRNA precursors (both Pol II and Pol IV and multiple dicers and other factors can interact to produce diverse siRNAs [Cuerda-Gil and Slotkin 2016]). In sperm, 21- to 22-nt siRNAs also exhibited a preference for a 5′ A, though weaker than that of 24-nt siRNAs, consistent with reports that Pol IV can produce 21-nt siRNAs in *Arabidopsis* pollen (Borges et al. 2018; Martinez et al. 2018).

Transposon-related mRNAs made up a higher proportion of the total mRNA in sperm than in egg, zygote, or seedling (Supplemental Fig. S13; Anderson et al. 2013, 2017). This was true both for heterochromatic transposons (represented by LTR retrotransposons of the *Gypsy* superfamily) and euchromatic ones (represented by TIR transposons, except CACTA). Analysis of siRNA and mRNA expression of individual transposon copies is complicated by the limitation in mapping short reads to them uniquely. However, a strong trend was evident in that transposon copies with high mRNA expression in gametes tended to have fewer 24-nt siRNAs and vice versa (Supplemental Fig. S14). Both *Gypsy* and TIR transposons exhibited this relationship, suggesting that 24-nt siRNAs might be generally antagonistic to transposon mRNA accumulation.

Gametes have similar DNA methylation as vegetative tissues

Twenty-four-nucleotide siRNAs are key components of RdDM in vegetative tissues. To test whether the unusual genomic distributions of siRNAs in sperm and egg were associated with RdDM, we prepared and sequenced whole-genome bisulfite sequencing libraries using a post-bisulfite adapter tagging (PBAT) method from isolated egg cell and sperm cell (Supplemental Table S2). We also analyzed published PBAT libraries from egg cell, sperm cell, central cell, and pollen vegetative nuclei (Park et al. 2016; Kim et al. 2019). The DNA methylation in all of these resembled typical vegetative tissues in that mCHH was highest near genes (Fig. 3A; Supplemental Figs. S15, S16). Despite the genome-wide differences in distributions of 24-nt siRNAs in egg cell, sperm cell, and

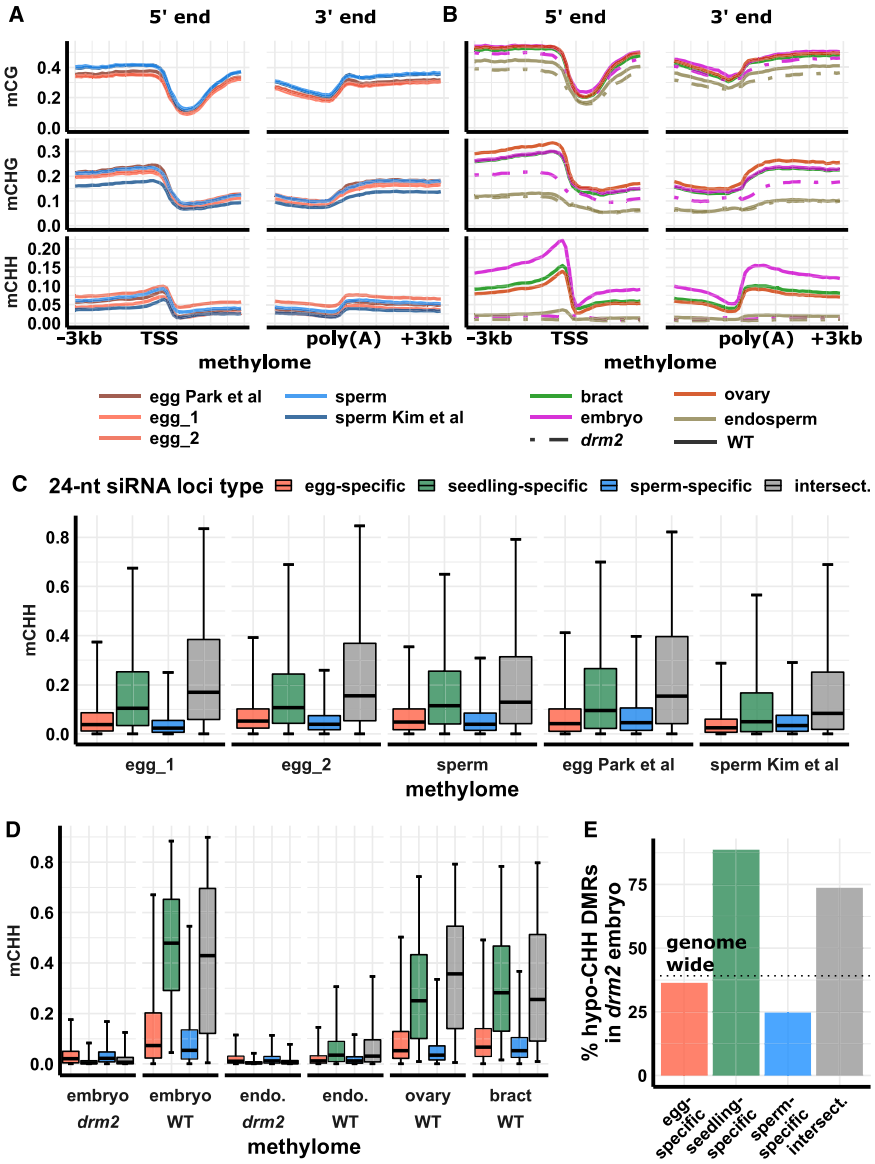


Figure 3. DNA methylation across tissues. (*A*) DNA methylation metagene plots for post-bisulfite adapter tagging (PBAT) libraries. Plots indicate average DNA methylation values over 100-bp intervals. Vertical grid lines indicate 500-bp intervals. DNA methylation is measured as the proportion of methylated cytosines relative to total cytosines in each sequence context. (TSS) Transcription start site, (poly[A]) polyadenylation site. Published data from Park et al. (2016) and Kim et al. (2019). (*B*) DNA methylation metagene plot for conventional WGBS libraries; *x*- and *y*-axes as in *A*. (*C*) DNA methylation of 24-nt siRNA loci for PBAT libraries. *Center* line is median; box spans interquartile range; whiskers span 2.5th to 97.5th percentiles. (*D*) DNA methylation of 24-nt siRNA loci for conventional WGBS libraries, as in *C*; color code as in *C*. (*E*) Abundance of hypomethylated CHH differential methylated regions (hypo-CHH DMRs) in *drm2* mutant relative to wild-type mature embryo. Abundance is measured as the percent of hypomethylated regions relative to total number of regions with sufficient read coverage for statistical significance. Dotted line indicates genome-wide average percent hypo-CHH DMRs. Color code as in *C*.

vegetative tissues, distributions of DNA methylation were similar between them, though the magnitude of methylation was variable between cell types and even between experiments. We also prepared libraries from whole ovary, bract (palea and lemma), embryo, and endosperm using a conventional WGBS method (Supplemental Table S2). All wild-type conventional libraries revealed the expected pattern of high mCHH flanking genes except mature endosperm (Fig. 3B). As with egg cell, the ovary retained robust mCHH at canonical RdDM loci, despite loss of 24-nt siRNAs. We also generated a knockout mutant for *drm2* using CRISPR-Cas9 gene editing and prepared conventional WGBS libraries from mutant endosperm and embryo. *DRM2* encodes a DNA methyltransferase of the Domains Rearranged Methyltransferase family, which function in RdDM. Near-gene mCHH was almost abolished in *drm2* embryo and endosperm (Fig. 3B), demonstrating that the mCHH peaks near genes require functional *DRM2* and consistent with prior work on methylation in *drm2* mutant leaf (Tan et al. 2016).

To test whether 24-nt siRNAs directed DNA methylation in gametes, we examined gamete DNA methylation in the tissue-specific 24-nt siRNA loci that we identified (see above). All tissue-specific classes of 24-nt siRNA loci had high mCG and mCHG in all wild-type tissues, including gametes (Supplemental Fig. S17). Endosperm mCHG was an exception, being reduced by more than 50% relative to other vegetative tissues. mCHH was more variable across tissues, with endosperm having the lowest levels. Across all tissues, seedling-specific and intersection loci had the highest mCHH levels, independently of 24-nt siRNAs. In other words, despite the relative abundance of 24-nt siRNAs in egg-specific 24-nt siRNA loci, these loci had lower mCHH in egg than seedling-specific loci in egg. The same was true for sperm: Sperm-specific 24-nt siRNA loci had lower mCHH in sperm than seedling-specific loci in sperm (Fig. 3C). Intersection loci, like seedling-specific loci, had high mCHH across tissues, including pollen vegetative cell, central cell, and vegetative tissues (Supplemental Fig. S17). Although variation between tissues and even between experiments exists (Fig. 3A), within-library patterns are reproducible and confirm that redistribution of 24-nt siRNAs is not coupled to redistribution of DNA methylation in gametes (Fig. 3C). In all cases, differences between gamete-specific loci and seedling-specific loci were highly significant ($P < 10^{-5}$, Tukey test using a generalized linear model with logit link). In addition, the *drm2* mutation abolished mCHH in seedling-specific 24-nt loci and intersection 24-nt loci in mature embryo and leaf, further confirming that these are canonical RdDM loci (Figs. 3D, 4A). We also confirmed our inferences by counting the proportion of differentially methylated regions (DMRs) in the CHH context between mature WT

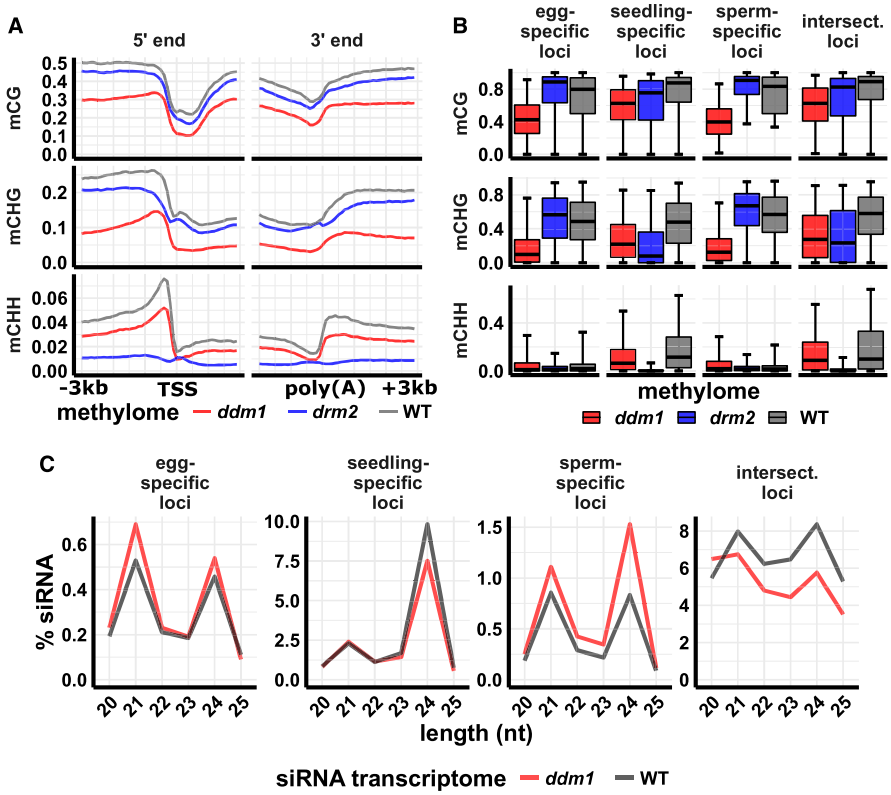


Figure 4. siRNA and DNA methylation in *ddm1* mutant leaf. (*A*) DNA methylation metagene plots. Plots indicate average DNA methylation values over 100-bp intervals in *ddm1* (*osddm1a osddm1b* double mutant), *drm2*, and wild-type leaf. Vertical grid lines indicate 500-bp intervals. DNA methylation is measured as the proportion of methylated cytosines relative to total cytosines in each sequence context. (TSS) Transcription start site, (poly[A]) polyadenylation site. Data source: Tan et al. (2016). (*B*) DNA methylation of 24-nt siRNA loci. *Center* line is median; box spans interquartile range; whiskers span 2.5th to 97.5th percentile. (*C*) siRNA abundance at 24-nt siRNA loci in *ddm1* and wild-type leaf. "Intersect." refers to intersection loci, those identified as 24-nt siRNA in all three tissues. *y*-axis values are percent of siRNAs of each length relative to the total number of siRNAs that mapped to the genome. Error bars are 95% confidence intervals. Data source: Tan et al. (2018).

and *drm2* embryo across 24-nt siRNA loci categories (Fig. 3E). Genome-wide, 39% of eligible loci (loci that had sufficient coverage to be included in DMR analysis) were hypomethylated for mCHH in the *drm2* embryo. Seedling-specific and intersection loci were enriched for hypo-mCHH DMRs (74% and 89%, respectively), whereas egg-specific and sperm-specific loci were depleted for hypo-mCHH DMRs (36% and 25%, respectively). Hypergeometric tests detected significant enrichment or depletion of DMRs in these sets of loci, respectively (P value = 0 for all four categories).

Gamete siRNAs mark loci dependent on DDM1, but not DRM2, for methylation in vegetative tissues

Loss of siRNAs from TIR DNA transposons and gain of siRNAs from LTR retrotransposons in gametes is reminiscent of *ddm1* mutant phenotypes in vegetative tissues in diverse angiosperms (Creasey et al. 2014; McCue et al. 2015; Corem et al. 2018; Fu et al. 2018; Long et al. 2018; Tan et al. 2018). In mutants such as *ddm1* that reduce heterochromatic DNA methylation, RNA polymerases may gain access to DNA from which they are normally restricted by inaccessible chromatin structure. The leaf phenotype of an *osddm1a osddm1b* double mutant in rice (which we will refer to simply as *ddm1*) includes substantial loss of mCHH and siRNAs flanking genes, coupled with increased DRM2 activity and siRNAs in heterochromatic regions (Tan et al. 2016, 2018). The reported differences in siRNAs between wild-type and *ddm1* mutant leaf were not as severe as the differences we found between wild-type seedling shoot and wild-type gametes. Since the magnitude of such differences can be influenced by analysis methods, we reexamined the published *ddm1* small RNA and methylation data in parallel with our gamete data. We also included published *drm2* methylation data from the same study as the *ddm1* mutant methylation data (Tan et al. 2016). Our analysis of methylation and siRNAs confirmed the prior study in that 24-nt siRNAs were reduced in gene flanks in the *ddm1* mutant, but not nearly as severely as in gametes (Supplemental Fig. S18). In agreement with the previous study, the *ddm1* mutant also had a strong reduction in methylation (~50% lower in all contexts), in gene flanks in leaf (Fig. 4A). The *drm2* mutant had a stronger reduction in mCHH but weaker reduction in mCG and mCHG near genes, as expected for a factor directly involved in RdDM (Fig. 4A).

To investigate whether gamete-specific 24-nt siRNA loci might be targets of DDM1 activity, we also evaluated DNA methylation in the mutants for each set of 24-nt siRNA loci. We found severe reductions in methylation—particularly in mCHG—in *ddm1* mutants ($P < 10^{-5}$, Tukey test using a generalized linear model with logit link), but not in *drm2* mutants, at sperm-specific and egg-specific 24-nt siRNA loci (Fig. 4B). At these loci, spread of mCHH values was more variable in the *ddm1* mutant than in wild type, and thus it was more difficult to confidently determine the overall trend. The seedling-specific and intersection 24-nt siRNA loci were strongly dependent on DRM2 for mCHH methylation ($P < 10^{-5}$, Tukey test using a generalized linear model with logit link), consistent with canonical RdDM activity. The relative abundance of both 21-nt and 24-nt siRNAs increased in gamete-specific 24-nt siRNA loci, but 24-nt siRNAs decreased in seedling-specific and intersection loci in *ddm1* leaf (Fig. 4C). Taken together, these analyses reveal that the *ddm1* mutant causes leaf to resemble gametes in terms of siRNA expression and that loci that gain siRNA expression in either gamete are dependent on DDM1, but not DRM2, for methylation in vegetative tissues.

Discussion

The male and female gametes of flowering plants are highly dimorphic, which is reflected in their divergent transcriptomes (Anderson et al. 2013), as well as in their chromatin and histones (Ingouff et al. 2010), suggestive of major differences in reprogramming of their genomes prior to fertilization. The results from this study show that the small RNA transcriptomes of rice gametes differ significantly from each other, as well as from vegetative tissues. Whereas both sperm and egg cells had reduced abundance of miRNAs relative to siRNAs, each gamete expressed a diverse set of miRNAs (Supplemental Fig. S1). Similarities between miRNA expression patterns between egg and ovary might be indicative of cell-to-cell miRNA mobility, but any such miRNA mobility is unlikely to be general since ovary-specific miRNAs such as miR166 family members were not detectable or were detected at very low levels in the egg cell (Supplemental Fig. S1C). In *Arabidopsis*, expression of miR845 in pollen leads to production of 21- and 22-nt *Gypsy* and *Copia* siRNAs, termed epigenetically activated siRNAs, or easiRNAs, which are hypothesized to be transferred from the pollen vegetative cell into the sperm cell (Creasey et al. 2014; Borges et al. 2018; Martinez et al. 2018). However, we could not find any annotated miRNA or any genomic locus with significant homology to miR845 in rice, suggesting that *Gypsy* siRNAs in rice gametes are initiated differently. Likewise, 21- and 22-nt siRNAs were rare in sperm cell, including at gamete-specific 24-nt siRNA loci, as in other rice tissues (Fig. 2E; Supplemental Fig. S4).

The most notable feature of both gametes was the depletion of 24-nt siRNAs from heterochromatin boundaries (also called mCHH islands) (Fig. 5), where they are found in typical vegetative tissues (Gent et al. 2013; Li et al. 2015), constituting key components of canonical RdDM (for review, see Cuerda-Gil and Slotkin

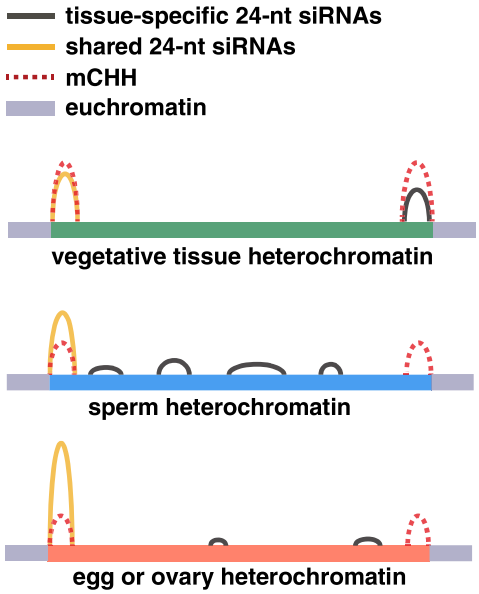


Figure 5. Schematic representation of 24-nt siRNAs and CHH methylation in different tissues. Gray segments represent euchromatic DNA and colored segments represent heterochromatic DNA, the same region in each tissue. Tissue-specific 24-nt siRNAs are abundant in either sperm cell, egg cell, or seedling leaf (tissue-specific loci), while shared 24-nt siRNAs are abundant in all three (intersection loci).

2016). In spite of this loss of 24-nt siRNAs from canonical RdDM loci in both gametes, the abundance of 24-nt siRNAs relative to other siRNA lengths was unchanged. Each gamete, however, had a distinct genomic distribution of 24-nt siRNAs (Fig. 2). The redistribution of 24-nt siRNAs into heterochromatin in the sperm cell might be related to sperm-specific chromatin modifications, which could facilitate access of RNA Pol IV or another polymerase to repetitive loci that would normally be inaccessible. Sperm cells have been previously reported to have a suite of long noncoding RNAs (lncRNAs) distinct from vegetative cells but overlapping with lncRNAs in the *emf2b* polycomb group mutant (Johnson et al. 2018). Alternatively, sperm cell siRNAs could be inherited from progenitor cells or transferred from the pollen vegetative cell, which is known to lose heterochromatic modifications in *Arabidopsis* (Mérai et al. 2014). In sperm cells, egg cells, their companion gametophytic cells, or their progenitors, RdDM may not be fully functional and only methylate DNA at high-affinity loci, which are also shared with vegetative tissues. Gamete-specific 24-nt siRNAs could instead arise from a pathway that processes a Pol II primary transcript rather than a Pol IV primary transcript. Such an interpretation would be consistent with the lack of DNA methylation at gamete-specific sRNA loci.

Changes in chromatin structure also occur in the production of female reproductive cells in plants, which may lead to corresponding changes in small RNA expression or activity (for review, see Wang and Köhler 2017). It is tempting to speculate that the absence of siRNAs from canonical RdDM loci in egg cell may be related to an absence of heterochromatin, and hence heterochromatin boundaries in egg cell, but this explanation would also require a similar absence of heterochromatin in ovaries to explain the similarities between egg cell and ovary 24-nt siRNAs. Alternatively, the chromatin status of egg may have more to do with chromatin in the ovary than egg cell. While it has been proposed that siRNAs are transferred from the central cell into the egg cell (Ibarra et al. 2012), the similarities between ovary and egg cell raise the possibility that 24-nt siRNAs in the egg cell might be acquired from the female reproductive tissues of the ovary.

A recent report on DNA methylation in rice pollen utilizing mutants in the DNA glycosylase ROS1 revealed a correlation between demethylation in the pollen vegetative nucleus and increased methylation in sperm (Ibarra et al. 2012; Kim et al. 2019), which is consistent with previous findings in *Arabidopsis* (Ibarra et al. 2012). In this study, comparison of rice sperm cell PBAT methylomes with a previously published rice egg cell PBAT methylome (Park et al. 2016) and methylomes of vegetative tissues by conventional WGBS methods detected different magnitudes of DNA methylation across these cell types, described as methylation reprogramming in the gametes (Park et al. 2016; Kim et al. 2019). We find from analysis of the published data from Kim et al. (2019) that they are consistent with our PBAT methylation libraries in this study but that the deduced differences in methylation in sperm and egg are subtle in comparison to the differences in 24-nt siRNAs (cf. Fig. 2 and Fig. 3). More specifically, our analysis also found magnitudes of DNA methylation to be different between the different cell types, but the overall genome-wide patterns remained largely unchanged in all wild-type tissues we examined except mature endosperm (Fig. 3; summarized in Fig. 5). We propose that distinct and major reprogramming of 24-nt siRNAs occurs both maternally and paternally in the germline, which are not correlated with a similar reprogramming of DNA methylation in the gametes. Since mCHH requires 24-nt siRNAs, one explanation for the presence of mCHH at heterochromatin boundaries without 24-nt siRNAs might be the removal of siRNAs after the deposition of DNA methylation. It is also possible that the observed mCHH pattern was inherited from progenitor cells, in which case immediate progenitor cells would have had double the levels of mCHH.

DNA methylation in *drm2* mutants confirmed that mCHH and 24-nt siRNAs in vegetative tissues defined canonical RdDM loci, as they were strongly dependent upon *DRM2* for mCHH. The siRNA expression patterns of neither sperm nor egg cell could be explained by simple loss of 24-nt siRNAs from canonical RdDM loci because 24-nt siRNAs remained abundant relative to other siRNA lengths and even increased in abundance relative to miRNAs. In addition, removal of 24-nt siRNAs from canonical RdDM loci in silico and analysis of the residual ones in seedling shoot yielded a genomic distribution unlike either gamete (Supplemental Fig. S6). Thus, the pattern in both egg and sperm cell requires both loss of 24-nt siRNAs from canonical RdDM loci and gain of siRNAs at novel loci that do not undergo RdDM. In this respect, gametes share similarities with *ddm1* mutants in *Arabidopsis*, tomato, rice, and maize (Creasey et al. 2014; McCue et al. 2015; Corem et al. 2018; Long et al. 2018; Tan et al. 2018) and with chromomethylase mutants in maize (Fu et al. 2018), where loss of siRNAs at heterochromatin boundaries is accompanied by gain elsewhere. DDM1 facilitates methylation of heterochromatic DNA and has no known direct function in RdDM (Zemach et al. 2013; Lyons and Zilberman 2017). Our analysis of published *ddm1* mutant rice data (Tan et al. 2016, 2018) revealed that the loci with novel 24-nt siRNA expression in either gamete tended to also gain siRNAs in *ddm1* mutant leaves, not just 24-nt siRNAs but also 21-nt siRNAs (Fig. 4). Both mCHH and 24-nt siRNAs in *ddm1* mutant leaves were reduced at canonical RdDM loci, as previously reported. The effect of *ddm1* on siRNAs in leaves was not as strong as the difference between wild-type leaf and gametes, but the mutants clearly shifted leaf toward a more gamete-like profile. In maize *ddm1* mutant embryos, the effect is much stronger than in *ddm1* mutant rice leaf and more similar to rice gametes in that 24-nt siRNAs are nearly completely lost from canonical RdDM loci (Fu et al. 2018).

The finding that a transgenic *DDM1::GFP* construct under the native *DDM1* promoter expresses in *Arabidopsis* sperm but not in pollen vegetative nucleus supports a model in which the absence of DDM1 in the pollen vegetative nucleus facilitates production of 21- and 22-nt siRNAs that are transferred into the sperm cell (Slotkin et al. 2009). While it is not clear how high levels of mCHG and mCG in pollen vegetative nucleus are produced in the absence of DDM1 (Calarco et al. 2012; Ibarra et al. 2012), it is clear that pollen vegetative nucleus undergoes a major loss of heterochromatin in *Arabidopsis* (Mérai et al. 2014). The mRNA expression data in rice (Anderson et al. 2013) do not clearly indicate the *DDM1* status in the pollen vegetative nucleus, as it has reduced but not absent mRNA expression (Supplemental Fig. S11). Whether siRNAs of any length are transferred from the pollen vegetative nucleus into the sperm cell in rice is not known. Regardless of the origin of 24-nt siRNAs in rice gametes, their lack of correlation with mCHH in these cells raises two possibilities. The first is that they direct chromatin modifications other than DNA methylation in gametes. The second is that the gametic siRNAs are primed for RdDM in the early embryo. In this case, one or more RdDM factors missing in each gamete might become available upon fertilization and allow for immediate activity at loci not usually targeted in vegetative tissues.

Methods

Kitaake variety rice was used for all experiments. Gametes were isolated as described (Anderson et al. 2013; Li et al. 2019). Ovaries were dissected from pre-anthesis flowers. Seedling shoot segments were collected from 7-d-old water-germinated rice seeds. Total RNA was used as input for small RNA libraries: ~30 ng for egg cells, ~1 ng for sperm cells, ~50 ng for ovaries, and ~20 ng for seedling shoots. Small RNA libraries were made using the NEXTflex Small RNA-seq kit v3 (PerkinElmer NOVA-5132-05). Sperm samples and one of the four seedling replicates were amplified with 25 PCR cycles, all other samples with 15–20 cycles. After adapter trimming small RNA reads, PCR duplicates were removed using PRINSEQ "prinseq-lite.pl -fastq -out_format 3 -out_good -derep 1" (Schmieder and Edwards 2011). Reads were aligned to the genome with BWA-backtrack (version 0.7.15) (Li and Durbin 2009). Except where indicated otherwise, multimapping reads were included in all analyses. The complete set of read alignments was compared with miRNA, phasiRNA, tRNA, 5S rRNA, and NOR RNA loci in the genome (Supplemental Tables S4, S5). All other reads were categorized as siRNA reads and used for subsequent siRNA analyses. To evaluate reproducibility of low-input libraries, pairwise Pearson's correlation coefficients were calculated using the cor() function in R (R Core Team 2019). For this analysis, 24-nt siRNAs were first counted on 50-kb intervals (Supplemental Fig. S19A–D). Hierarchical clustering was performed using the hclust() function in R on these data and revealed the expected clustering by tissue (Supplemental Fig. S19E).

Whole-genome small RNA heat maps were made on 50-kb intervals using IGVtools (Thorvaldsdóttir et al. 2013). For better visualization of midrange values, heat map intensity was maxed out at 1.25× coverage (per 10 million siRNAs). To identify representative 24-nt siRNA loci, reads alignments were subsampled then combined from each sperm cell sample, from each egg cell sample, and from each seedling shoot sample to get as equal a representation as possible from each sample and a final combined number of two million in each. The genome was divided into nonoverlapping 100-bp loci, and siRNAs were counted for every locus. All loci that had at least three mapping 24-nt siRNAs that spanned at least one-third (34 bp) of the 100 bp were categorized as 24-nt siRNA loci; 5′ nucleotide frequencies were calculated with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

PBAT WGBS libraries were prepared using Pico Methyl-Seq Library Prep kits (Zymo, D5456). Conventional WGBS libraries from endosperm, embryo, and ovary were prepared using the methylC-seq method (Urich et al. 2015). MethylC-seq and PBAT reads were quality-filtered and trimmed of adapters using Cutadapt (Martin 2011), and reads were aligned to the genome with BS-Seeker2 (version 2.1.5) (Guo et al. 2013). For all reads except paired-end, previously published reads (Tan et al. 2016), PCR duplicates were removed prior to alignment with the BS-Seeker2 FilterReads.py module. DMRs were identified with CGmapTools, version 0.1.1 (Guo et al. 2018).

Additional details on tissue collection, CRISPR-Cas9 mutagenesis of *drm2*, small RNA library construction and sequencing analysis, mRNA sequencing analysis, mock egg isolations and qPCR quantification, genome annotations, and WGBS library construction and sequencing analysis are included in Supplemental Methods.

Data access

All raw sequencing data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA533115. All R scripts for statistical analyses and data visualization can be found at https://github.com/cxli233/gamete_smRNA_revision and as Supplemental Code.

Acknowledgments

We thank Daniel Jones for assistance in sperm isolation; Zach Liechty, Christian Santos, and Joseph Edwards for assistance in R programming; and Alina Yalda and Jake Anichowski for greenhouse maintenance. The UC Davis Genome Center provided Illumina sequencing services. The UC Davis Plant Transformation Facility provided plant transformation services. 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).

References

Allen RS, Li J, Alonso-Peral MM, White RG, Gubler F, Millar AA. 2010. MicroR159 regulation of most conserved targets in *Arabidopsis* has negligible phenotypic effects. *Silence* **1:** 18. doi:10.1186/1758-907X-1-18

Anderson SN, Johnson CS, Jones DS, Conrad LJ, Gou X, Russell SD, 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. doi:10.1111/tpj.12336

Anderson SN, Johnson CS, Chesnut J, Jones DS, Khanday I, Woodhouse M, Li C, Conrad LJ, Russell SD, Sundaresan V. 2017. The zygotic transition is initiated in unicellular plant zygotes with asymmetric activation of parental genomes. *Dev Cell* **43:** 349–358.e4. doi:10.1016/j.devcel.2017.10.005

Autran D, Baroux C, Raissig MT, Lenormand T, Wittig M, Grob S, Steimer A, Barann M, Klostermeier UC, Leblanc O, et al. 2011. Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. *Cell* **145:** 707–719. doi:10.1016/j.cell.2011.04.014

Blevins T, Podicheti R, Mishra V, Marasco M, Wang J, Rusch D, Tang H, Pikaard CS. 2015. Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in *Arabidopsis*. *eLife* **4:** e09591. doi:10.7554/eLife.09591

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

Borges F, Parent JS, van Ex F, Wolff P, Martínez G, Köhler C, Martienssen RA. 2018. Transposon-derived small RNAs triggered by miR845 mediate genome dosage response in *Arabidopsis*. *Nat Genet* **50:** 186–192. doi:10.1038/s41588-017-0032-5

Calarco JP, Borges F, Donoghue MT, Van Ex F, Jullien PE, Lopes T, Gardner R, Berger F, Feijó JA, Becker JD, et al. 2012. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell* **151:** 194–205. doi:10.1016/j.cell.2012.09.001

Corem S, Doron-Faigenboim A, Jouffroy O, Maumus F, Arazi T, Bouché N. 2018. Redistribution of CHH methylation and small interfering RNAs across the genome of tomato *ddm1* mutants. *Plant Cell* **30:** 1628–1644. doi:10.1105/tpc.18.00167

Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, Martienssen RA. 2014. miRNAs trigger widespread epigenetically activated siRNAs from transposons in *Arabidopsis*. *Nature* **508:** 411–415. doi:10.1038/nature13069

Cuerda-Gil D, Slotkin RK. 2016. Non-canonical RNA-directed DNA methylation. *Nat Plants* **2:** 16163. doi:10.1038/nplants.2016.163

Del Toro-De León G, García-Aguilar M, Gillmor CS. 2014. Non-equivalent contributions of maternal and paternal genomes to early plant embryogenesis. *Nature* **514:** 624–627. doi:10.1038/nature13620

Dooner HK, Wang Q, Huang JT, Li Y, He L, Xiong W, Du C. 2019. Spontaneous mutations in maize pollen are frequent in some lines and arise mainly from retrotranspositions and deletions. *Proc Natl Acad Sci* **116:** 10734–10743. doi:10.1073/pnas.1903809116

Erdmann RM, Satyaki PRV, Klosinska M, Gehring M. 2017. A small RNA pathway mediates allelic dosage in endosperm. *Cell Rep* **21:** 3364–3372. doi:10.1016/j.celrep.2017.11.078

Fei Q, Yang L, Liang W, Zhang D, Meyers BC. 2016. Dynamic changes of small RNAs in rice spikelet development reveal specialized reproductive phasiRNA pathways. *J Exp Bot* **67:** 6037–6049. doi:10.1093/jxb/erw361

Fu FF, Dawe RK, Gent JI. 2018. Loss of RNA-directed DNA methylation in maize chromomethylase and DDM1-type nucleosome remodeler mutants. *Plant Cell* **30:** 1617–1627. doi:10.1105/tpc.18.00053

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

Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, Dawe RK. 2013. CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. *Genome Res* **23:** 628–637. doi:10.1101/gr.146985.112

Grant-Downton R, Kourmpetli S, Hafidh S, Khatab H, Le Trionnaire G, Dickinson H, Twell D. 2013. Artificial microRNAs reveal cell-specific differences in small RNA activity in pollen. *Curr Biol* **23:** R599–R601. doi:10.1016/j.cub.2013.05.055

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

Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, Chen PY, Pellegrini M. 2013. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* **14:** 774. doi:10.1186/1471-2164-14-774

Guo W, Zhu P, Pellegrini M, Zhang MQ, 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. doi:10.1093/bioinformatics/btx595

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

Hsieh PH, He S, Buttress T, Gao H, Couchman M, Fischer RL, Zilberman D, Feng X. 2016. *Arabidopsis* male sexual lineage exhibits more robust maintenance of CG methylation than somatic tissues. *Proc Natl Acad Sci* **113:** 15132–15137. doi:10.1073/pnas.1619074114

Ibarra CA, Feng X, Schoft VK, Hsieh TF, Uzawa R, Rodrigues JA, Zemach A, Chumak N, Machlicova A, Nishimura T, et al. 2012. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* **337:** 1360–1364. doi:10.1126/science.1224839

Ingouff M, Rademacher S, Holec S, Šoljić L, Xin N, Readshaw A, Foo SH, Lahouze B, Sprunck S, Berger F. 2010. Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in *Arabidopsis*. *Curr Biol* **20:** 2137–2143. doi:10.1016/j.cub.2010.11.012

Ingouff M, Selles B, Michaud C, Vu TM, Berger F, Schorn AJ, Autran D, Van Durme M, Nowack MK, Martienssen RA, et al. 2017. Live-cell analysis of DNA methylation during sexual reproduction in *Arabidopsis* reveals context and sex-specific dynamics controlled by noncanonical RdDM. *Genes Dev* **31:** 72–83. doi:10.1101/gad.289397.116

Johnson C, Kasprzewska A, Tennessen K, Fernandes J, Nan GL, Walbot V, Sundaresan V, Vance V, Bowman LH. 2009. Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. *Genome Res* **19:** 1429–1440. doi:10.1101/gr.089854.108

Johnson C, Conrad LJ, Patel R, Anderson S, Li C, Pereira A, Sundaresan V. 2018. Reproductive long intergenic noncoding RNAs exhibit male gamete specificity and Polycomb Repressive Complex 2-mediated repression. *Plant Physiol* **177:** 1198–1217. doi:10.1104/pp.17.01269

Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, et al. 2013. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice (N Y)* **6:** 4. doi:10.1186/1939-8433-6-4

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

Kirkbride RC, Lu J, Zhang C, Mosher RA, Baulcombe DC, Chen ZJ. 2019. Maternal small RNAs mediate spatial-temporal regulation of gene expression, imprinting, and seed development in *Arabidopsis*. *Proc Natl Acad Sci* **116:** 2761–2766. doi:10.1073/pnas.1807621116

Komiya R, Ohyanagi H, Niihama M, Watanabe T, Nakano M, Kurata N, Nonomura K. 2014. Rice germline-specific Argonaute MEL1 protein binds to phasiRNAs generated from more than 700 lincRNAs. *Plant J* **78:** 385–397. doi:10.1111/tpj.12483

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

Li Q, Gent JI, Zynda G, Song J, Makarevitch I, Hirsch CD, Hirsch CN, Dawe RK, Madzima TF, McGinnis KM, et al. 2015. RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. *Proc Natl Acad Sci* **112:** 14728–14733. doi:10.1073/pnas.1514680112

Li X, Shahid MQ, Xia J, Lu Z, Fang N, Wang L, Wu J, Chen Z, Liu X. 2017. Analysis of small RNAs revealed differential expressions during pollen and embryo sac development in autotetraploid rice. *BMC Genomics* **18:** 129. doi:10.1186/s12864-017-3526-8

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

Long JC, Xia AA, Liu JH, Jing JL, Wang YZ, Qi CY, He Y. 2018. *Decrease in DNA methylation 1* (*DDM1*) is required for the formation of ^{m}CHH islands in maize. *J Integr Plant Biol* **61:** 749–764. doi:10.1111/jipb.12733

Lyons DB, Zilberman D. 2017. DDM1 and Lsh remodelers allow methylation of DNA wrapped in nucleosomes. *eLife* **6:** e30674. doi:10.7554/eLife.30674

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17:** 3. doi:10.14806/ej.17.1.200

Martinez G, Wolff P, Wang Z, Moreno-Romero J, Santos-González J, Conze LL, DeFraia C, Slotkin RK, Köhler C. 2018. Paternal easiRNAs regulate parental genome dosage in *Arabidopsis*. *Nat Genet* **50:** 193–198. doi:10.1038/s41588-017-0033-4

McCue AD, Panda K, Nuthikattu S, Choudury SG, Thomas EN, Slotkin RK. 2015. ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. *EMBO J* **34:** 20–35. doi:10.15252/embj.201489499

Mérai Z, Chumak N, García-Aguilar M, Hsieh TF, Nishimura T, Schoft VK, Bindics J, Slusarz L, Arnoux S, Opravil S, et al. 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. doi:10.1073/pnas.1418564111

Nagaki K, Talbert PB, Zhong CX, Dawe RK, Henikoff S, Jiang J. 2003. Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. *Genetics* **163:** 1221–1225.

Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, Rohr NA, Rambani A, Burke JM, Udall JA, et al. 2016. Widespread natural variation of DNA methylation within angiosperms. *Genome Biol* **17:** 194. doi:10.1186/s13059-016-1059-0

Nonomura K, Morohoshi A, Nakano M, Eiguchi M, Miyao A, Hirochika H, Kurata N. 2007. A germ cell–specific gene of the *ARGONAUTE* family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. *Plant Cell* **19:** 2583–2594. doi:10.1105/tpc.107.053199

Park K, Kim MY, Vickers M, Park JS, Hyun Y, Okamoto T, Zilberman D, Fischer RL, Feng X, Choi Y, et al. 2016. DNA demethylation is initiated in the central cells of *Arabidopsis* and rice. *Proc Natl Acad Sci* **113:** 15138–15143. doi:10.1073/pnas.1619047114

Pillot M, Baroux C, Vazquez MA, Autran D, Leblanc O, Vielle-Calzada JP, 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. doi:10.1105/tpc.109.071647

R Core Team. 2019. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.

Russell SD, Gou X, Wong CE, Wang X, Yuan T, Wei X, Bhalla PL, Singh MB. 2012. Genomic profiling of rice sperm cell transcripts reveals conserved and distinct elements in the flowering plant male germ lineage. *New Phytol* **195:** 560–573. doi:10.1111/j.1469-8137.2012.04199.x

Satyaki PRV, Gehring M. 2019. Paternally acting canonical RNA-directed DNA methylation pathway genes sensitize *Arabidopsis* endosperm to paternal genome dosage. *Plant Cell* **31:** 1563–1578. doi:10.1105/tpc.19.00047

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

Schoft VK, 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. *EMBO Rep* **10:** 1015–1021. doi:10.1038/embor.2009.152

She W, Grimanelli D, Rutowicz K, Whitehead MW, Puzio M, Kotliński M, Jerzmanowski A, Baroux C. 2013. Chromatin reprogramming during the somatic-to-reproductive cell fate transition in plants. *Development* **140:** 4008–4019. doi:10.1242/dev.095034

Slotkin RK, Vaughn M, Borges F, Tanurdžić M, Becker JD, Feijó JA, Martienssen RA. 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136:** 461–472. doi:10.1016/j.cell.2008.12.038

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE. 2014. Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat Struct Mol Biol* **21:** 64–72. doi:10.1038/nsmb.2735

Tan F, Zhou C, Zhou Q, Zhou S, Yang W, Zhao Y, Li G, Zhou DX. 2016. Analysis of chromatin regulators reveals specific features of rice DNA

methylation pathways. *Plant Physiol* **171:** 2041–2054. doi:10.1104/pp.16.00393
Tan F, Lu Y, Jiang W, Wu T, Zhang R, Zhao Y, Zhou DX. 2018. DDM1 represses noncoding RNA expression and RNA-directed DNA methylation in heterochromatin. *Plant Physiol* **177:** 1187–1197. doi:10.1104/pp.18.00352
Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14:** 178–192. doi:10.1093/bib/bbs017
Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR. 2015. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat Protoc* **10:** 475–483. doi:10.1038/nprot.2014.114
Walker J, Gao H, Zhang J, Aldridge B, Vickers M, Higgins JD, Feng X. 2018. Sexual-lineage-specific DNA methylation regulates meiosis in *Arabidopsis*. *Nat Genet* **50:** 130–137. doi:10.1038/s41588-017-0008-5
Wang G, Köhler C. 2017. Epigenetic processes in flowering plant reproduction. *J Exp Bot* **68:** 797–807. doi:10.1093/jxb/erw486
Xia R, Chen C, Pokhrel S, Ma W, Huang K, Patel P, Wang F, Xu J, Liu Z, Li J, et al. 2019. 24-nt reproductive phasiRNAs are broadly present in angiosperms. *Nat Commun* **10:** 627. doi:10.1038/s41467-019-08543-0
Yelagandula R, Stroud H, Holec S, Zhou K, Feng S, Zhong X, Muthurajan UM, Nie X, Kawashima T, Groth M, et al. 2014. The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in *Arabidopsis*. *Cell* **158:** 98–109. doi:10.1016/j.cell.2014.06.006
Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D. 2013. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153:** 193–205. doi:10.1016/j.cell.2013.02.033
Zhai J, Bischof S, Wang H, Feng S, Lee TF, Teng C, Chen X, Park SY, Liu L, Gallego-Bartolome J, et al. 2015a. A one precursor one siRNA model for Pol IV-dependent siRNA biogenesis. *Cell* **163:** 445–455. doi:10.1016/j.cell.2015.09.032
Zhai J, Zhang H, Arikit S, Huang K, Nan GL, Walbot V, Meyers BC. 2015b. Spatiotemporally dynamic, cell-type–dependent premeiotic and meiotic phasiRNAs in maize anthers. *Proc Natl Acad Sci* **112:** 3146–3151. doi:10.1073/pnas.1418918112
Zhao Y, Wang S, Wu W, Li L, Jiang T, Zheng B. 2018. Clearance of maternal barriers by paternal miR159 to initiate endosperm nuclear division in *Arabidopsis*. *Nat Commun* **9:** 5011. doi:10.1038/s41467-018-07429-x
Zhao P, Zhou X, Shen K, Liu Z, Cheng T, Liu D, Cheng Y, Peng X, Sun MX. 2019. Two-step maternal-to-zygotic transition with two-phase parental genome contributions. *Dev Cell* **49:** 882–893.e5. doi:10.1016/j.devcel.2019.04.016

Received June 13, 2019; accepted in revised form December 23, 2019.

Genome-wide redistribution of 24-nt siRNAs in rice gametes

Chenxin Li, Hengping Xu, Fang-Fang Fu, et al.

Genome Res. published online January 2, 2020
Access the most recent version at doi:10.1101/gr.253674.119

Table of contents

Supplemental Methods

Tissue collection

Rice (*Kitaake* variety) was grown in soil in greenhouses under natural light. Plants were irrigated with deionized water twice a week and supplemented with fertilized water every other week. Gametes were isolated as described (Anderson et al. 2013; Li et al. 2019). Briefly, ovaries were dissected from pre-anthesis flowers. A transverse cut was made at the middle region of the ovary. The lower part of the cut ovary was gently pushed by an acupuncture needle under a phase inverted microscope. Once the egg cell floated out from the ovary incision, it was captured by a fine capillary in a volume of ~1 µl and frozen in liquid nitrogen. For small RNA, 35 – 50 cells were used for each biological replicate. For whole genome bisulfite sequencing (WGBS), 100 egg cells were used for each biological replicate. Six biological replicates were collected for egg cells and two for sperm cells. For sperm cell small RNA and WGBS libraries, around 50 panicles with mature flowers were used for each biological replicate. Eight biological replicates were collected for small RNA and one for WGBS. Ovaries were dissected from pre-anthesis flowers. For both small RNA and WGBS, five ovaries were pooled to make each biological replicate. Three biological replicates were collected for small RNA and six for WGBS. Bracts (lemma and palea) were collected from pre-anthesis flowers. For WGBS, three pairs of bracts were pooled to make each biological replicate and six biological replicates were collected. Seedling shoot segments were collected from 7-day-old water-germinated rice seeds for small RNA libraries, and one seedling was used for each of the four biological replicates. Individual mature endosperms and embryos were used for each biological replicate (three for each genotype). Dry seeds were soaked in 6% NaOH in water at 57°C for 8 min, and pericarps were removed with forceps. The endosperm and embryo were separated then ground to a powder with a mini pestle in a 2-ml Eppendorf tube. DNA was extracted from ovary, bract, endosperm, and embryo with a DNeasy Plant Mini Kit (Qiagen, 69104). Embryos and endosperm mutant for *drm2* were genotyped using the primers DRM2-Ri (TCTCACTACAAAGGCACCATAAAG) and DRM2-48F (CGAGGAGGAGGATGATACTAT) for the 48-bp deletion allele and primers DRM2-Ri and DRM2-52F (CGAGGAGGAGGATGATATG) for the allele 52-bp deletion allele.

Generation of *drm2* mutation using CRISPR-Cas9 and genotyping

CRISPR-Cas9 was used to generate targeted mutations in the rice DRM2 gene (MSU: *LOC_Os03g02010*). Single-guide RNA (sgRNA) sequences, GGAGGAGGATGATACTAATT and GACAGGACTCCTCACTCTGA, were designed using the web tool https://www.genome.arizona.edu/crispr/ (Xie et al. 2014). CRISPR construct assembly was performed as described previously (Khanday et al. 2019). Rice transformation was performed at the UC Davis Plant Transformation Facility. One transgenic line carrying one in-frame 48-bp deletion and one frame-shift 52-bp deletion was selected for further analysis. Since homozygous rice *drm2* mutants were sterile [also previously reported (Moritoh et al. 2012)], we maintained the *drm2* mutation in a segregating population with the in-frame 48-bp deletion. The 48-bp allele was functional, as the plants carrying one or both 48-bp alleles were phenotypically indistinguishable from wild-type. Genotyping was performed using two forward primers, F48 (CGAGGAGGAGGATGATACTAT) and F52 (CGAGGAGGAGGATGATAT). Each specifically amplifies either the 48-bp or 52-bp alleles, with one reverse primer (TCTCACTACAAAGGCACCATAAAG). To genotype each sample, two separate PCR

reactions were performed: F48 + R and F52 + R, at 59°C annealing temperature, 30 PCR cycles. Amplicon sizes were ~500-bp.

RNA extraction and small RNA library construction

RNA extractions were performed using Ambion RNaqueous Total RNA kit (AM1931), including an on-column DNase treatment using Qiagen DNase (79254). Total RNA was run on a Bioanalyzer to check for RNA integrity, using the eukaryotic total RNA-pico program. RNA input was around 30 ng total RNA for egg cells, 1 ng for sperm cells, 50 ng for ovaries, and 20 ng for seedlings. Small RNA libraries were made using the NEXTflex Small RNA-seq kit v3 (NOVA-5132-05), with the following modifications. Since RNA input was low, a 1/4 dilution of adapters was used. The 3' adapter ligation step was performed at 20°C overnight. Sperm libraries were amplified with 25 PCR cycles. All other libaries were amplified with 15 – 20 cycles, except one of the four seedling replicates was amplified with 25 cycles. After amplification, libraries were run on a Bioanalyzer DNA High Sensitivity Assay. Libraries with a 130-bp peak (adapter dimer peak) constituting less than 10% of the 150-bp peak (expected small RNA peak) were ran on a 10% TBE-Acrylamide gel (100 V for 1 hr). Gels were stained and the area around 150 bp was excised and purified according to recommendations of the NEXTflex Small RNA-seq kit.

Mock egg isolations and qPCR quantification

Ovaries were dissected as during egg cell isolation. Rather than collecting an egg cell, about 1 µL of cell-free solution was collected into a microcentrifuge tube and immediately frozen in liquid nitrogen. Thirty collections were combined as a single replicate, and two independent replicates were collected. RNA extraction and library construction were performed as described above. A strong 150-bp band could be seen on a Bioanalyzer gel for a positive control ovary library, but no band for negative water control or mock samples (**Supplemental Fig. S2B**), indicating that the lack of a band is not due to failed library preparation, but lack of sufficient input RNA. To produce a DNA standard for qPCR absolute quantification, the following primers were used: P5_F (AATGATACGGCGACCACCGACATGACATTGACTATAAGGATGACG) and P7_R (CAAGCAGAAGACGGCATACGAGATCGAGGCCGATGCTATACTTT) using a plasmid template (Khanday et al. 2018). The target amplicon was 153bp and 49% CG content. The standard was PCR amplified, gel purified, and quantified using a Nanodrop spectrophotometer. qPCR was performed using SYBR Green Master Mix (Biorad 1725270), with Illumina P5 and P7 universal primers, AATGATACGGCGACCACCGA and CAAGCAGAAGACGGCATACGAGAT, with a 60°C annealing temperature, 30 second elongation time and 35 cycles. Serial dilutions of the standard (10^{-1} to 10^{-7}, ten-fold dilution each step, two technical reps each) were used to fit the standard curve (**Supplemental Fig. S2C**). A 1/10 dilution of each library (three technical reps each) was used as templates for qPCR in the same run. The number of molecules in each library was calculated using the standard curve (**Supplemental Fig. S2D**).

Genome annotations

The Os-Nipponbare-Reference-IRGSP-1.0 reference genome was used for all analyses (Kawahara et al. 2013). MSU7 Rice gene annotations were extracted from the all.gff genome annotation file downloaded from

http://rice.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/all.dir/. Genes that were flagged as transposons were removed, leaving a set of 39,953 genes. Transposons were annotated using RepeatMasker version 4.0.05 (http://www.repeatmasker.org/), parameters as follows: "-gff -species rice -s -pa 8". miRNA annotations were downloaded from miRBase version 22 (Kozomara et al. 2019). To identify locations of the tandem repeat *CentO*, a consensus sequence (Zhang et al. 2013) was aligned to the genome using using Bowtie2 version 2.3.4.1 (Langmead and Salzberg 2012) parameters "--local -a -f". All alignments separated by 100 bp or fewer were merged using BEDTools merge (Quinlan and Hall 2010). Locations of 5S rRNA and tRNA repeats were identified in the same way as *CentO* but using the GenBank reference sequence KM036285.1 for 5S rRNA and a set of tRNA sequences from The tRNAscan-SE Genomic tRNA Database for tRNAs (http://gtrnadb.ucsc.edu/GtRNAdb2/genomes/eukaryota/Osati/Osati-tRNAs.fa.) (Chan and Lowe 2016). For NOR annotation, an 18S ribosomal RNA gene, internal transcribed spacer 1, 5.8S ribosomal RNA gene, internal transcribed spacer 2, and 26S ribosomal RNA gene complete sequence from GenBank (KM036285.1) was aligned to the genome with Bowtie2 version 2.3.4.1, parameters "--local --ma 1 --mp 24,8 --rdg 20,48 --rfg 20,48 -a f". All alignments separated by 100 bp or fewer were merged using BEDTools.

Small RNA sequencing analysis

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 "-u 4" followed by cutadapt "-u -4". Previously published small RNA libraries that did not include 4 random nucleotides at each end were processed similarly, but without removal of PCR duplicates. Reads were aligned to the genome with BWA-backtrack (version 0.7.15) (Li and Durbin 2009), parameters "aln -t 8 -l 10." A single mapped position was kept per input read, regardless of the possibility of mapping to multiple locations. Except where indicated otherwise, multi-mapping reads were included in all analyses. Locations of 21-nt and 24-nt phasiRNA loci were identified using PHASIS version 3.3 (https://www.biorxiv.org/content/10.1101/158832v1). The phasdetect module was run on small RNA reads of each length from 20 to 25nt separately, followed by the phasemerge module, parameters "-mode merge -pval 1e-5" with each length. Only 21 and 24-nt lengths produced detectable phasiRNA loci (2364 21-nt phasiRNA loci and 68 24-nt phasiRNA loci, **Supplemental Table S4**). The complete set of read alignments was compared with miRNA, phasiRNA, tRNA, 5S rRNA, and NOR RNA loci in the genome, and all reads that aligned by at least 90% with any of these was categorized as such using BEDTools intersect. All other reads were categorized as siRNA reads and used for subsequent siRNA analyses. The uniquely mapping subset of siRNAs was defined by having MAPQ values of at least 20 using SAMtools (Li et al. 2009). 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 (**Supplemental Table S5**). Whole-genome small RNA heatmaps were made on 50-Kb intervals using IGVtools (Thorvaldsdóttir et al. 2013). For better visualization of midrange values, heatmap intensity was maxed out at 1.25X coverage (per 10 million 24-nt siRNAs). To identify representative 24-nt siRNA loci, reads

alignments were subsampled then combined from each sperm cell sample, from each egg cell sample, and from each seedling shoot sample to get as equal a representation as possible from each sample and a final combined number of 2 million in each using SAMtools view –s followed by SAMtools merge. The genome was divided into non-overlapping 100-bp loci, and siRNAs were counted for each locus using BEDTools coverage. All loci that had at least three mapping 24-nt siRNAs that spanned at least a third (34 bp) of the 100 bp were categorized as 24-nt siRNA loci. siRNAs that mapped to the intersection of two 100-bp loci contributed toward the siRNA count for both loci. Adjacent 100-bp loci that both qualified as 24-nt siRNA loci were not merged, but instead counted as individual 24-nt siRNA loci. 5' nucleotide frequencies were calculated with FastQC, version 0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

miRNA expression analyses

miRNA expression data were organized into a matrix, with each row as an individual miRNA and each column as a library (**Supplemental Dataset 1**). R package EdgeR was used to analyze miRNA expression (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. Libraries were then further normalized by the TMM method (Robinson and Oshlack 2010), as recommended by the EdgeR package. 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. Clustering analyses (**Supplemental Fig S1** and **Supplemental Table S3**) were performed using hierarchical clustering, and assignment of miRNA into clusters were done using the Dynamic Tree Cut R package (Langfelder et al. 2008).

Preparation of WGBS libraries

PBAT libraries were prepared using Pico Methyl-Seq Library Prep Kits (Zymo D5456). Sperm and egg cell isolates of approximately 100 cells each were diluted in 200 µl of 10 mM Tris, pH 8 in a 1.5-ml tube, then centrifuged for 10 minutes at 16,000 x G at 4°C. The supernatant was removed except for 9 µl at the bottom of the tube. The 9 µl was pipetted up and down 10 times, then transferred to a 0.2-ml tube. 10 µl M-Digestion Buffer and 1 ml Proteinase K (Zymo D3001-2-5) were added and incubated for 20 minutes at 50°C. For pollen vegetative cells, 12 µl was retained after the initial dilution in Tris buffer, and 13 µl Zymo M-Digestion Buffer and 1 ml Zymo Proteinase K (20 mg/mL) were added. Following incubation for 20 minutes at 50°C, pollen vegetative cells were centrifuged again for 5 minutes at 16,873 x G, then 20 µl of supernatant transferred to a new tube. Subsequent steps were as directed in the Pico Methyl-Seq Library Prep kit protocol, version 1.2.0, with a 16-minute incubation in L-Desulphonation buffer and 5:1 ratios of DNA binding buffer in DNA purification steps. Conventional WGBS libraries from endosperm, embryo, ovary, and bract were prepared using the methylC-seq method (Urich et al. 2015).

WGBS analysis

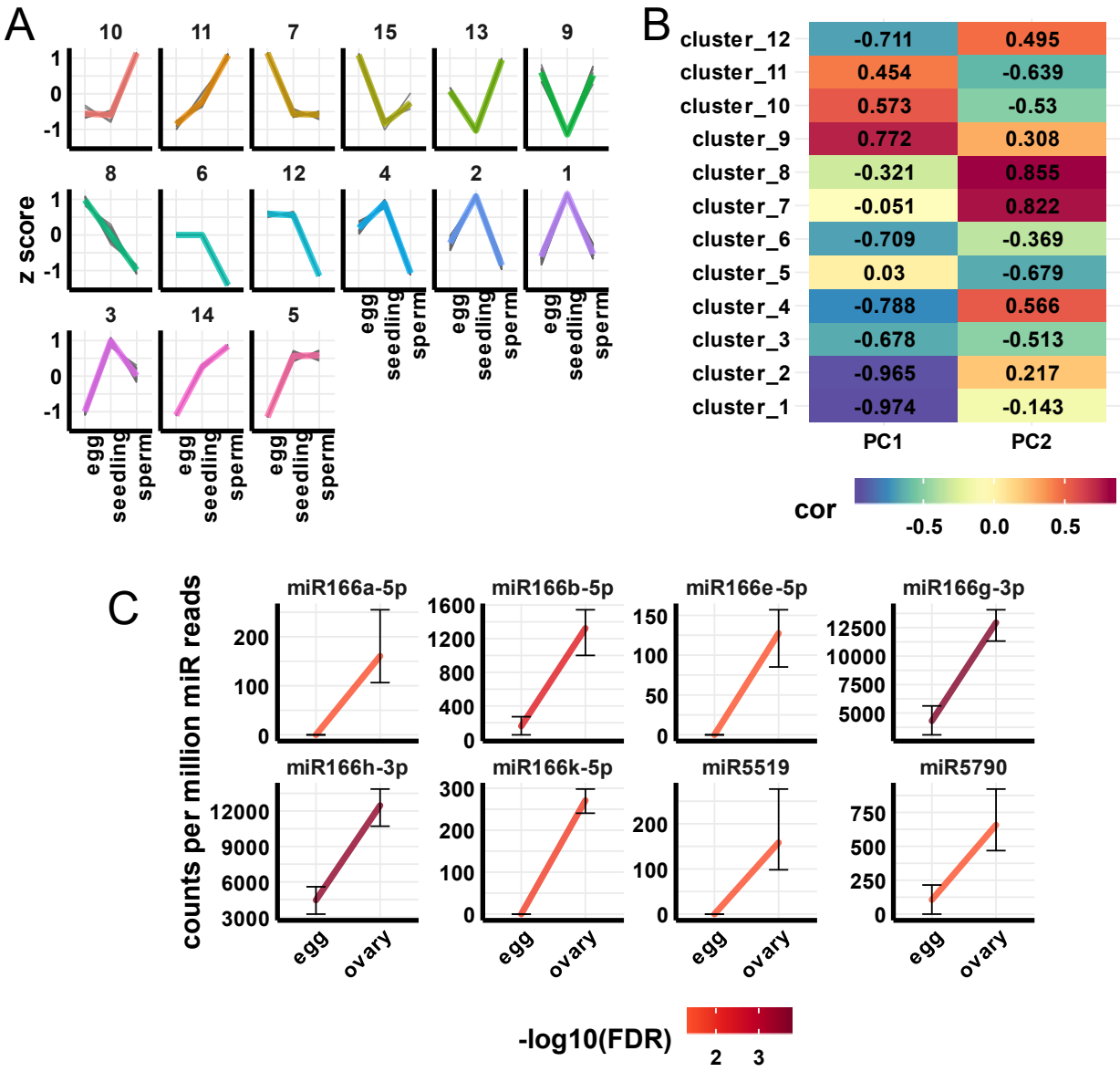
MethylC-seq and PBAT reads were quality filtered and trimmed of adapters using cutadapt (Martin 2011), parameters as follows: "-q 20 -a AGATCGGAAGAGC -e .1 -O 1 -m 50". PBAT reads were aligned to the genome with BS-Seeker2 (version 2.1.5) (Guo et al. 2013) with parameters as follows: "--aligner=bowtie2 --bt2--end-to-end -m 1 -t Y -s 5 -e 100"

methylC-seq reads were aligned similarly, but BS-Seeker2 parameters were modified to "-m 1 --aligner=bowtie2". For all reads except paired-end, previously published reads (Tan et al. 2016), PCR duplicates were removed prior to alignment with the BS-Seeker2 FilterReads.py module. DMRs were identified with CGmapTools version 0.1.1 (Guo et al. 2018). Biological replicates were first merged using the CGmapTools mergelist tosingle module, then the set of cytosines covered by reads in both samples were identified using the CGmapTools intersect module. Methylation comparisons were made using the CGmapTools dmr module, parameters "-c 3 -C 50 -s 100 -S 100 -n 5". DMRs were selected based on four criteria: at least five measured cytosines in the region; a P-value less than 0.001; absolute difference in methylation proportion of greater than 0.25 (wild-type value minus *drm2* mutant value), and a relative difference in methylation proportion of less than 0.3 (wild-type value divided by *drm2* mutant value).

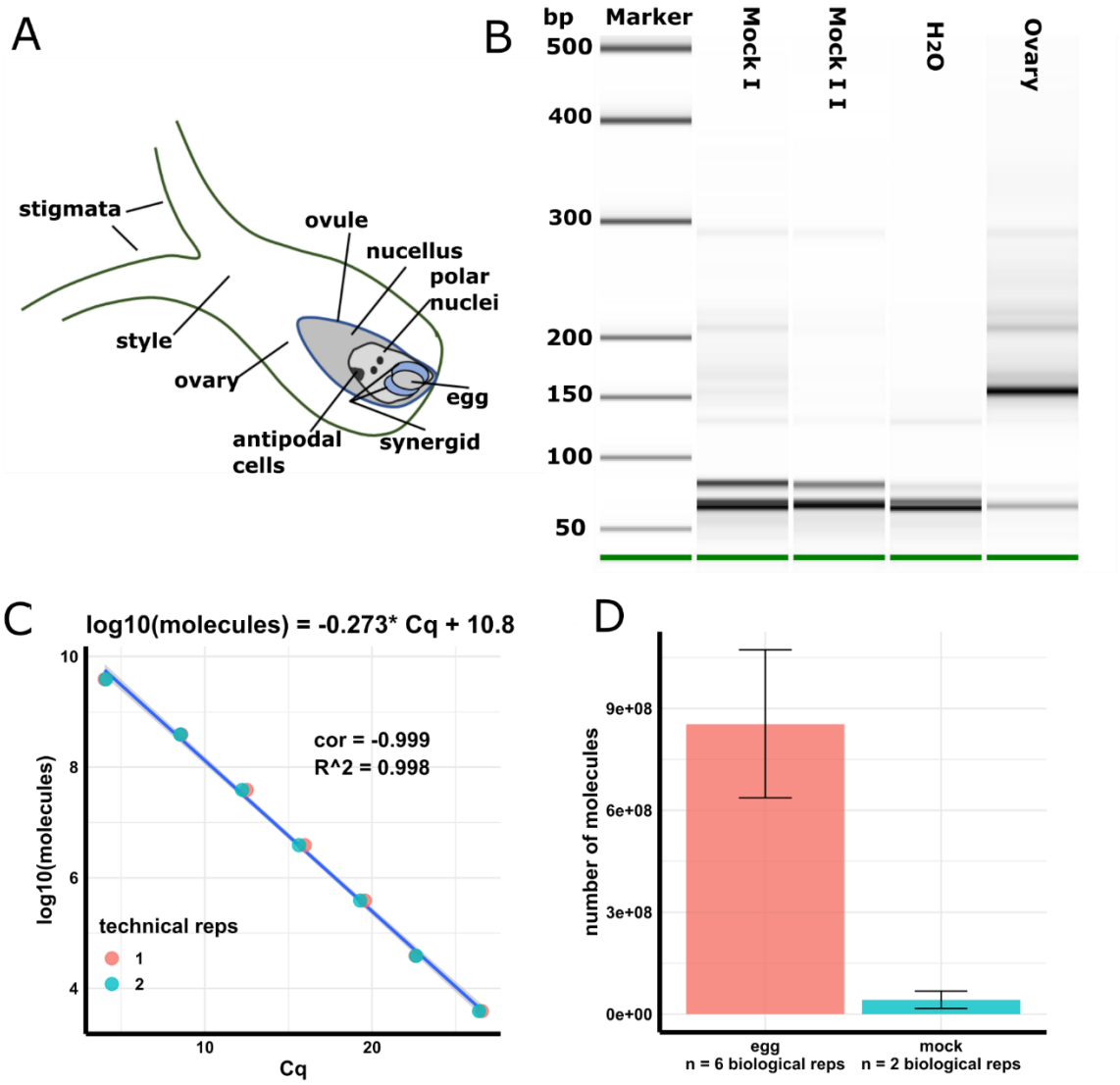
mRNA expression analysis

Previously published mRNA reads (Anderson et al. 2013; Anderson et al. 2017) were aligned to the genome using Tophat, version 2.0.13 (Kim et al. 2013), parameters as follows: "--read-realign-edit-dist=0 --min-intron-length=15 --max-intron-length=20000 --max-multihits=1 --microexon-search --library-type=fr-unstranded --b2-very-sensitive". For uniquely mapping reads only, "--prefilter-multihits" was also included. The number of reads that overlapped with genomic features was counted using BEDTools intersect, requiring that half of each read's mapped length overlapped with a feature using the "-f .5" parameter to be counted.

All R-script for statistical analyses and data visualization can be found at https://github.com/cxli233/gamete_smRNA_revision



Supplemental Figure 1: Differential miRNA expression and clustering
A: Clusters of miRNAs based on expression in egg cell, sperm cell, and seedling shoot. miRNAs were clustered by hierarchical clustering of z-scores. The z-score of each miRNA was calculated relative to the mean and standard deviation of each miRNA across tissues. Colored lines are the average of each cluster. Grey lines are individual miRNAs. See also **Supplemental Table S3**.
B: "Cor" indicates correlation of miRNA clusters to each principal component (PC) axis (see also **Fig 1A**). Color and text reflect Pearson's correlation coefficient.
C: Representative differentially expressed miRNAs between egg cell and ovary. Differential expression was determined by 2-fold change and FDR (false discovery rate) < 0.05 cutoffs. Y-values are relative to the total number of miRNA reads in each sample. Error bars are 95% confidence intervals. Color of line reflects -log10 of FDR.



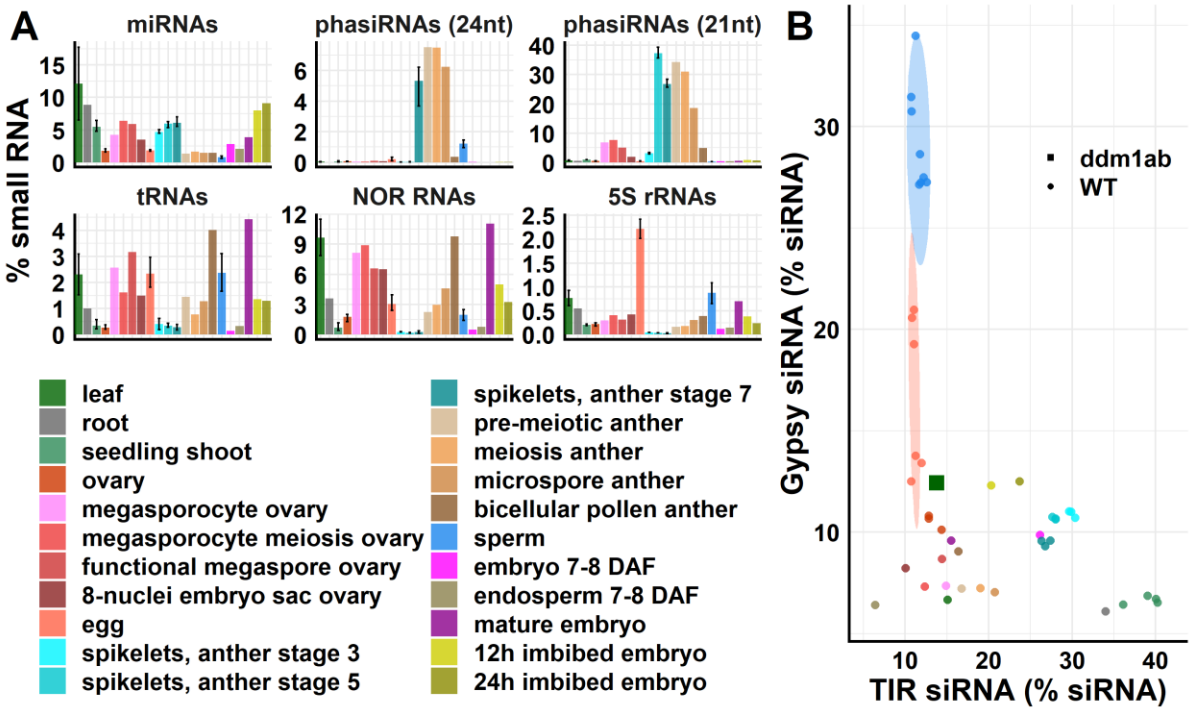
Supplemental Figure 2: Mock egg cell isolations and qPCR quantification of small RNA libraries

A: Scheme of rice ovary, redrawn from Li et al (2019).

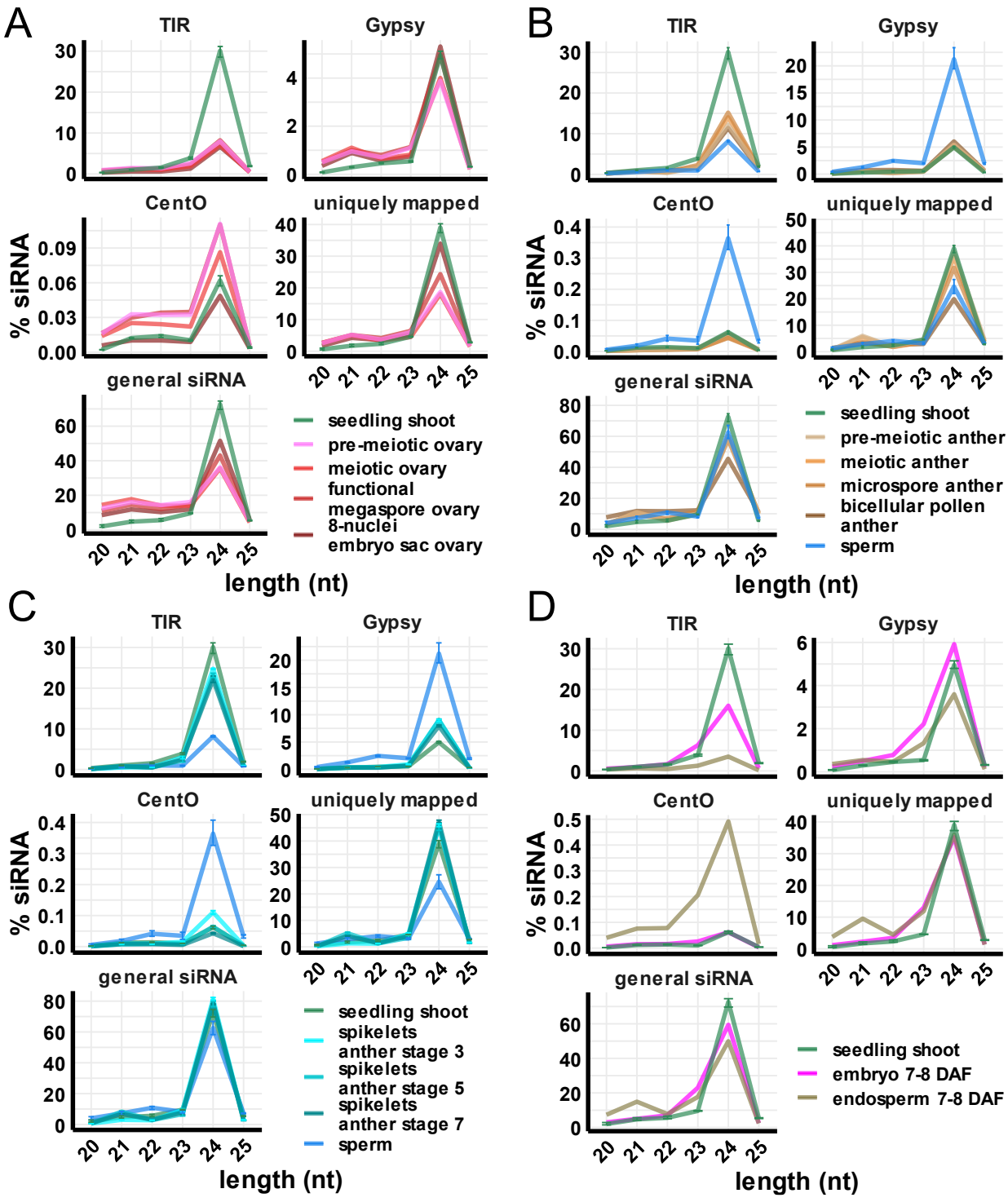
B: Bioanalyzer gel image of mock libraries. Expected library product size is about 150 bp.

C: Standard curve for qPCR quantification.

D: Number of molecules in PCR-amplified egg cell libraries and mock libraries.

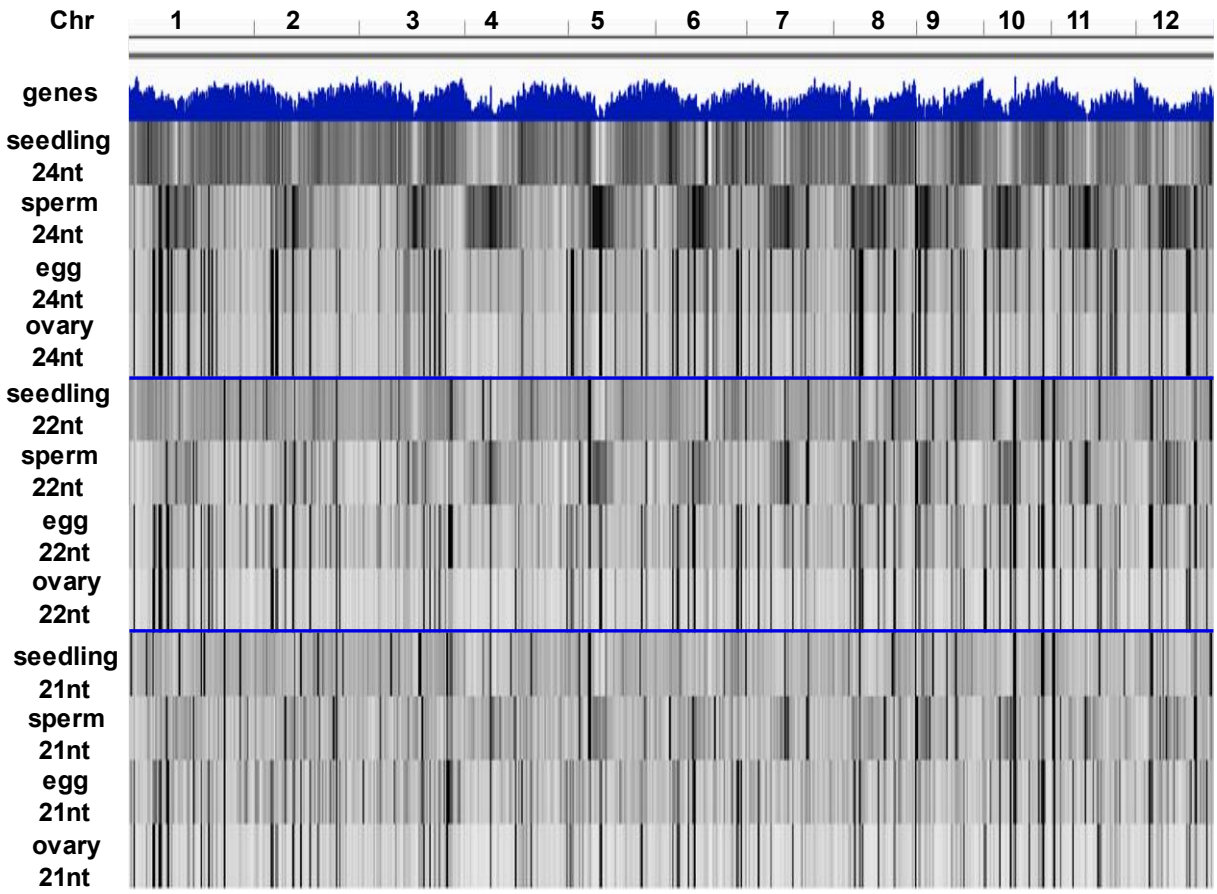


Supplemental Figure 3: Small RNA compositions across tissues
A: small RNA compositions across sample types, as in **Fig 1C**. Y-axis values are relative to the total number of reads that mapped to the genome. For samples with two or more biological replicates, error bars are 95% confidence intervals.
B: Scatter plot showing TIR siRNAs on X-axis and Gypsy siRNAs on Y-axis, each measured as percent of the total siRNAs in each library. Egg and sperm cell replicates formed unique clusters, highlighted by pink and blue eclipse, respectively. Color code as in **A**. Published data sources: leaf (Tan et al. 2018), root (Shin et al. 2018), ovary and anther (Li et al. 2017); spikelet (Fei et al. 2016), 7-8 DAF embryo and endosperm (Rodrigues et al. 2013), and mature and imbibed embryo (He et al. 2015).

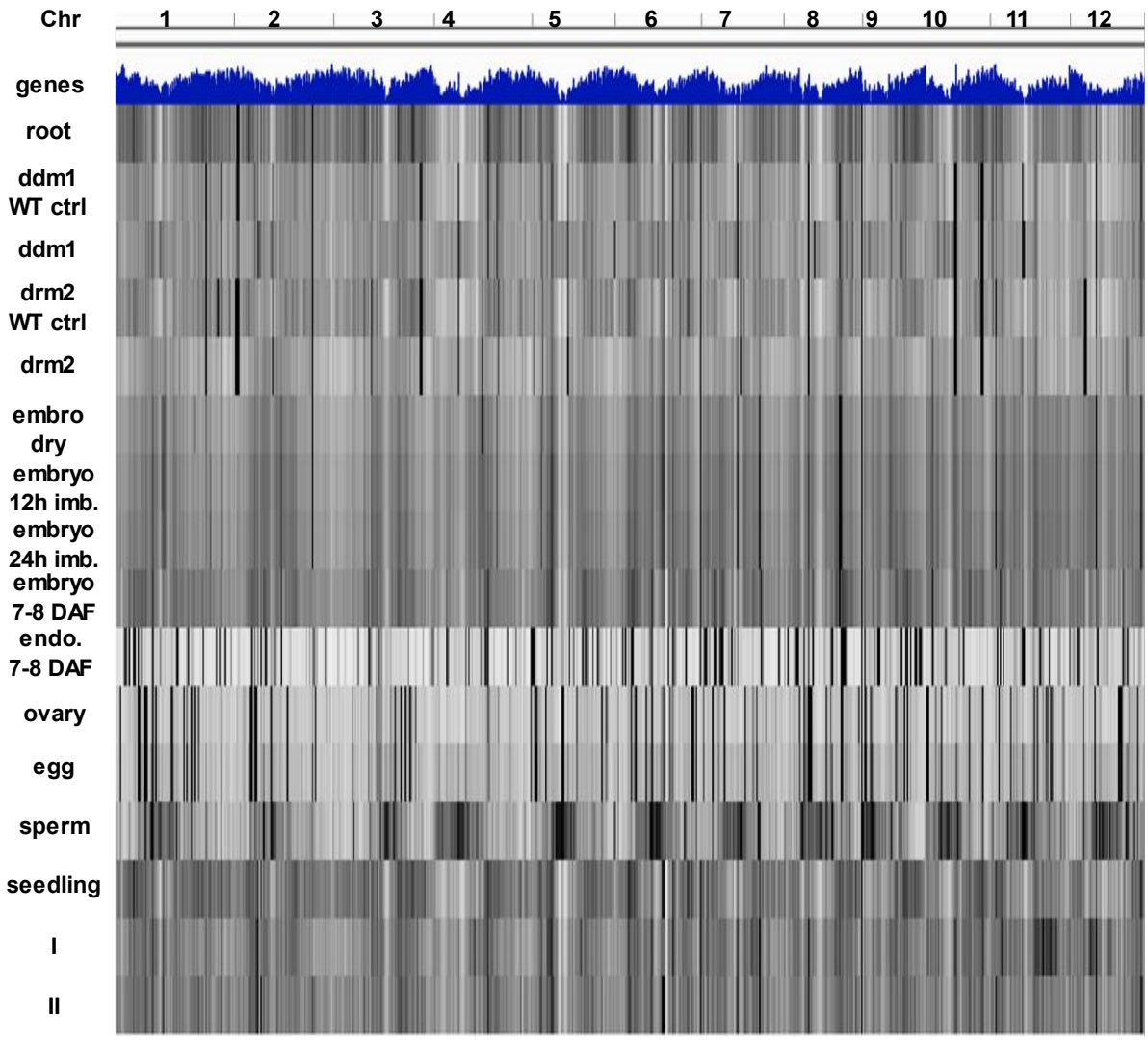


Supplemental Figure 4: siRNA abundance by length and category
Y-axis values are the number of siRNAs for each length normalized by the total number of siRNAs. Error bars are 95% confidence intervals.
A: Ovary developmental series (Li et al. 2017).
B: Anther developmental series (Li et al. 2017).
C: Spikelet developmental series (Fei et al. 2016).

D: Developing embryo and endosperm, 7 – 8 DAF (Rodrigues et al. 2013).



Supplemental Figure 5: Whole-genome heatmaps of 21-nt, 22-nt, and 24-nt siRNAs
Top track indicates gene density, all others siRNAs.

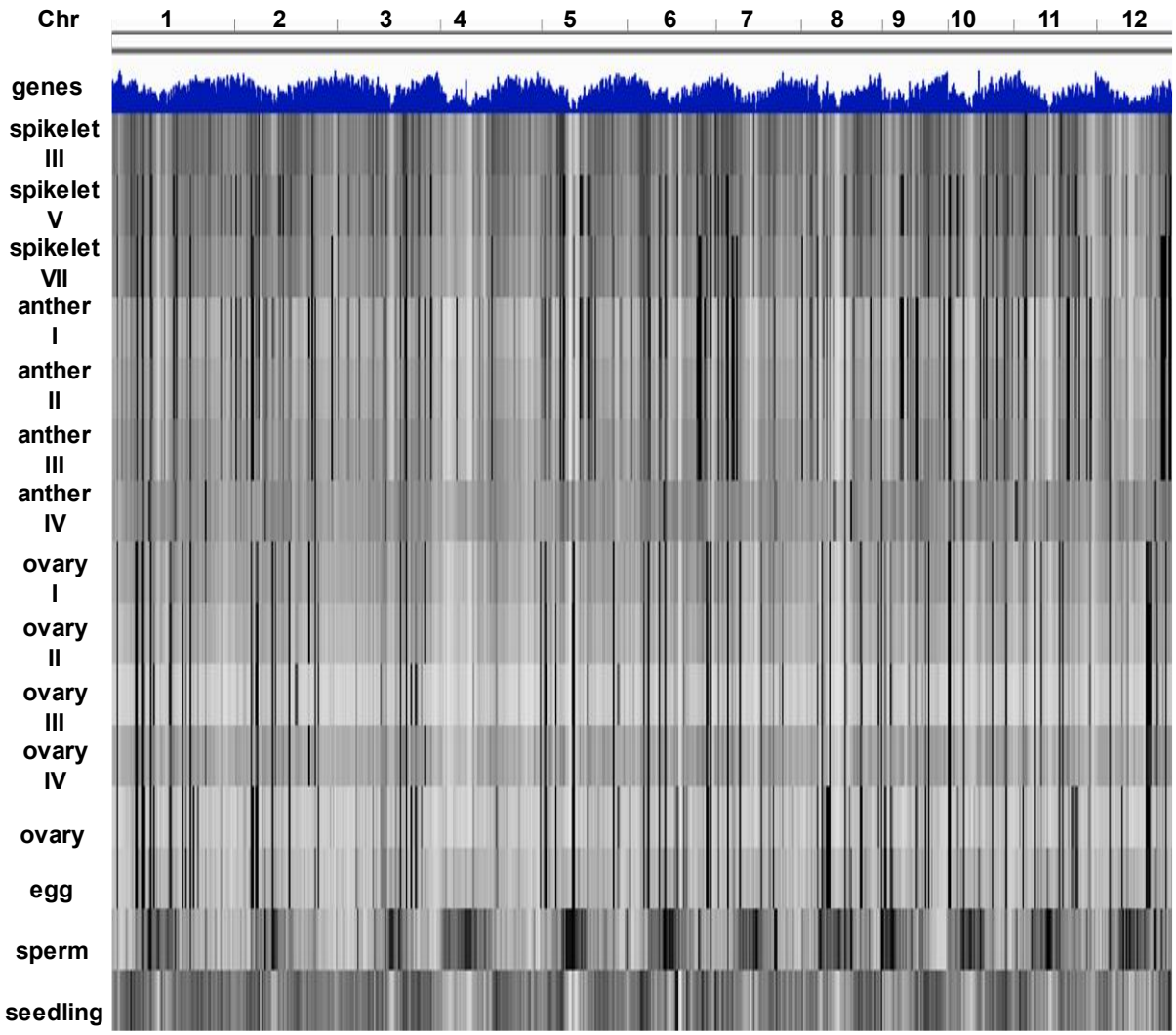


Supplemental Figure 6: Whole-genome heatmaps of vegetative tissues, embryo and endosperm 24-nt siRNAs

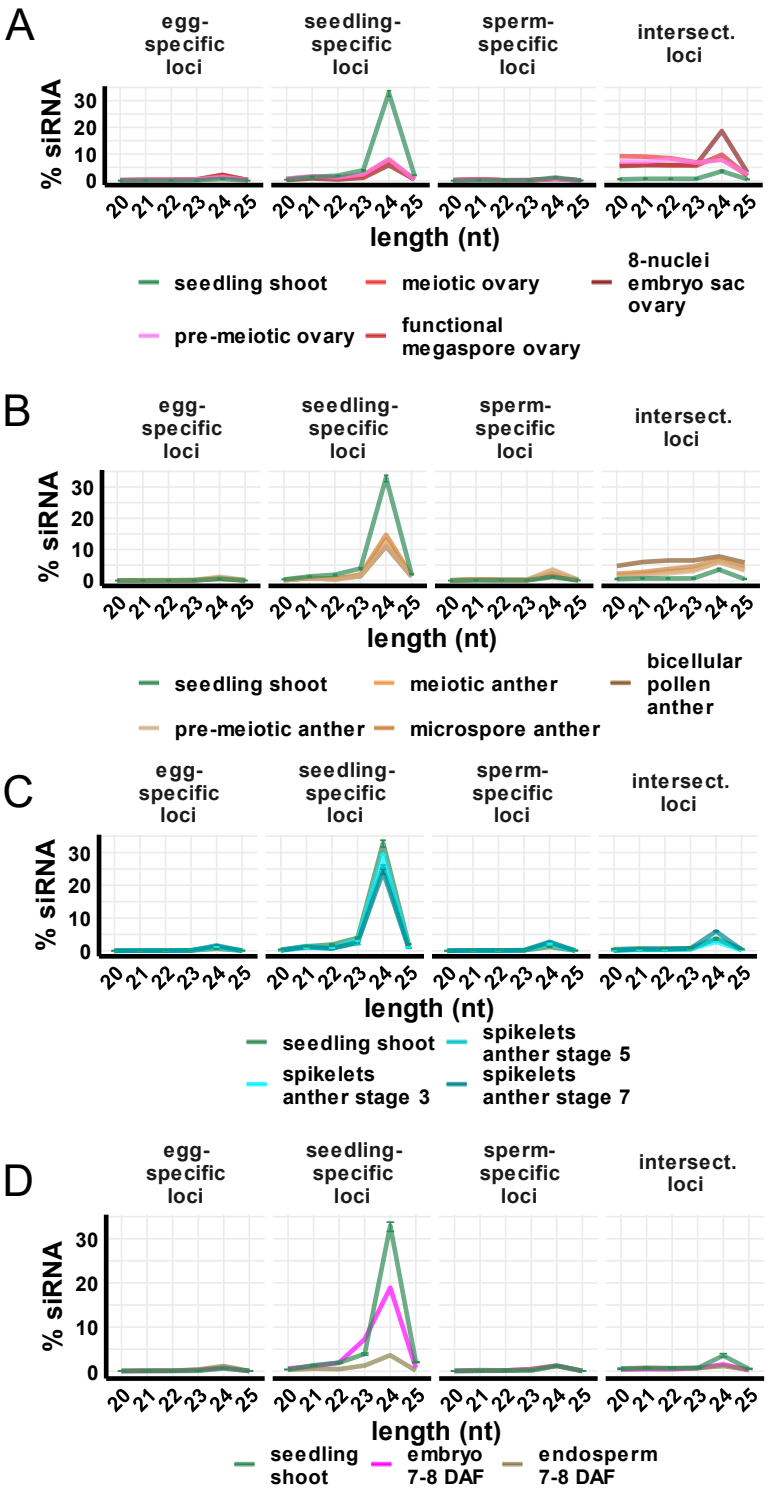
Top track indicates gene density, all others siRNAs. Published data sources: *drm2* and wildtype leaf (Tan et al. 2016), *ddm1* and wildtype leaf (Tan et al. 2018), root (Shin et al. 2018), 7-8 DAF embryo and endosperm (Rodrigues et al. 2013), and mature and imbibed embryo (He et al. 2015).

I: 24-nt siRNAs from seedling shoot after removing all siRNAs that overlapped with seedling-specific 24-nt siRNA loci.

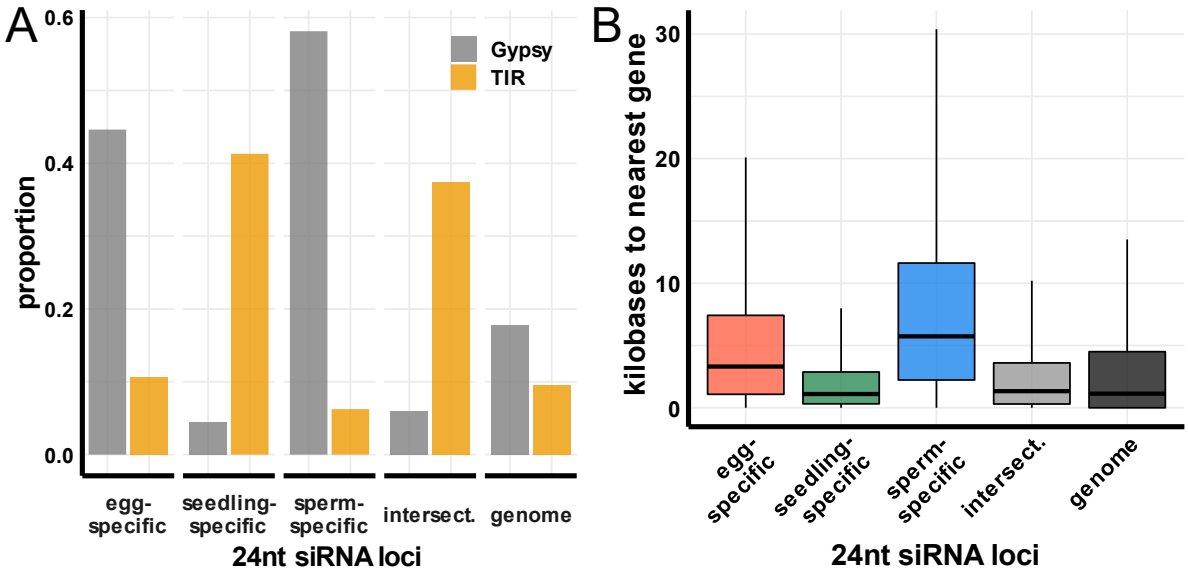
II: 24-nt siRNAs from seedling shoot after removing all siRNAs that overlapped with mCHH loci, the set of 100-bp loci in the genome with mCHH > 5% from leaf (Tan et al. 2016).



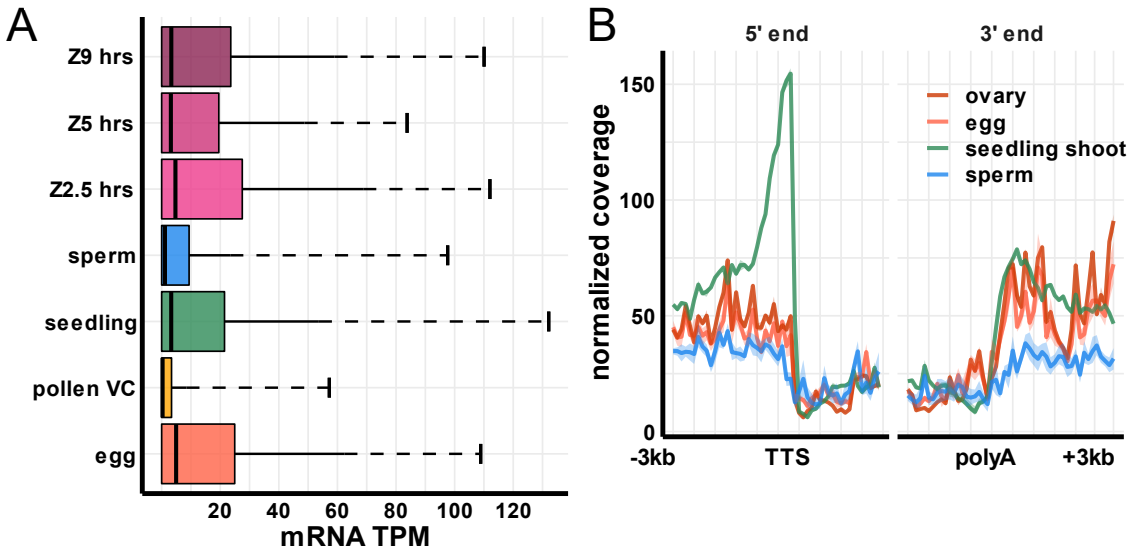
Supplemental Figure 7: Whole-genome heatmaps of reproductive tissue 24-nt siRNAs
Top track indicates gene density, all others siRNAs. Spikelet III, IV and VII: spikelet of anther developmental stages 3, 5 and 7, respectively (Fei et al. 2016); Anther I, II, III and IV: pre-meiotic anther; meiotic anther; microspore anther and bicellular pollen anther, respectively (Li et al. 2017);
Ovary I, II, III, IV: pre-meiotic ovary; meiotic ovary; functional megaspore ovary and 8-nuclei embryo sac ovary, respectively (Li et al. 2017).



Supplemental Figure 8: siRNA abundance at 24-nt siRNA loci.
Y-values are the percent of siRNAs relative to total siRNAs. Error bars are 95% confidence intervals for sample types with more than one replicates, as in **Fig. 1D**. Data sources as indicated in **Fig. S4**.
A: Ovary developmental series.
B: Anther developmental series.
C: Spikelet developmental series.
D: Developing embryo and endosperm, 7 – 8 DAF.

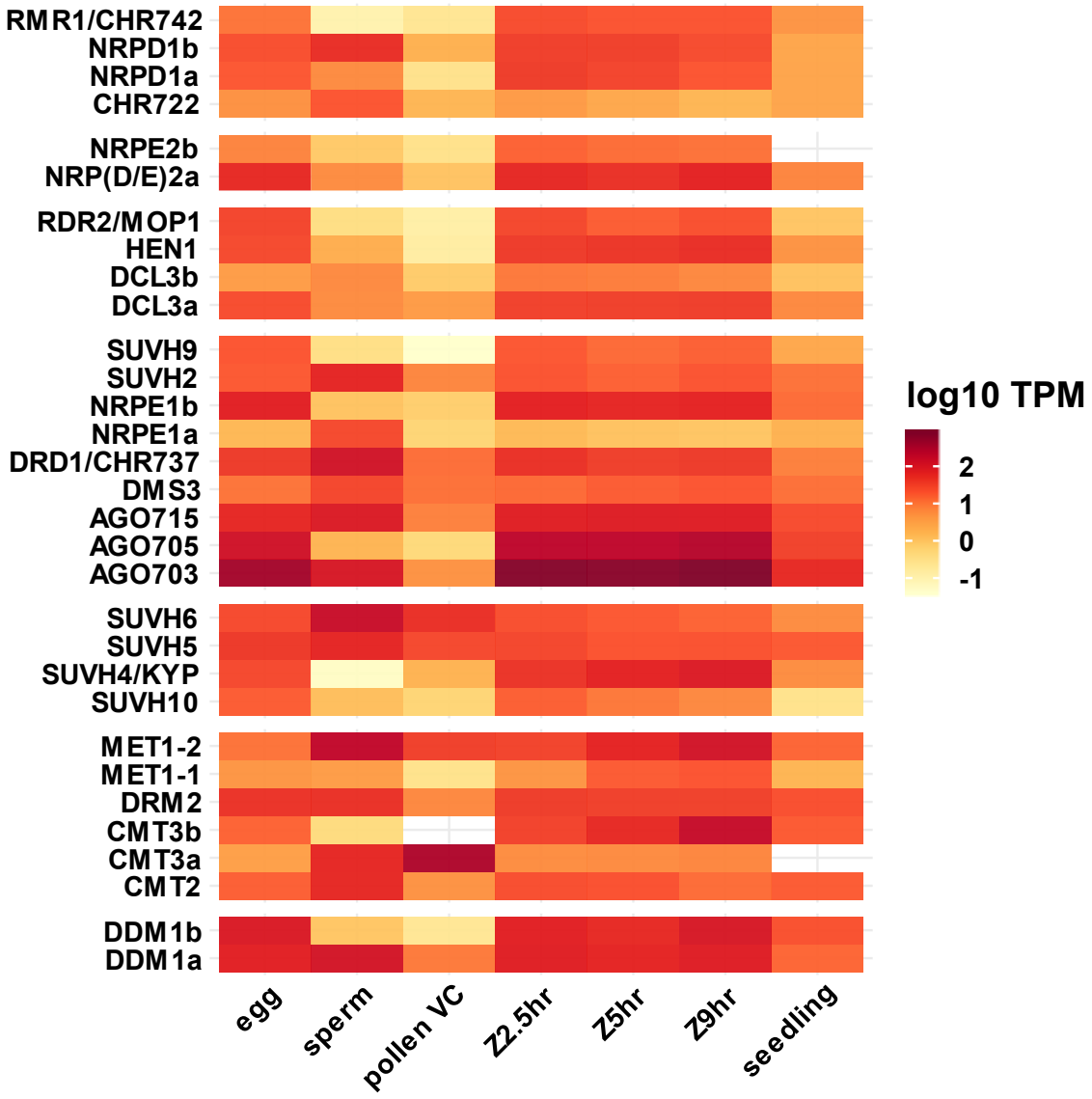


Supplemental Figure 9: Overlaps of sample-specific siRNA loci with repeats and distances to nearest genes
A: Proportion of sample-specific or intersection 24-nt siRNA loci overlapping a *Gypsy* or TIR transposon.
B: Distance in kilobases from 24-nt siRNA loci to the nearest gene. Boxes are interquartile range (IQR), upper whisker extends to upper quartile + 2 * IQR, lower whisker to zero.

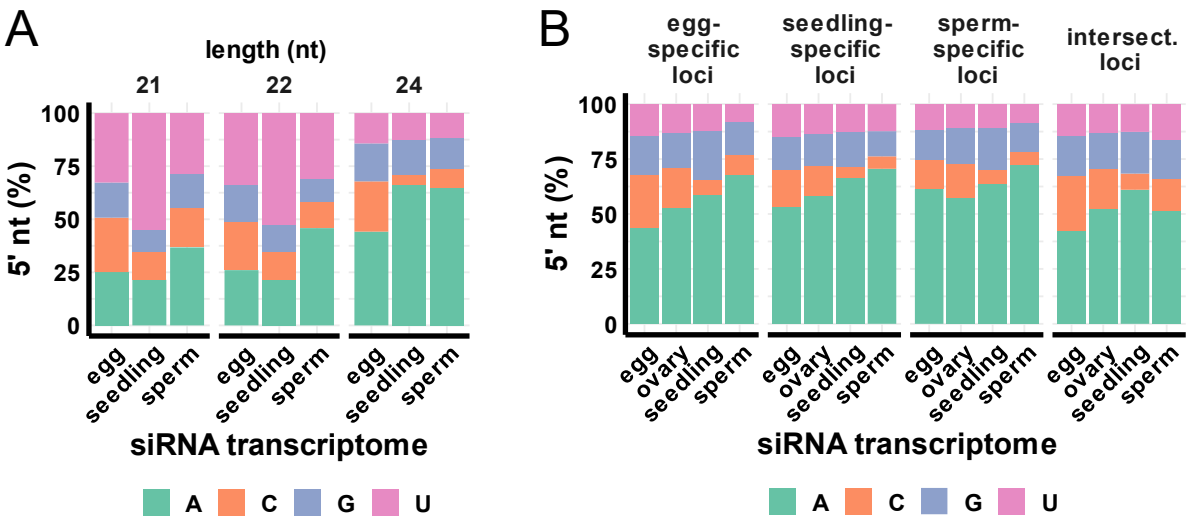


Supplemental Figure 10: Depletion of flanking 24-nt siRNAs for highly expressed sperm genes (>10 TPM)
A: Boxplot showing distribution of mRNA expression. Boxes are interquartile range (IQR), and whiskers extend to upper quartile + 1.5 * IQR. Dotted lines extend to 95th percentile. A TPM (transcripts per million) of 10 approximately corresponds to upper quartile in sperm.

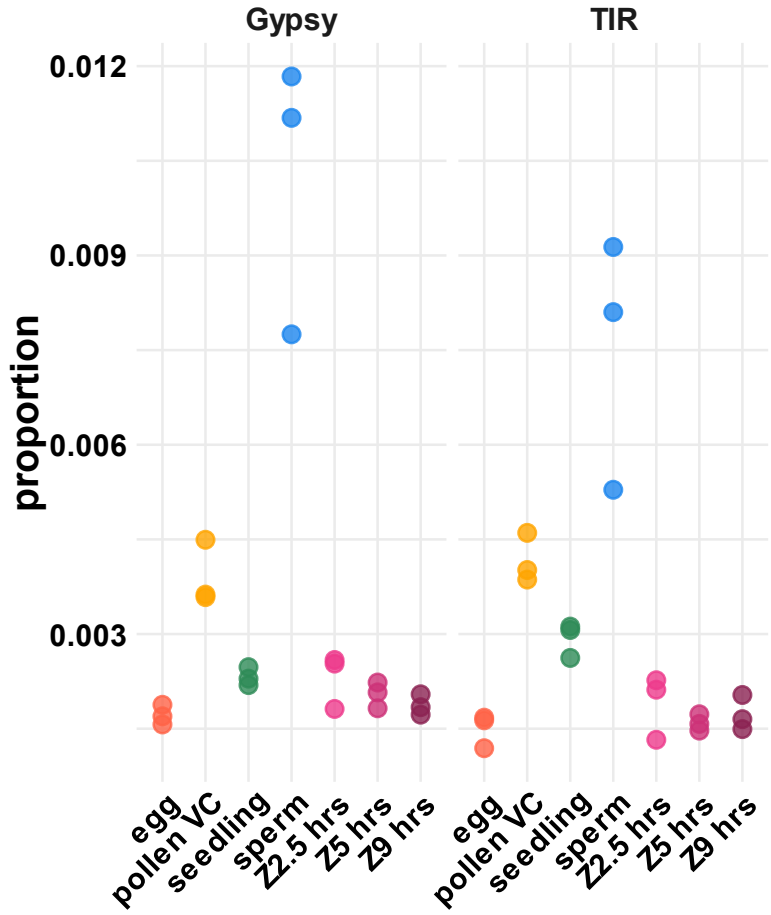
B: Metagene plot for 24-nt siRNAs for genes expressing at >10 TPM in sperm, as in **Fig 2B**. Plots indicate 24-nt siRNA coverage with 100-bp resolution from 3-KB upstream to 3-KB downstream of genes, normalized per 1000 total siRNAs. Tick marks indicate 500-bp intervals. TSS: Transcription start site; poly A: polyadenylation signal.



Supplemental Figure 11: Expression of RdDM and methylation related factors in gametes
Pollen VC: pollen vegetative cell; Z: zygote, 2.5hr, 5hr, and 9hr, at completion of karyogamy, nucleoli fusion and S-phase, respectively. Colors reflect transcripts per million values in log10 scale. mRNA data source: (Anderson et al. 2013; Anderson et al. 2017).

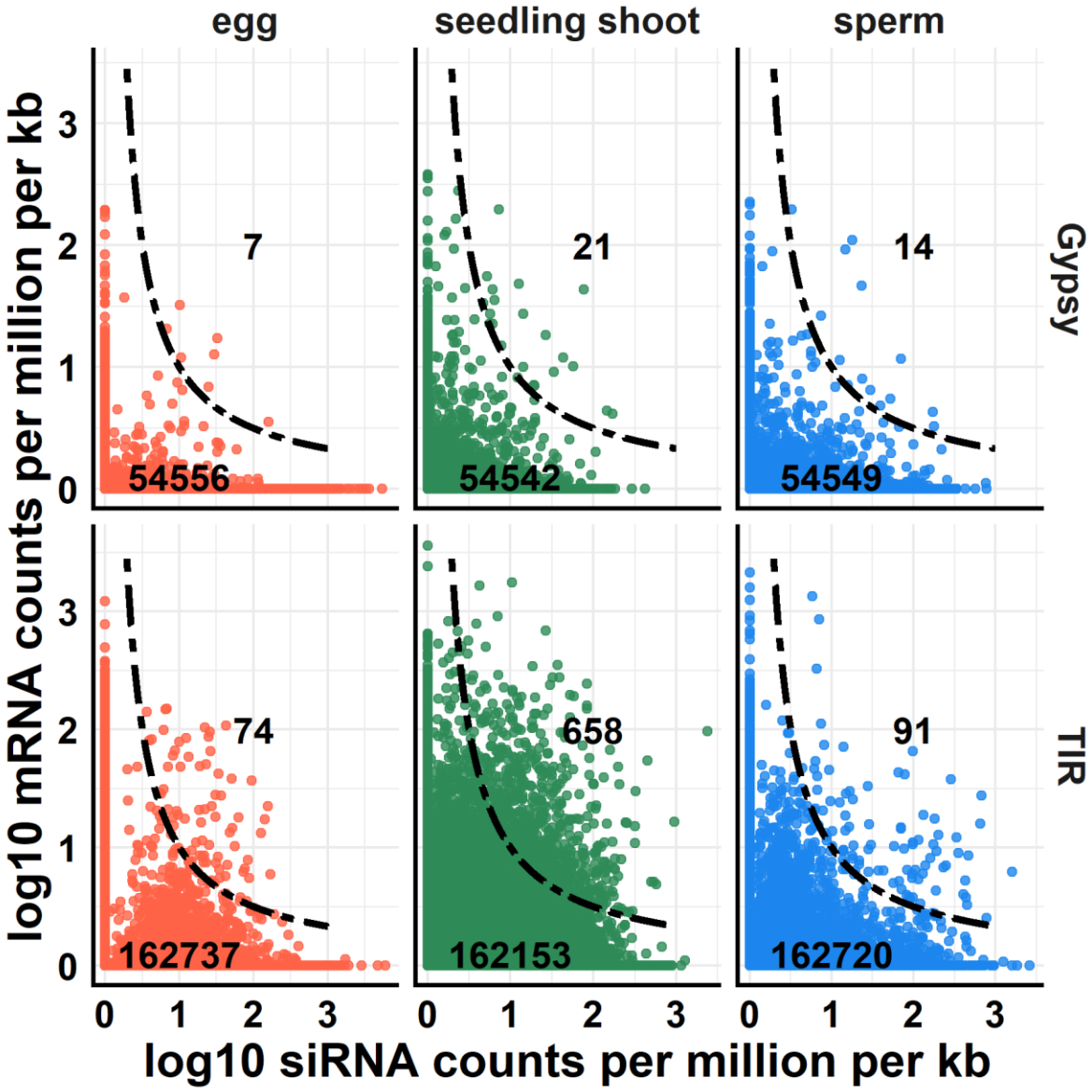


Supplemental Figure 12: siRNA 5' nucleotide preferences
A: The stacked bar charts show the percent of 24-nt siRNAs from the tissue-specific 24-nt siRNA loci categories that begin with each nucleotide. **B:** 5' nucleotide preference of 24-nt siRNAs across different siRNA loci.



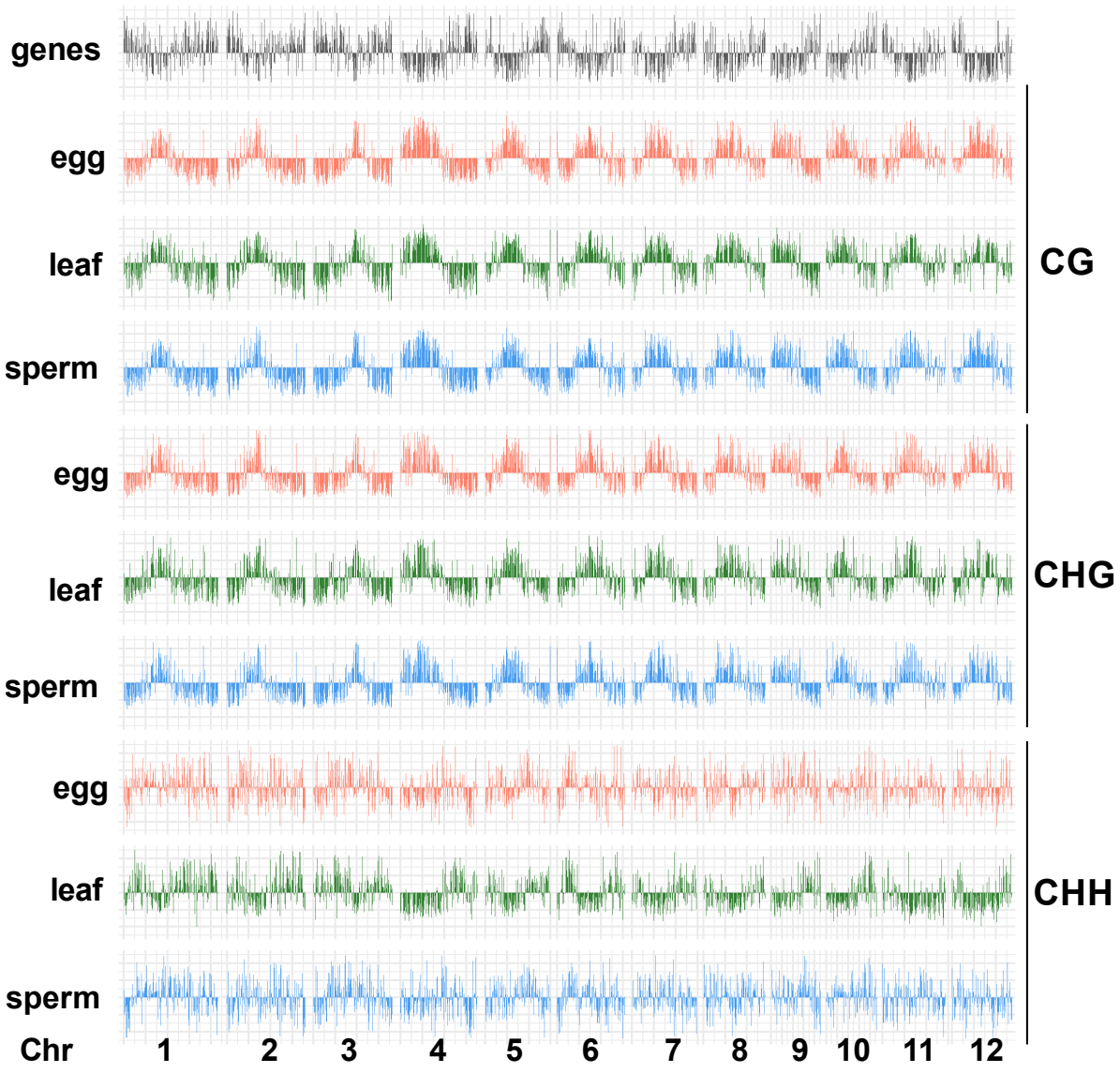
Supplemental Figure 13: Proportion of mRNA reads mapping to transposons in gametes
All reads were included in this analysis, uniquely and multi-mapping. Read counts were normalized by total number of mapping reads per library. Pollen VC: pollen vegetative cell; Z:

zygote, 2.5hr, 5hr, and 9hr, at completion of karyogamy, nucleoli fusion and S-phase, respectively. mRNA data source: (Anderson et al. 2013; Anderson et al. 2017).

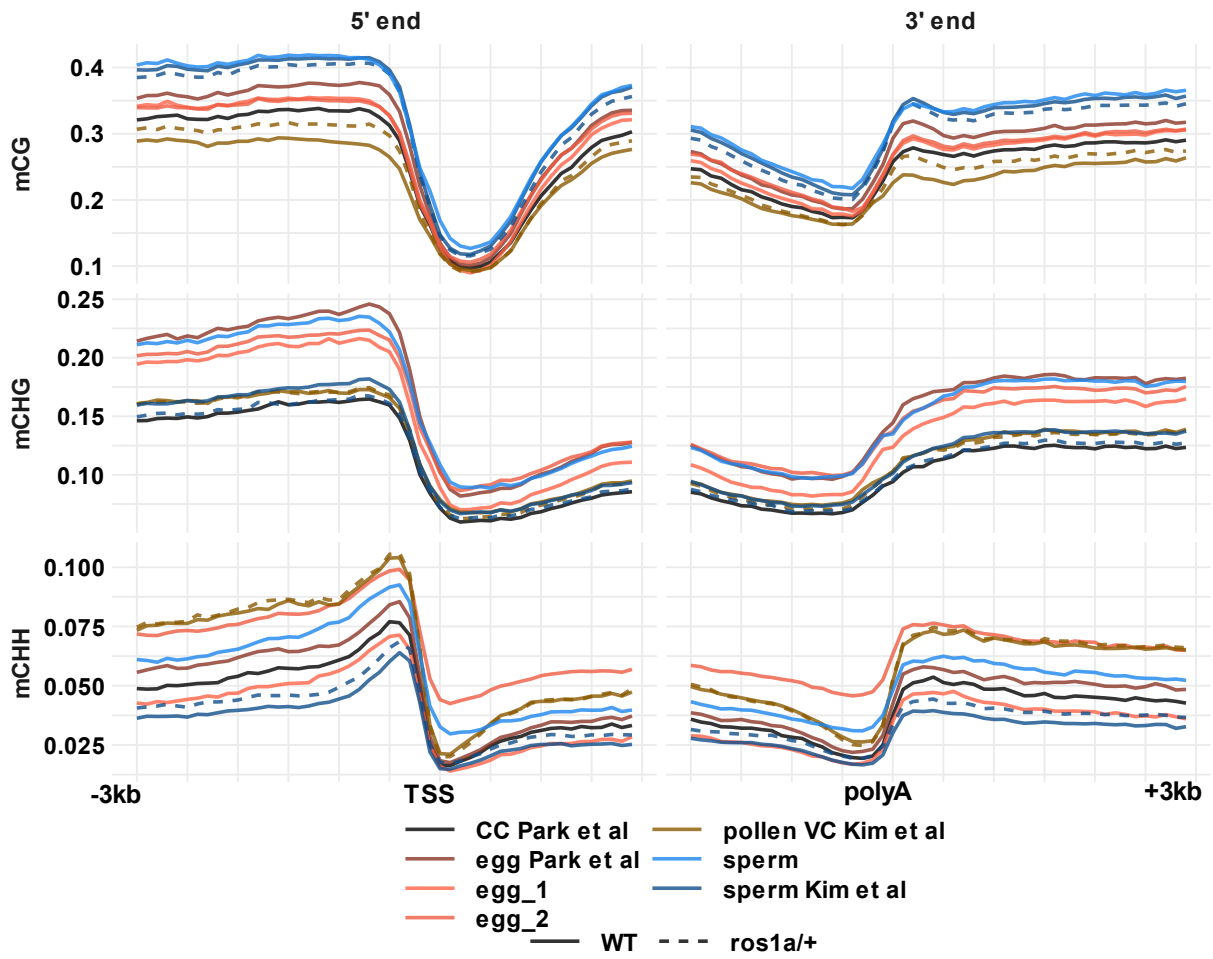


Supplemental Figure 14: mRNA read counts vs. 24-nt siRNA read counts for individual transposon copies

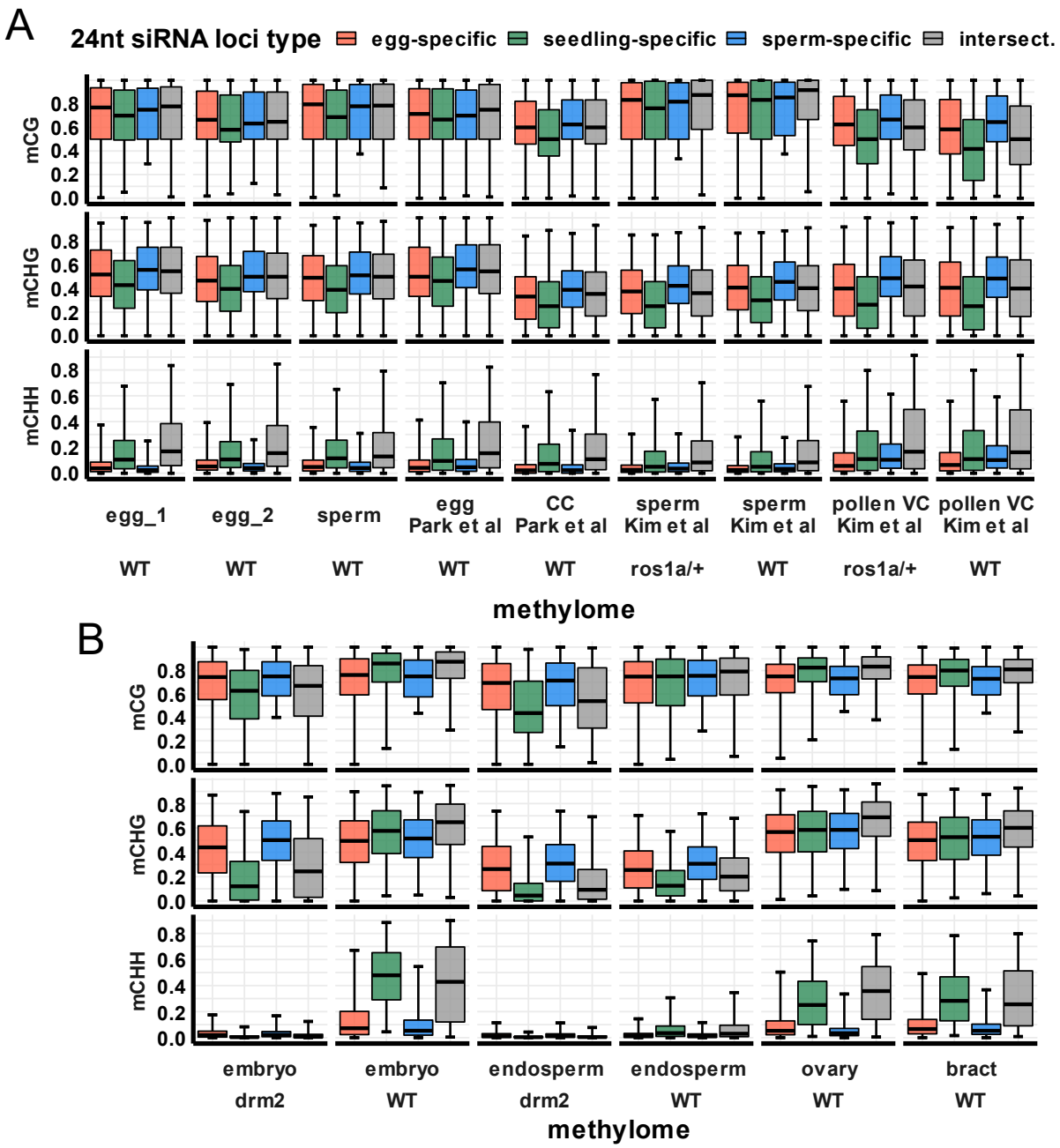
Only uniquely-mapping reads were included in this analysis. mRNA counts were normalized by total number of mapping reads and siRNA reads were normalized per million 24-nt siRNAs. Dotted line plots y = 1/x relationship. Numbers above and below the curve show number of transposon copies outside or inside the y = 1/x relationship, respectively. mRNA data source: (Anderson et al. 2013; Anderson et al. 2017).



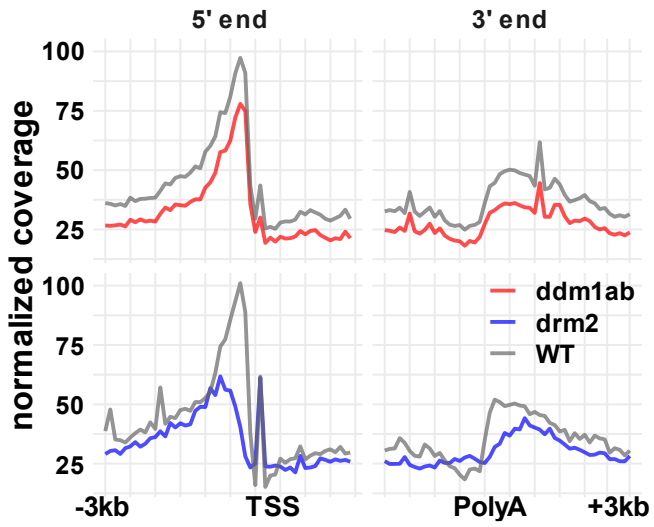
Supplemental Figure 15: Genome-wide view of DNA methylation
Y-axis values are z-scores of average DNA methylation values on 50 kilobase intervals. Z-scores were calculated from mean and standard deviation of each sample type and each context. Leaf data source: (Tan et al. 2016).



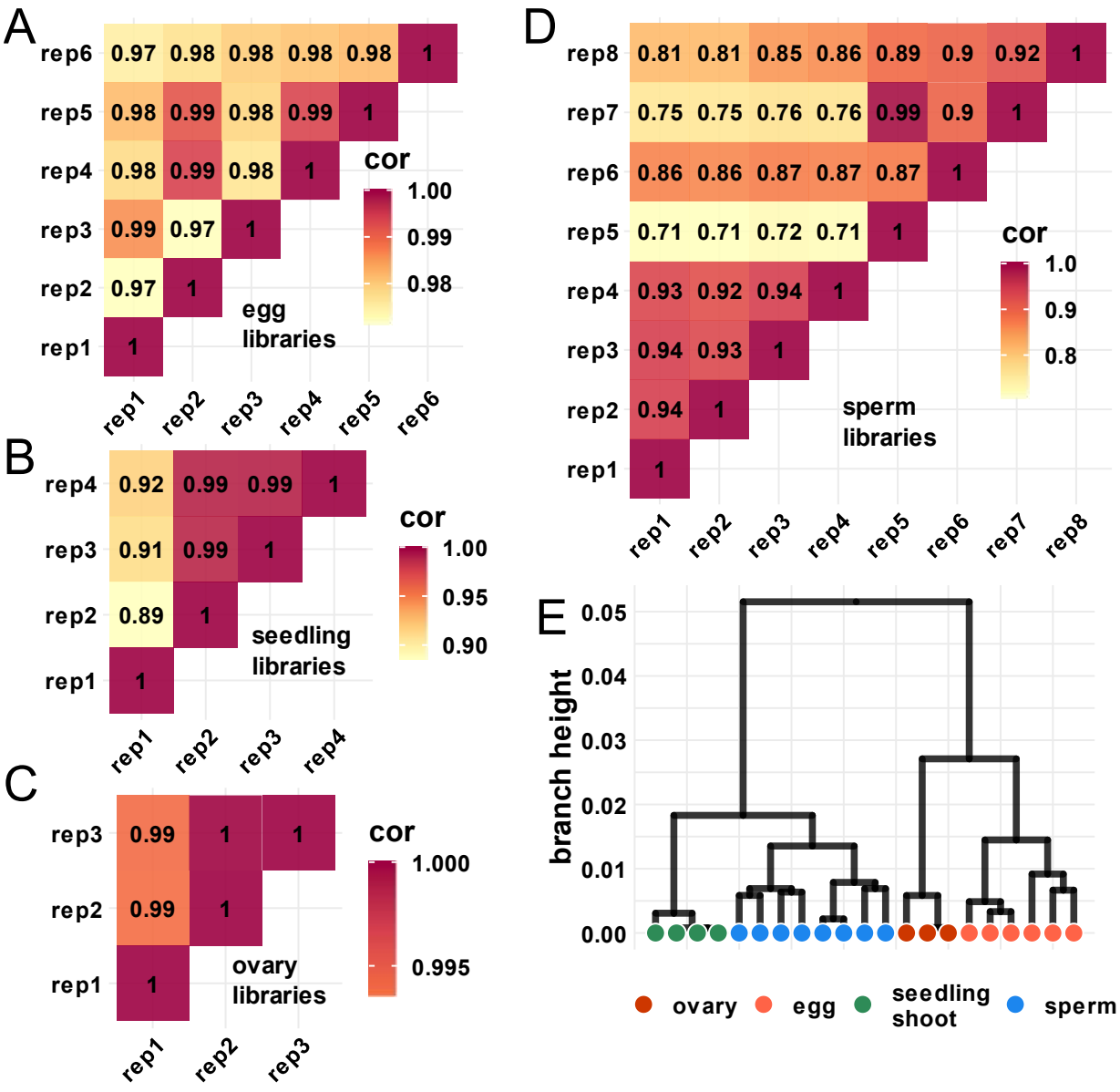
Supplemental Figure 16: Methylation metaplots of all PBAT libraries analyzed
Plots indicate average DNA methylation values over 100-bp intervals from 3-KB upstream to 3-KB downstream of genes. Tick marks indicate 500-bp intervals. DNA methylation is measured as the proportion methylated cytosines relative to total cytosines in each sequence context. TSS: Transcription start site; polyA: polyadenylation signal. CC, central cell (Park et al. 2016); egg cell (Park et al. 2016); sperm (Kim et al. 2019); and pollen VC, vegetative cell (Kim et al. 2019).



Supplemental Figure 17: DNA methylation of 24-nt siRNA loci
A: PBAT libraries. Center line is median methylation; box represents interquartile range; whiskers extend from 2.5th to 97.5th percentile.
B: Conventional libraries as in **A**. Bract: lemma and palea of rice florets. CC, central cell (Park et al. 2016); egg cell (Park et al. 2016); sperm (Kim et al. 2019); and pollen VC, vegetative cell (Kim et al. 2019).



Supplemental Figure 18: Metagene plot for 24-nt siRNAs in *ddm1* and *drm2* mutants Plots indicate 24-nt siRNA coverage with 100-bp resolution from 3-KB upstream to 3-KB downstream of genes, normalized per 1000 total siRNAs. Tick marks indicate 500-bp intervals. TSS: Transcription start site; polyA: polyadenylation signal. Data sources: (Tan et al. 2016; Tan et al. 2018).



Supplemental Figure 19: Correlation and clustering of 24-nt siRNA libraries
A: Pairwise Pearson's correlation coefficient between six egg cell libraries.
B: Pairwise Pearson's correlation coefficient between four seedling shoot libraries.
C: Pairwise Pearson's correlation coefficient between three ovary libraries.
D: Pairwise Pearson's correlation coefficient between eight sperm cell libraries.
E: Clustering of small RNA libraries based on hierarchical clustering.

Supplemental References

Anderson SN, Johnson CS, Chesnut J, Jones DS, Khanday I, Woodhouse M, Li C, Conrad LJ, Russell SD, Sundaresan V. 2017. The Zygotic Transition Is Initiated in Unicellular Plant Zygotes with Asymmetric Activation of Parental Genomes. *Dev Cell* **43**: 349-358.e344.

Anderson SN, Johnson CS, Jones DS, Conrad LJ, Gou X, Russell SD, 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.

Chan PP, Lowe TM. 2016. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* **44**: D184-189.

Fei Q, Yang L, Liang W, Zhang D, Meyers BC. 2016. Dynamic changes of small RNAs in rice spikelet development reveal specialized reproductive phasiRNA pathways. *J Exp Bot* **67**: 6037-6049.

Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, Chen PY, Pellegrini M. 2013. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* **14**: 774.

Guo W, Zhu P, Pellegrini M, Zhang MQ, 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.

He D, Wang Q, Wang K, Yang P. 2015. Genome-Wide Dissection of the MicroRNA Expression Profile in Rice Embryo during Early Stages of Seed Germination. *PLoS One* **10**: e0145424.

Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S et al. 2013. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice (N Y)* **6**: 4.

Khanday I, Skinner D, Yang B, Mercier R, Sundaresan V. 2019. A male-expressed rice embryogenic trigger redirected for asexual propagation through seeds. *Nature* **565**: 91-95.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36.

Kim MY, Ono A, Scholten S, Kinoshita T, Zilberman D, Okamoto T, Fischer RL. 2019. DNA demethylation by ROS1a in rice vegetative cells promotes methylation in sperm. *Proc Natl Acad Sci U S A* doi:10.1073/pnas.1821435116.

Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to function. *Nucleic Acids Res* **47**: D155-D162.

Langfelder P, Zhang B, Horvath S. 2008. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**: 719-720.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Li C, Xu H, Russell SD, Sundaresan V. 2019. Step-by-step protocols for rice gamete isolation. *Plant Reprod* **32**: 5-13.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Li X, Shahid MQ, Xia J, Lu Z, Fang N, Wang L, Wu J, Chen Z, Liu X. 2017. Analysis of small RNAs revealed differential expressions during pollen and embryo sac development in autotetraploid rice. *BMC Genomics* **18**: 129.

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

McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**: 4288-4297.

Moritoh S, Eun CH, 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. *Plant J* **71**: 85-98.

Park K, Kim MY, Vickers M, Park JS, Hyun Y, Okamoto T, Zilberman D, Fischer RL, Feng X, Choi Y et al. 2016. DNA demethylation is initiated in the central cells of Arabidopsis and rice. *Proc Natl Acad Sci U S A* **113**: 15138-15143.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**: R25.

Rodrigues JA, Ruan R, Nishimura T, Sharma MK, Sharma R, Ronald PC, Fischer RL, 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 U S A* **110**: 7934-7939.

Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**: 863-864.

Shin SY, Jeong JS, Lim JY, Kim T, Park JH, Kim JK, Shin C. 2018. Transcriptomic analyses of rice (Oryza sativa) genes and non-coding RNAs under nitrogen starvation using multiple omics technologies. *BMC Genomics* **19**: 532.

Tan F, Lu Y, Jiang W, Wu T, Zhang R, Zhao Y, Zhou DX. 2018. DDM1 Represses Noncoding RNA Expression and RNA-Directed DNA Methylation in Heterochromatin. *Plant Physiol* **177**: 1187-1197.

Tan F, Zhou C, Zhou Q, Zhou S, Yang W, Zhao Y, Li G, Zhou DX. 2016. Analysis of Chromatin Regulators Reveals Specific Features of Rice DNA Methylation Pathways. *Plant Physiol* **171**: 2041-2054.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178-192.

Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR. 2015. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat Protoc* **10**: 475-483.

Xie K, Zhang J, Yang Y. 2014. Genome-wide prediction of highly specific guide RNA spacers for CRISPR-Cas9-mediated genome editing in model plants and major crops. *Mol Plant* **7**: 923-926.

Zhang T, Talbert PB, Zhang W, Wu Y, Yang Z, Henikoff JG, Henikoff S, Jiang J. 2013. The CentO satellite confers translational and rotational phasing on cenH3 nucleosomes in rice centromeres. *Proc Natl Acad Sci U S A* **110**: E4875-4883.

Chapter 4

Resetting of the 24-nt siRNA landscape in rice zygotes

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

4.1 Abstract

The zygote, a totipotent stem cell, is crucial to the life cycle of sexually reproducing organisms. It is produced by the fusion of two differentiated cells — the egg and sperm, which in plants have radically different siRNA transcriptomes from each other, and from multicellular embryos. Due to technical challenges, the epigenetic changes that accompany the transition from differentiated gametes to totipotent zygote are poorly understood. Since siRNAs serve as both regulators and outputs of the epigenome, we performed here the successful characterization of small RNA transcriptomes of zygotes from rice. Zygote small RNAs exhibited extensive maternal carryover and an apparent lack of paternal contribution, indicated by absence of sperm signature siRNAs. Zygote formation was accompanied by widespread redistribution of 24-nt siRNAs relative to gametes, such that ~70% of the zygote siRNA loci did not overlap any egg

cell siRNA loci. Newly-detected siRNA loci in zygote are gene proximal and not associated with centromeric heterochromatin, similar to canonical siRNAs, in sharp contrast to gametic siRNA loci which are gene-distal and heterochromatic. In addition, zygote but not egg siRNA loci were associated with high DNA methylation in the mature embryo. Thus, the zygote begins transitioning before the first embryonic division to an siRNA profile that is associated with future RdDM in embryogenesis. These findings indicate that in addition to changes in gene expression, the transition to totipotency in the plant zygote is accompanied by resetting of the epigenetic reprogramming that occurred during gamete formation.

Keywords: Small RNAs, DNA methylation, Epigenome, Plant Reproduction, Zygotic genome activation

Author contributions

HX and HF isolated zygote and post-fertilization ovaries. CL produced all small RNA libraries. CL drafted the outline of figures and determined the analyses that would be included in figures. JIG performed all command line analyses (from raw reads to data tables). CL performed all downstream analyses in R (from data tables to figures and statistics). CL produced and assembled all figures. CL drafted the manuscript.

This chapter was posted as a preprint on BioRxiv (2021) and submitted to *Genome Research* for peer review. Supplemental tables and dataset can be found with the online version of this article: https://www.biorxiv.org/content/10.1101/2020.08.31.275958v4

Resetting of the 24-nt siRNA landscape in rice zygotes

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, is crucial to the life cycle of sexually reproducing organisms. It is produced by the fusion of two differentiated cells — the egg and sperm, which in plants have radically different siRNA transcriptomes from each other, and from multicellular embryos. Due to technical challenges, the epigenetic changes that accompany the transition from differentiated gametes to totipotent zygote are poorly understood. Since siRNAs serve as both regulators and outputs of the epigenome, we performed here the successful characterization of small RNA transcriptomes of zygotes from rice. Zygote small RNAs exhibited extensive maternal carryover and an apparent lack of paternal contribution, indicated by absence of sperm signature siRNAs. Zygote formation was accompanied by widespread redistribution of 24-nt siRNAs relative to gametes, such that ~70% of the zygote siRNA loci did not overlap any egg cell siRNA loci. Newly-detected siRNA loci in zygote are gene proximal and not associated with centromeric heterochromatin, similar to canonical siRNAs, in sharp contrast to gametic siRNA loci which are gene-distal and heterochromatic. In addition, zygote but not egg siRNA loci were associated with high DNA methylation in the mature embryo. Thus, the zygote begins transitioning before the first embryonic division to an siRNA profile that is associated with future RdDM in embryogenesis. These findings indicate that in addition to changes in gene expression, the transition to totipotency in the plant zygote is accompanied by resetting of the epigenetic reprogramming that occurred during gamete formation.

Keywords: Small RNAs, DNA methylation, Epigenome, Plant Reproduction, Zygotic genome activation

Background

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, in which a second sperm cell fuses with the central cell and gives rise to the endosperm, a nutritive tissue that nurtures the developing embryo or germinating seedling [reviewed in (Lord and Russell, 2002)]. 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 have occurred (Tadros and Lipshitz, 2009). Recent studies show that material-to-zygote-transition 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 zygotes of maize and *Arabidopsis* (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 zygotic genome activation (ZGA) occurs in the zygote. However, similar to animals, ZGA in plants is gradual. The initial transcriptome of flowering plant zygotes is thus 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 flowering 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). 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 loaded onto an argonaute protein (AGO), which recruits 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 both rice and maize (Tan et al., 2016, 2018; Gent et al., 2013), but not in all plants (Zemach et al., 2013). 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 the germline, 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 seedlings, 24-nt siRNAs coincide with mCHH islands - short regions with high CHH methylation - that 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 [reviewed in Cuerda-Gil, and Slotkin (2016)], including rice gametes (Li et al., 2020, **Fig 1B**). We have previously shown that the siRNA transcriptome is reprogrammed 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 is unknown. A recent study in *Arabidopsis* revealed that siRNAs from heterochromatic TEs are transiently upregulated during embryogenesis, while siRNAs from euchromatic TEs peak at mature green embryos (Papareddy et al., 2020). However, currently there are no data available for siRNA transcriptomes before the preglobular embryo stage, and consequently very little is known about the siRNA landscape in plant zygotes. Since

siRNA production is influenced by histone modifications and DNA methylation, and siRNAs in turn can direct histone modifications and DNA methylation (Law and Jacobsen, 2010; Matzke and Mosher, 2014; Parent et al., 2021), the siRNA transcriptome is an output and indicator of the epigenome. Due to the extreme difficulties associated with plant zygote isolation and corresponding low-input sequencing, epigenome profiles of zygotes have remained poorly characterized in plants. Given the importance of genome wide epigenetic changes associated with the acquisition of totipotency, we undertook a detailed characterization of the small RNA transcriptome of rice zygotes, to investigate the changes that occur soon after fertilization.

Results

We collected 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 post-fertilization ovary of the same developmental stage as zygote (9 hap) and prepared 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 (**Fig S1A**).

<u>The global siRNA pattern in zygotes displays siRNA transcript carryover from the egg cell, and 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 siRNAs are spread out across wide heterochromatic regions, including centromeric tandem repeats. The egg cell and ovary have a pattern different from both sperm and vegetative tissues, in which 24-nt siRNAs are concentrated at discrete loci (**Fig 1A**). We found that in a whole-genome view, the zygote had a similar pattern to the egg cell (**Fig 1A**, zygote vs. egg track). 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**, see **Methods** for additional details). 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 also free of pre-fertilization ovary contamination (Li et al., 2020).

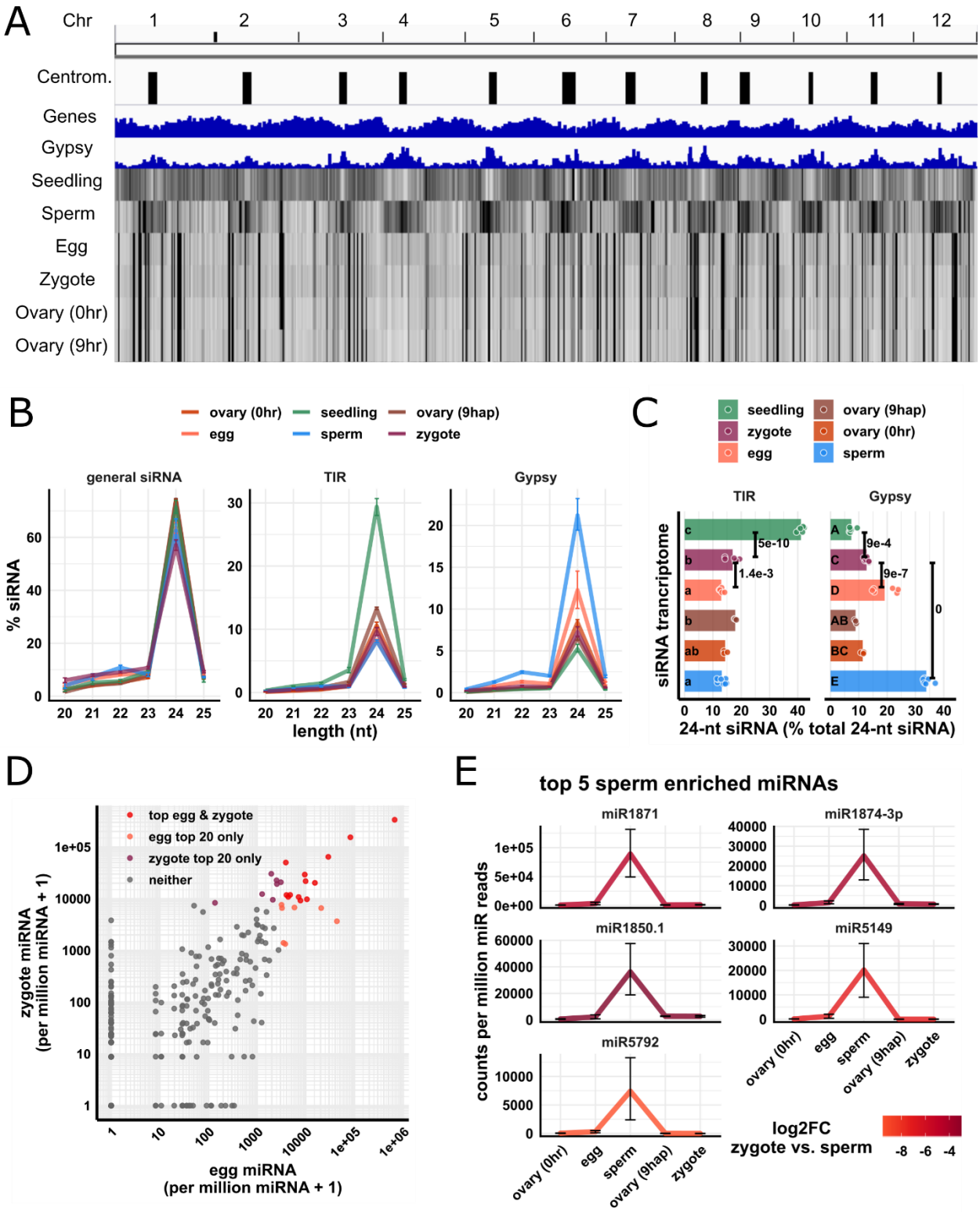


Fig 1: Genome-wide distribution of zygote small RNAs and comparisons of zygote, egg and sperm small RNAs

(A) Heat map showing abundance of 24-nt siRNA across genome at 50-kb resolution. The first three tracks are centromeres [as defined by (Mizuno et al., 2018)], genes, and *Gypsy* retrotransposons.

(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. miRNA and phasiRNA are not included in this analysis (**Fig S1A**).

(C) Quantification of TIR and *Gypsy* panels in (B). Each data point is an siRNA transcriptome. Bar heights are averages. x-axis values are relative to total 24-nt siRNAs. Letter grouping ($\alpha = 0.05$) and P values are based on 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 published 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 lower 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 $P = 5e\text{-}10$, 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 $P = 0$, Tukey tests). However, we noted that while

the zygote and egg cell were similar, there were some clear differences. Zygote had significantly more siRNAs overlapping TIR elements, and significantly less siRNA overlapping *Gypsy* retrotransposons than the egg cell (**Fig 1C,** P = 1.4e-3 and P = 9e-7 respectively, 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 were low expressed in the zygote (Anderson et al., 2017, 2013). Indeed, 13 out of the 20 most abundant miRNAs in egg cells remained among the 20 most abundant miRNAs in zygote (P = 3e-14, Exact test, **Fig 1D**). However, the miRNA accumulation patterns were not identical between zygote and egg. 32 miRNAs were detected in the zygote but not in the egg cell (> 50 reads per million miRNA reads in zygote and undetected in 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 and undetected in zygote). The presence of 32 miRNAs detected in zygote but not egg cell 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, consistent with dilution after fertilization (**Fig 1E** and **Fig S1C**). 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, and < 500 reads per million miRNA reads in the egg cell for this classification. 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**). Taken together, these results imply 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.

Unusual siRNA loci with abundant siRNAs in egg, ovary, zygote, and endosperm

The zygote, egg cell and ovary displayed an unusual set of highly abundant siRNAs that appeared to be concentrated at discrete sites across the genome, 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 (siRNA in the endosperm). 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 2A**). In endosperm and ovaries (pre- and post-fertilization), ~0.1% (n = 73, 213 and 102, respectively) of the siRNA loci accounted for 60% of the total siRNA accumulation in all siRNA loci for each tissue type (**Fig 2A**). Similarly, in egg cell and zygote, ~1% (n = 1881 and 1429, respectively) of the siRNA loci accounted for 60% of the total siRNA accumulation in all siRNA loci for each cell type (**Fig 2A**). We call these highly expressing loci 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 reported in rice endosperm, at least for the specific endosperm stage described , i.e., 7-8 days after fertilization (Li et al., 2020; Rodrigues et al., 2013). Importantly, egg siren siRNAs did not show a significant difference in relative abundance between egg and zygote, without a significant decrease after fertilization (**Fig 2B**), a factor that was taken into account in subsequent analysis (see below).

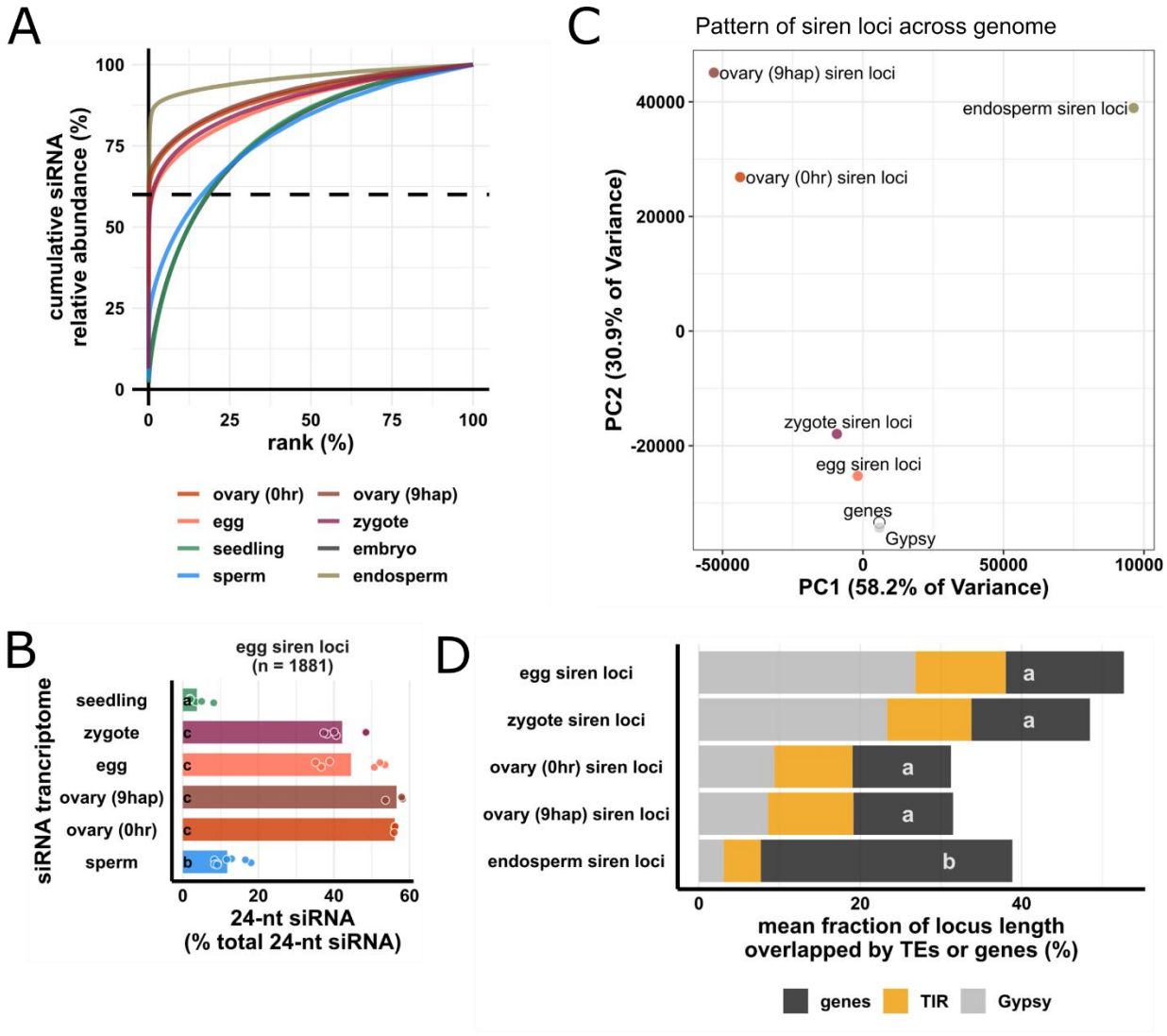

A
cumulative siRNA relative abundance (%)
rank (%)
ovary (0hr)
egg
seedling
sperm
ovary (9hap)
zygote
embryo
endosperm
C
Pattern of siren loci across genome
ovary (9hap) siren loci
endosperm siren loci
ovary (0hr) siren loci
zygote siren loci
egg siren loci
genes
Gypsy
PC2 (30.9% of Variance)
PC1 (58.2% of Variance)
B
egg siren loci
(n = 1881)
siRNA trancriptome
seedling
zygote
egg
ovary (9hap)
ovary (0hr)
sperm
24-nt siRNA
(% total 24-nt siRNA)
D
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
TIR
Gypsy
E
siRNA
transcriptome
ovary (0hr)
egg
seedling
sperm
ovary (9hap)
zygote
embryo
endosperm
24-nt siRNA (RPM)
ovary (0hr)
siren loci
ovary (9hap)
siren loci
egg
siren loci
zygote
siren loci
endosperm
siren loci

Fig 2: Zygote siren loci are similar to siren loci detected in ovary and egg cell and stably expressed between egg and zygote 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%. 0.1% of siRNA loci accounted for 60% of siRNA reads in all siRNA loci in endosperm and ovary. 1% of siRNA loci accounted for 60% of siRNA reads in all siRNA loci in egg and zygote.

(B) Bar plot showing relative abundances of 24-nt siRNA at egg siren loci. Each data point is an siRNA transcriptome. Bar heights are averages. x-axis values are relative to total 24-nt siRNAs.

(C) Principal component plot for 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.

(D) Stacked bar plots showing mean fraction of locus length overlapped by TEs or genes. TIR: terminal inverted repeat transposons, CACTA superfamily excluded. *Gypsy*: *Gypsy* retrotransposons.

(E) 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 siRNAs in each siRNA transcriptome. Whiskers span 2.5^{th} and 97.5^{th} percentiles.

Letter grouping ($\alpha = 0.05$) and P values are based on 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 abundances in 50-kb windows genome-wide (**Fig 2C**). As reference points, the genomic distributions of genes and Gypsy retrotransposons were included. On the PC plot, endosperm siren loci were well separated from all the others along PC1, which accounts for 58% of the variance in their genomic distributions. The rest of the siren loci categories were separated along PC2, which accounts for 31% of the variance, much

less than what was explained by PC1. Ovary siren loci categories (pre- and post-fertilization) had similar genomic distributions, clustering closely together (**Fig 2C**). Egg and zygote siren loci also had nearly the same genomic distribution, clustering closely together (**Fig 2C**). All siren categories have distinct genomic distributions from distributions of genes or Gypsy elements (**Fig 2C**). Consistent with its unique genomic distribution, endosperm siren loci were more likely to overlap a gene (**Fig 2D**). On average, ~30% of the locus length of an endosperm siren locus was covered by a gene, whereas all the other siren categories display a similar fraction of locus length covered by genes (~13%, $P < 1.4e-4$, Tukey tests). Lastly, we compared the relative abundances of 24-nt siRNAs at different siren categories across different cell types. 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 higher than the levels in all other cell types we examined (**Fig 2E**). In contrast, the other siren classes shared a siRNA accumulation pattern across cell types (**Fig 2E**). Ovaries (pre- and post-fertilization), egg cell and zygote all had high abundances of 24-nt siRNAs at ovary/egg/zygote siren loci, consistent with the stable expression of egg siren siRNAs in zygote (**Fig 2B**), 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 reveal that zygote siRNAs were concentrated at discrete sites similar to egg and ovary, and that the persistence of egg siren siRNAs explains the overall similarity between zygote and egg (**Fig 1A**).

The genome distribution pattern of zygote 24-nt siRNA loci is distinct from egg

Although the siren siRNAs result in similarity of the zygote siRNA profile in the zygote to that of 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 3A**). 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 cells (**Fig 3A-B**, $P = 3e-8$, Tukey tests). In contrast, there was no significant changes between pre- and post-fertilization ovaries (**Fig 3A-B**, $P = 0.98$, Tukey tests). Thus, the differences between zygote and

egg could not be due to trafficking of the newly-transcribed siRNAs from 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-signature loci as seedling siRNA loci that did not overlap any egg siRNA loci or sperm siRNA loci (seedling loci ∉ egg loci ∉ sperm loci, **Fig 3C**). Overlapping siRNA loci were defined as at least 1-bp overlap in genomic coordinates (see also **Methods**). Likewise, we classified sperm-signature loci as sperm siRNA loci that did not overlap any egg or seedling siRNA loci (sperm loci ∉ egg loci ∉ seedling loci, **Fig 3C**), and lastly, egg-signature loci as egg siRNA loci that did not overlap any seedling or sperm siRNA loci (egg loci ∉ seedling loci ∉ sperm loci, **Fig 3C**). At egg-signature loci, zygote experienced a 10-fold reduction of 24-nt siRNAs (**Fig 3D**, P = 5e-14, Tukey tests). At seedling-signature loci, zygote had 4.7-fold more 24-nt siRNAs than egg cell (**Fig 3D**, P = 1e-13, Tukey tests). Gaining siRNAs at gene-proximal regions and seedling-signature loci is consistent with an increase of TIR siRNAs in zygote (**Fig 1C**). Since these seedling-signature loci did not overlap any egg siRNA loci or sperm siRNA loci, the increase of 24-nt siRNAs at seedling-signature loci in zygote was unlikely due to carryover from either gamete. At sperm-signature loci, zygote had very few 24-nt siRNAs (**Fig 3D**, zygote vs. sperm P = 4e-14, Tukey tests), much like the results for miRNAs (**Fig 1E**), confirming small RNA contribution from sperm cell is very limited relative to egg. There was little difference in the ovaries before and after fertilization for any of these locus categories (**Fig 3D**, P = 0.76, P = 0.84 and P = 0.84 at egg-, seedling- and sperm-signature loci, respectively). It is important to note that the zygote siRNA transcriptome was not used to define these locus categories. Lastly, we bioinformatically removed siren siRNAs from egg and zygote libraries (**Fig 3E**) and re-analyzed their genome-wide 24-nt siRNA distributions. This analysis revealed that outside of the siren loci (which as defined previously constitute ~1% of all 24-nt siRNA loci), zygote and egg were indeed distinct from each other in genome-wide 24-nt siRNA distribution (**Fig 3E** and **Fig S2F,** P = 0, Tukey Tests). Egg cell has a slight enrichment of 24-nt siRNAs at centromeric regions, while zygote showed a relative depletion of siRNAs at centromeric regions (**Fig S2G**, egg vs. zygote P = 0, Tukey tests), much like embryo (zygote vs embryo P = 0.5, Tukey tests; **Fig 3E**, pink boxes indicate two examples, see also **Fig S2G**). Taken together, these results indicate that the zygote has an siRNA transcriptome that is distinct from that of the egg cell, and further, that the changes from egg cell to zygote were independent of post-fertilization changes in the ovary.

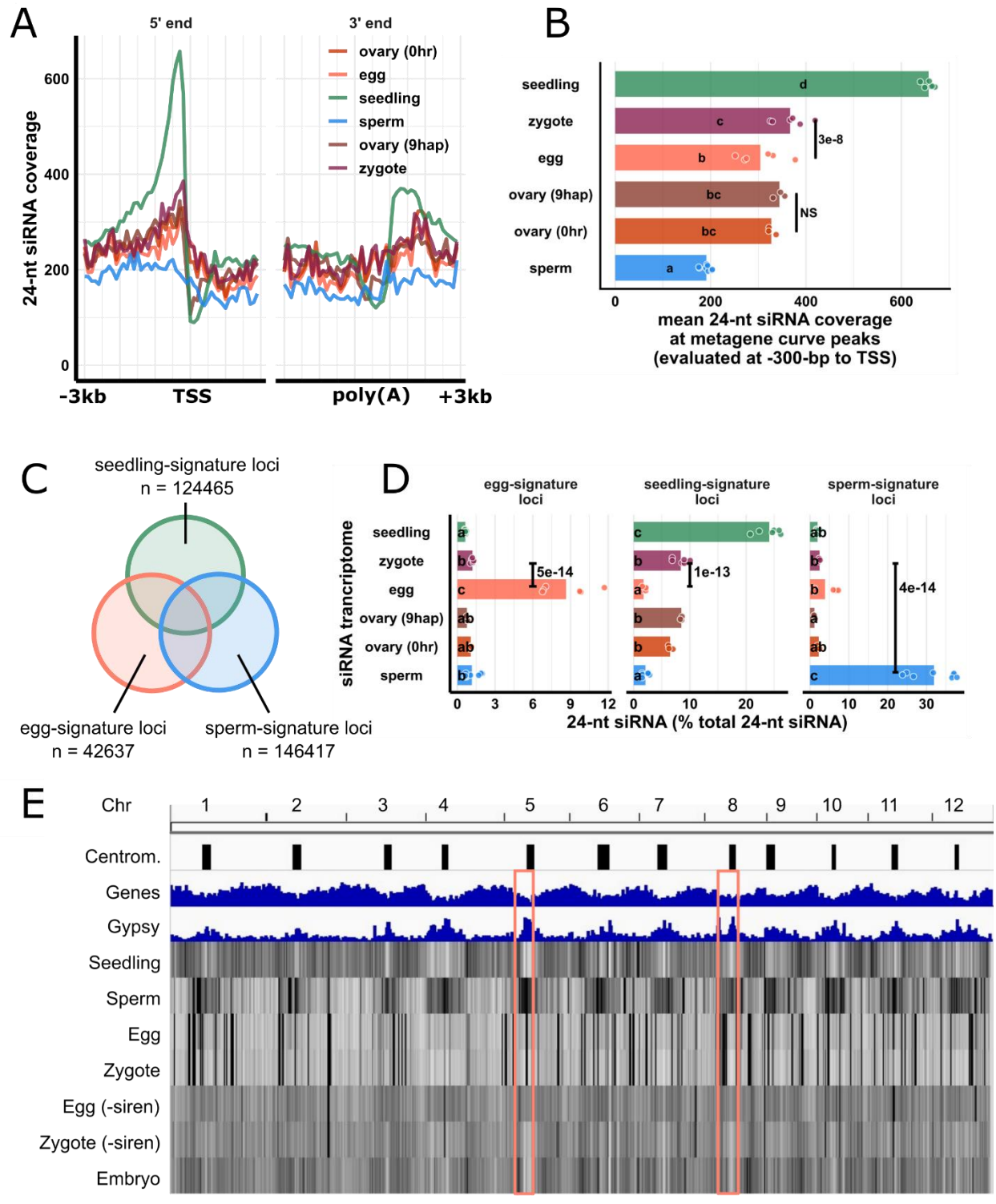


Fig 3: Changes in the zygote siRNA transcriptome after fertilization are independent from the ovary.

(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.

(C) Venn diagram illustrating egg-signature loci (egg siRNA loci that do not overlap any seedling or sperm siRNA loci), seedling-signature loci (seedling siRNA loci that do not overlap any egg or sperm siRNA loci), and sperm-signature loci (sperm siRNA loci that do not overlap any egg or seedling siRNA loci). Sizes of overlap in Venn diagrams are not to scale.

(D) Bar plot showing relative abundances of 24-nt siRNA across siRNA loci categories defined in (C). 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 total 24-nt siRNAs.

(E) Heat map showing abundance of 24-nt siRNA across genome at 50-kb resolution. The first three tracks are centromeres [as defined by (Mizuno et al., 2018)], genes, and *Gypsy* retrotransposons. '-siren' refers to siren siRNAs removed. Pink boxes highlight examples where egg and zygote are distinct.

Letter grouping ($\alpha = 0.05$), and P values are based on Tukey tests. Zygote and 9 hap ovary data are from this study, embryo (7-8 DAF) from Rodrigues et al. (2013), 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 next defined zygote siRNA loci using Shortstack with zygote siRNAs. We then classified Z-E loci as zygote siRNA loci that did not overlap any egg cell siRNA loci (Z loci $\notin$ E loci in set operation), E-Z loci as egg siRNA loci that did not overlap any zygote siRNA loci (E loci $\notin$ Z loci), and Z/E loci intersect as zygote siRNA loci that overlapped egg siRNA loci (Z loci $\cap$ E loci, **Fig 4A**). Despite the similarities between egg and zygote at the high abundance siRNA loci (**Fig 1A**, **Fig 2B**), widespread distinct siRNA loci were detected in one cell type but

not the other. There were 101,841 E-Z loci (newly diminished siRNA loci in zygote), 94,591 Z-E loci (newly detected siRNA loci in zygote), but only 42,437 Z/E loci intersect.

When 24-nt siRNA reads at individual loci were tallied and normalized to total 24-nt siRNAs, as expected, we found that at E-Z loci, egg had ~10-fold more 24-nt siRNAs than the zygote (**Fig 4B**, $P = 0$, Tukey tests); at Z-E loci, zygote had ~10-fold more 24-nt siRNAs than egg (**Fig 4B**, $P = 0$, Tukey tests); and no difference at Z/E loci intersect. There were siRNAs not captured by siRNA loci. These siRNAs resided at genomic regions with insufficient siRNAs and did not meet the 0.5 RPM threshold for assignment as loci on Shortstack (see also **Methods**), which explains the small number of egg siRNAs at Z-E loci and the small number of zygote siRNAs at E-Z loci. There were no differences between ovaries before and after fertilization in any of the three locus categories (**Fig 4B**), again suggesting changes in the zygote siRNA transcriptome were not coupled with the ovary (**Fig 3**). In addition, the abundance of seedling siRNAs in Z-E loci and scarcity in E-Z loci revealed the emergence of a seedling-like siRNA pattern in zygote (**Fig 4B**). Since the seedling siRNA transcriptome was not used to classify Z-E loci, this observation suggests that the zygote has initiated a return to the canonical siRNA profile, consistent with the increase in 24-nt siRNAs from TIR transposons (**Fig 1C**) as well as at gene-proximal regions (**Fig 3A-B**) and seedling-signature loci in zygote (**Fig 3D**).

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 transcript carryover (Anderson et al., 2017; Chen et al., 2017; Zhao et al., 2019). Thus, if the siRNA transcriptome transitions similarly in the zygote, one would expect to see an initial widespread detection of low abundance 24-nt siRNAs at new loci, relative to a background of more abundant maternal carryover siRNAs corresponding to egg siren loci. Indeed, in contrast to the high abundance siRNAs of intersect loci, Z-E loci and E-Z loci overall had lower siRNA abundances than zygote/egg intersect loci (**Fig 4B**). Nevertheless, on average, one in every five zygote 24-nt siRNAs (~20%) resided at Z-E loci in the zygote. Together with the numerical abundance of Z-E loci (70% of all zygote loci) these results suggest that newly detected siRNA loci in zygote are widespread and explain a substantial fraction of 24-nt siRNAs in zygote.

The highly expressed siren loci in egg and zygote raise the concern of whether the apparent upregulation of Z-E loci could be explained by downregulation of egg siren loci. Consistent with the stable expression of egg siren loci in zygote, including or excluding siren

siRNAs did not change the results of analyses (**Fig S2**, **Fig 3A-D**, **Fig 4A-B**), supporting the distinct distributions of non-egg-siren 24-nt siRNAs in egg and zygote (**Fig 3E**). Taken together, changes in the zygote siRNA transcriptome are not explained by downregulation of abundance egg siren siRNAs, but due to up- and downregulation of other siRNA loci that are widespread across genome.

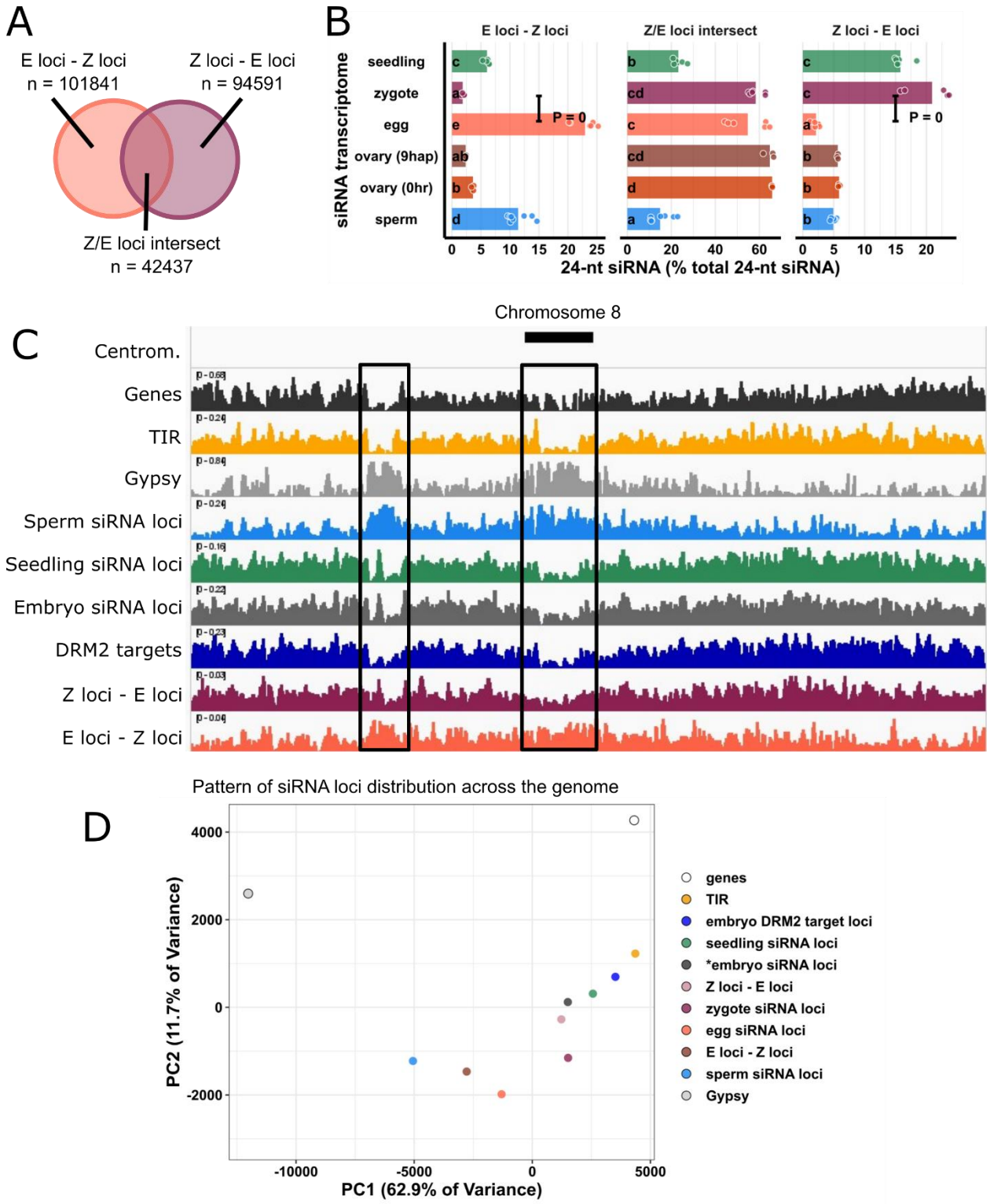


Fig 4: Widespread distribution of newly detected zygote siRNA loci across the rice genome

(A) Venn diagram illustrating E-Z loci (egg siRNA loci that do not overlap any zygote siRNA loci, E loci ∉ Z loci), Z-E loci (zygote siRNA loci that do not overlap any egg

siRNA loci Z loci ∉ E loci), and Z/E loci intersect (zygote siRNA loci that overlap egg siRNA loci, Z loci ∩ E loci). Sizes of overlap in Venn diagrams are not to scale.

(B) 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 24-nt siRNA reads. Letter grouping ($\alpha = 0.05$), and P values are based on Tukey tests.

(C) Distribution of siRNA loci along a chromosome. Chromosome 8 is chosen because it is one of the chromosomes with a completed sequenced centromeric region (Mizuno et al., 2018). Centrom. Centromeric regions; TIR: terminal inverted repeat transposons, CACTA superfamily excluded. *Gypsy*: *Gypsy* retrotransposons. Black boxes highlight regions with abundant *Gypsy* retrotransposons and relative depletion of TIR, seedling siRNA loci, embryo siRNA loci, DRM2 targets, and Z-E loci.

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

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

<u>Newly-detected siRNA loci in zygote diverge from gamete siRNA loci and resemble canonical siRNA loci in genomic location and DNA methylation</u>

To investigate the patterns and characteristics of zygote siRNA loci, we compared the genomic distribution of zygote siRNA loci and Z-E loci against a set of different siRNA loci categories, including E-Z loci, egg siRNA loci, embryo siRNA loci (Rodrigues et al., 2013), seedling siRNA loci, and sperm siRNA loci. Our efforts to generate robust DNA methylome profiles for zygotes were not successful, possibly because zygotes are fragile as compared to egg cells, and the output of random-primed based methylome sequencing methods are highly sensitive to library preparation conditions (Li et al., 2019). We have previously shown that the sperm and egg methylomes are similar, and that gamete-specific 24nt siRNA loci are associated with sites of DDM1-dependent methylation and not DRM2-dependent methylation (Li et al. 2020). To determine whether the zygote siRNAs might be correlated with DRM2-dependent methylation during embryogenesis, we analyzed a rice *drm2* mutant generated by CRISPR-Cas9 genome editing. By comparing mCHH between mature wildtype and *drm2* embryos, we had previously identified a set of DRM2-dependent methylation sites in the embryo (Li et al., 2020).

For reference points in the genome-wide analysis, we also included genes, TIR transposons and *Gypsy* retrotransposons (Kawahara et al., 2013). We observed that the genome-wide distribution of Z-E loci follow the distribution of TIR elements, as well as embryo DRM2 targets, and resemble the distribution of seedling and embryo siRNA loci (**Fig 4C**). In contrast, the E-Z loci representing the newly diminished loci in zygote had a distinct pattern, more similar to sperm siRNA loci, which are predominantly heterochromatic (**Fig 4C**). There was a relative depletion of siRNA loci from centromeric regions for zygote siRNA loci, much like canonical siRNA loci, and unlike egg siRNA loci (**Fig S3A**, P = 4e-10, Tukey tests). Consistent with a more similar distribution to canonical siRNA loci, zygote siRNA loci and Z-E loci had higher degrees of overlap with seedling siRNA loci and embryo DRM2 targets, while egg siRNA and E-Z loci had low degrees of overlap, as did sperm siRNA loci (**Fig S3B**). Zygote siRNA loci and Z-E loci overlapped larger numbers of DRM2 targets per Mb genome space, much like seedling siRNA loci, and unlike egg siRNA loci, E-Z loci or sperm siRNA loci (**Fig S3C**).

To gain more information on the factors underlying the variation in siRNA loci distributions, we used principal component analysis (PCA) to cluster the genomic distributions of the above locus categories based on their abundance in 50-kb windows genome-wide (**Fig 4D**). The separation of locus categories along PC1 had a near-perfect rank order correlation with median distance to nearest genes (**Fig S3D**, rho = -0.98, P = 0), and PC2 was correlated with median length of locus categories (**Fig S3E**, rho = 0.88, P = 7e-4). PC1, which explained 63% of variance in genomic distributions across loci categories, was strongly correlated with various aspects of rice genome organization. PC1 was strongly correlated with TIR transposon overlap as well as mCHH level in wildtype embryo (**Fig S4A** and **C**). PC1 was also strongly anti-correlated with *Gypsy* retrotransposon overlap and mCG, and to a lesser extent mCHG in wildtype embryo (**Fig S4B, D** and **E**). These genomic features are mutually correlated (**Fig S4F**), consistent with the prior understanding of cereal genome organization (Gent et al., 2013; Han et al., 2013).

Strong correlations between PC1 (gene proximity), TE overlap, and DNA methylation led us to statistically assess the differences of these attributes among siRNA loci categories. TIR transposons, where RdDM is known to take place in cereal genomes, was gene proximal, consistent with the gene proximal distribution of embryo DRM2 targets (**Fig 5A**). Canonical siRNA loci, such as seedling siRNA loci and embryo siRNA loci, were closer to genes than non-canonical siRNA loci, such as sperm siRNA loci (**Fig 5A**). E-Z loci, the newly diminished

siRNA loci in zygote, were on average much farther away from genes than Z-E loci were (**Fig 5A**, 2.8-kb vs. 1.6-kb, $P = 0$ Tukey tests), consistent with their heterochromatic genomic distributions (**Fig 4C-D**, **Fig S3A**). Total zygote siRNA loci were closer to genes than total egg siRNA loci ($P = 0$). From egg cell to zygote, there was a 30% decrease in median distance (2.4-kb vs. 1.6-kb). In contrast, from zygote to embryo (7 days after fertilization, data from Rodrigues et al., 2013), there was an 6% decrease (1.6-kb vs. 1.5-kb). In gametes, siRNA loci were more likely to overlap a *Gypsy* retrotransposon than a TIR transposon (**Fig 5B**). However, in sporophytes, including zygote itself, siRNA loci are more likely to overlap a TIR transposon than a *Gypsy* retrotransposon (**Fig 5B**). These observations are consistent with the results where zygote had more TIR siRNAs and less *Gypsy* siRNAs (**Fig 1C**), more gene-proximal 24-nt siRNAs than egg cell (**Fig 3A-B**), zygote had increased siRNAs in seedling-signature loci (**Fig 3C**), and seedling had comparable siRNA level with zygote in Z-E loci (**Fig 4B**).

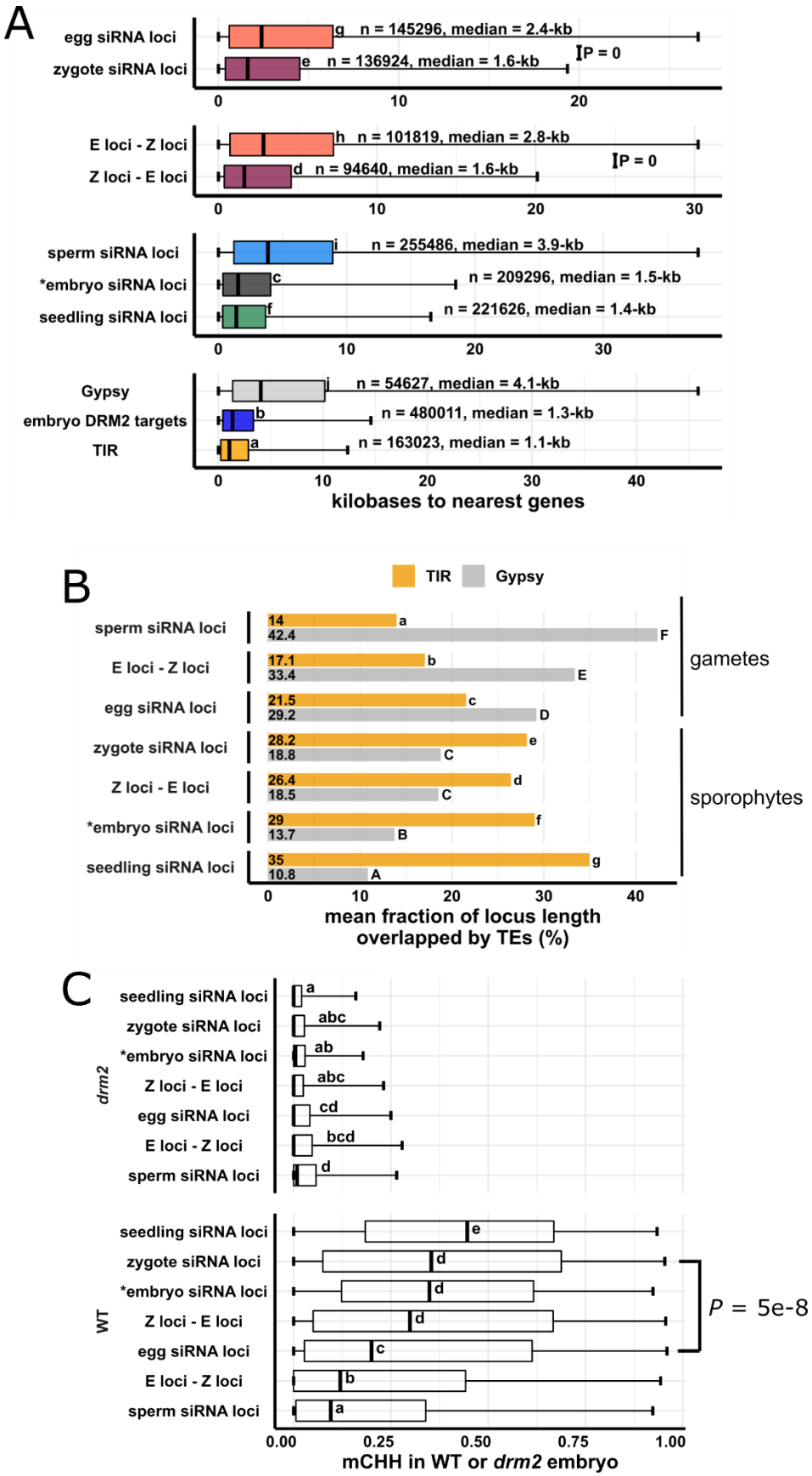

A
egg siRNA loci
zygote siRNA loci
g n = 145296, median = 2.4-kb
e n = 136924, median = 1.6-kb
P = 0
E loci - Z loci
Z loci - E loci
h n = 101819, median = 2.8-kb
d n = 94640, median = 1.6-kb
P = 0
sperm siRNA loci
*embryo siRNA loci
seedling siRNA loci
i n = 255486, median = 3.9-kb
c n = 209296, median = 1.5-kb
f n = 221626, median = 1.4-kb
Gypsy
embryo DRM2 targets
TIR
j n = 54627, median = 4.1-kb
b n = 480011, median = 1.3-kb
a n = 163023, median = 1.1-kb
kilobases to nearest genes
B
TIR
Gypsy
sperm siRNA loci
E loci - Z loci
egg siRNA loci
zygote siRNA loci
Z loci - E loci
*embryo siRNA loci
seedling siRNA loci
14 a
42.4 F
17.1 b
33.4 E
21.5 c
29.2 D
28.2 e
18.8 C
26.4 d
18.5 C
29 f
13.7 B
35 g
10.8 A
gametes
sporophytes
mean fraction of locus length overlapped by TEs (%)
C
drm2
WT
seedling siRNA loci
zygote siRNA loci
*embryo siRNA loci
Z loci - E loci
egg siRNA loci
E loci - Z loci
sperm siRNA loci
a
abc
ab
abc
cd
bcd
d
e
d
d
d
c
b
a
P = 5e-8
mCHH in WT or drm2 embryo

Fig 5: Newly detected siRNA loci in zygote reset to the canonical siRNA profile and predict CHH methylation in embryo in an RdDM-dependent manner.

(A) 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.

(B) Bar plots showing mean locus length overlapped by TIR or *Gypsy* transposons across siRNA loci categories. Statistical comparisons are made across siRNA loci categories within a TE superfamily.

(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. E-Z loci: n = 101,841, Z-E loci: n = 94,591 (69% of all zygote siRNA loci).

Letter groupings ($\alpha = 0.05$) and P values are based on Tukey tests. *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).

Resetting to a 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 5C**, see also **Fig S5**). Although all siRNA loci categories were associated with RdDM during embryogenesis, as median mCHH levels were all higher in wildtype embryo than *drm2* embryo, zygote siRNA loci had much higher level of mCHH than egg siRNA loci did in wildtype embryo (**Fig 5C**, P = 5e-8), much like embryo and seedling siRNA loci. Importantly, newly detected siRNA loci (as represented by Z-E loci) had high mCHH levels in wildtype embryo (P = 0), whereas newly diminished siRNA loci (E-Z loci) had low mCHH levels in wildtype embryo. In addition, while there were smaller differences in mCHG and especially in mCG across locus categories in wildtype embryo when compared to mCHH, the full extent of methylation across siRNA loci categories also depended on DRM2 for all three contexts (**Fig S5**). It is important to note that zygote siRNA loci had high degrees of overlap with canonical RdDM loci (seedling siRNA loci and embryo DRM2 targets, **Fig S3B**), and that zygote siRNA loci overlapped large number of DRM2 targets per Mb genome space (**Fig S3C**). Together with elevated mCHH level in embryo, these results suggest that newly detected zygote siRNA loci mark canonical siRNA loci that will undergo hypermethylation

during embryogenesis, rather than defining hypermethylated regions that are distinct from canonical siRNA loci.

Lastly, the fact that siren loci were defined by abundant 24-nt siRNAs (**Fig 2A**) led us to speculate that they would have high mCHH. In fact, we found the opposite, i.e., they had lower mCHH levels relative to other siRNA loci in each tissue (**Fig S6**). In ovary, siren loci had lower mCHH level than canonical siRNA loci (seedling siRNA loci), and ovary siRNA loci that were not siren loci had comparable mCHH level to siren loci, if not higher (**Fig S6A**). Similar results were found for egg cell and embryo as well (**Fig S6B-C**). At siren loci as well as siRNA loci that were not siren loci, DRM2 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 S6D**). Unlike mCHH, mCG and mCHG did not produce any notable pattern across cell types (**Fig S6**). Taken together, these results suggest that although the highly abundant siRNAs produced by siren loci may also target DNA methylation in a DRM2-dependent manner, they do so inefficiently as compared to siRNAs from canonical RdDM loci.

Discussion

The parental gametes have unequal contributions to the zygote siRNA transcriptome in rice

The zygote is an indispensable stage of the plant life cycle, in which the genomes of the differentiated gametes have to be reprogrammed to specify a totipotent cell that will regenerate a new plant (McClintock, 1983). However, the epigenetic changes that accompany this critical transition are poorly understood, due to the technical challenges of working with a cell type that presents major difficulties arising from the highly limiting and fragile nature of the material. To gain insights into these epigenetic changes, we present here the successful characterization of the small RNA transcriptome of plant zygotes. The results have notable implications for the current models of post-fertilization silencing through the male germline. 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) in *Arabidopsis*. Our findings that the zygote is depleted for sperm-signature siRNAs (Fig 3D) as well as sperm-enriched miRNAs (Fig 1E, Fig S1C) indicate that there is negligible paternal contribution to the

zygote small RNA transcriptome. Thus, at least in rice embryogenesis, any proposed 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 chromatin 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).

A special class of highly abundant maternal siRNAs persists in the zygote

We found that a small number of loci accounted for most of the siRNAs in egg cells and zygotes (**Fig 2A**). We refer to these loci as siren loci, using the term applied to similar loci in rice endosperm (Rodrigues et al., 2013), and more recently in *Brassica rapa* and *Arabidopsis* ovules (Grover et al., 2020). Importantly, egg siren siRNAs remained similarly highly expressed in the zygote (**Fig 2B**). Thus, upregulation of siRNAs at newly detected zygote siRNA loci cannot be explained by large downregulation of egg siren loci, which was further demonstrated by similar results for these newly detected loci obtained after excluding the siren siRNAs from the analysis (**Fig S2**). Siren loci were first discovered in rice endosperm (Rodrigues et al., 2013). In *Arabidopsis* and *Brassica rapa* (Grover et al., 2020), siren loci detected in ovules are also highly expressed in the endosperm; however, siren loci in rice ovary are distinct from those detected in rice endosperm (**Fig 2B-E**). Siren loci in the zygote were distinct from endosperm siren loci in endosperm collected 7-8 days after fertilization, instead coinciding with siren loci detected in ovary and egg cell (**Fig 2B-E**). 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. Although 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 S6**). Therefore, it is unlikely that siren siRNAs play a role in embryogenesis through directing DNA methylation. However, we cannot exclude 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.

The 24-nt siRNA landscape of rice zygotes reveals resetting towards a canonical siRNA pattern

We detected widespread new zygote siRNA loci relative to the egg cell, representing ~69% of all zygote siRNA loci. There were 94,591 zygote siRNA loci that did not overlap any egg siRNA loci (Z loci ∉ E loci or Z-E loci, **Fig 4A**), as compared to 42,437 siRNA loci that overlapped egg siRNA loci (Z loci ∩ E loci or Z/E loci intersect, **Fig 4A**). In addition, 101,841 egg siRNA loci were diminished in zygote (E loci ∉ Z loci or E-Z loci, **Fig 4A**). In relative abundance, most of the siRNA reads were accounted for by egg siRNA carryover and stably-expressed egg siren siRNAs (**Fig 2B**), and thus siRNA abundance was lower at Z-E loci than at Z/E loci intersect which contains the siren siRNAs (**Fig 4B**). 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, either by RNA polymerase IV or RNA polymerase II, 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. In fact, the relative abundance of the zygote siRNAs at Z-E loci (~20%) is similar to that of egg cell siRNAs at E-Z loci (~22.5%; **Fig 4B**), but they differ significantly in their genome-wide distribution as discussed below.

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 siRNA from TIR transposons, and decreased siRNAs from *Gypsy* retrotransposons, as compared to the egg cell (**Fig 1C**). Second, zygote had increased 24-nt siRNAs at seedling-signature loci (**Fig 3D**), as compared to the egg cell. In contrast, there were

no significant changes to the 24-nt siRNAs at seedling-signature loci in ovaries pre- and post-fertilization (**Fig 3D**). Moreover, seedling had comparable 24-nt siRNAs to zygote at Z-E loci (**Fig 4B**). As the zygote siRNA transcriptome was not used to define seedling-signature loci, and the seedling siRNA transcriptome was not used to define Z-E loci, these results serve as an objective indication that the zygote shifted towards a more seedling-like siRNA transcriptome. Third, 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 3A-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 3A**). Fourth, the genomic distribution of Z-E loci is more similar to TIR transposons, embryo siRNA loci, embryo DRM2 targets and seedling siRNA loci, while that of E-Z loci is not (**Fig 4C-D**). Consistent with the major contribution of Z-E 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 4C**). Fifth, similarities in genomic distribution were confirmed by distances to the nearest genes and TE overlaps (**Fig 5A-B**). Notably, there was a 30% decrease in median distance to genes from egg to zygote, which took place over the course of less than one cell cycle. Only an 6% decrease occurred during the transition from zygote to embryo (7 days after fertilization), occurred over the course of numerous cell cycles. Consistent with distance to nearest genes, gamete siRNA loci were more likely to overlap a *Gypsy* element than a TIR element; while zygote siRNA were more likely to overlap a TIR element instead, like the rest of sporophyte siRNA loci categories.

Newly detected zygote siRNAs mark future CHH hypermethylation sites in mature embryos

Hypermethylation of embryo has been reported in a number of angiosperm species, including *Arabidopsis*, soybean, chickpea, *Brassica rapa*, and rice (Bouyer et al., 2017; Chakraborty et al., 2021; Kawakatsu et al., 2017; Li et al., 2020; Lin et al., 2017; Rajkumar et al., 2020; Zhou et al., 2021). We found that newly detected siRNA loci have abundant CHH methylation in embryos that is dependent on the RdDM methyltransferase DRM2 (**Fig 5C**). Although all siRNA loci categories had higher mCHH levels in wildtype embryo than in *drm2* embryo, zygote siRNA loci had higher mCHH levels than egg siRNA loci in mature wildtype embryo, resembling embryo and seedling siRNA loci. These results indicate that newly detected

siRNA loci in zygote not only reset to canonical siRNA pattern, but also that the corresponding 24-nt siRNAs are capable of targeting high CHH methylation during embryogenesis. Since zygote siRNA loci have a similar distribution to canonical siRNA loci (**Fig 4D**) and had substantial degrees of overlap with seedling siRNA loci and embryo DRM2 targets (**Fig S3B-C**), zygote siRNAs are associated with high CHH methylation at regions similar to seedling siRNA loci and embryo DRM2 targets, instead of regions independent from canonical siRNA loci. Thus, resetting of the gametic 24nt siRNA loci to a distribution that results in embryo hypermethylation appears to be initiated in the zygote before the first embryonic division. Reminiscent of the increased heterochromatic siRNAs in rice gametes, a recent paper in *Arabidopsis* revealed that heterochromatin becomes decondensed during embryogenesis and promotes a transient production of siRNAs from heterochromatic TEs at the preglobular stage (Papareddy et al., 2020). The *Arabidopsis* embryo siRNAs from euchromatic TE and canonical siRNA loci peaked towards the end of embryogenesis. Because neither egg nor zygote siRNAs have been sequenced yet in *Arabidopsis*, the relationship of the heterochromatic siRNAs in older *Arabidopsis* embryos to the reprogramming of siRNAs in egg cells and zygotes, as described here for rice, remains to be determined.

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). In mammals, studies have found a progressive change in epigenomes after the two-cell embryo stage and concluded by the blastocyst stage (Xu and Xie, 2018). As detailed above, this study indicates that in plants, the zygote inherits maternal but not paternal 24-nt siRNAs from the gametes, and initiates a resetting towards a canonical siRNA pattern that sets the stage for the methylation pattern in the embryo. Lastly, as siRNA expression is influenced by histone modifications, 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 appears likely that resetting of the other features of the epigenome, such as histone modifications and chromatin conformation, may also be initiated in plant zygotes as a component of their transition to totipotency. It would be of interest to investigate whether such epigenetic resetting is associated with other examples of acquisition of pluripotency in plants,

especially from germline cells, such as the regeneration of haploid plants from anther cells after heat stress (Ibáñez et al., 2020)

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 0.3 M mannitol. The lower part of the cut ovary was gently pushed using an acupuncture needle to separate selected cells under a phase contrast inverted microscope. Once the zygote was separated and floated out of the ovary incision, it was captured by a fine glass capillary and immediately frozen in liquid nitrogen. We routinely culled any unfertilized egg cells that did not conform to zygotic cell morphology during our collections (Anderson et al., 2017). 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 analyzed using a Bioanalyzer (Agilent) to check for RNA integrity, with the eukaryotic total RNA-pico program. RNA input for library construction was ~30 ng. 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., 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 coverage tool (Quinlan and Hall, 2010). For analysis of overlaps of siRNAs at *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 heat maps were made on 50-kb intervals using IGVtools (Thorvaldsdottir et al., 2013). For better visualization of midrange values, heatmap intensity was maxed out at 1.25× coverage per 10 million 24-nt siRNAs.

miRNA analysis

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.

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. Each cell type was downsampled to 3.5 million small RNAs first. For each tissue type (pre- and post-fertilization ovary, egg cell, sperm cell, zygote, seedling, embryo and endosperm), siRNA loci were defined as RPM > 0.5, 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 60% 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 60% of the cumulative RPM in the ovary. The 60% cutoff was selected based on the turning point of cumulative expression vs. percentage rank plot of ovary (**Fig 2A**). Seedling-signature loci were identified as seedling siRNA loci that did not overlap any sperm siRNA loci or egg siRNA loci (seedling loci ∉ egg loci ∉ sperm loci, **Fig 3C**) using the BEDTools intersect tool (Quinlan and Hall, 2010). Overlaps were defined as at least 1-bp overlapping genomic coordinates. Similarly, sperm-signature loci were identified as sperm siRNA loci that did not overlap any egg siRNA loci or sperm siRNA loci (sperm loci ∉ egg loci ∉ seedling loci, **Fig 3C**). Egg-signature loci were identified as egg siRNA loci that did not overlap any seedling siRNA loci or sperm siRNA loci (egg loci ∉ seedling siRNA loci ∉ sperm siRNA loci, **Fig 3C**). Z-E loci were zygote siRNA loci that did not overlap egg siRNA loci (Z loci ∉ E loci). E-Z loci were egg siRNA loci that did not overlap zygote siRNA loci (E loci ∉ Z loci). Z/E loci intersect were zygote siRNA loci that overlapped egg siRNA loci (Z loci ⋂ E loci, **Fig 4A**).

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 2E**, **Fig 3D**, **Fig 4B**, **Fig 5B**, **Fig S1A**, **Fig S2D**, **Fig S2F**, and **Fig S3B**, and at the level of genotype and context for **Fig 5C**, **Fig S5** and **Fig S6**. For analyses of siRNA relative abundances or siRNA coverage across siRNA locus category across siRNA transcriptomes, a linear model was fitted using logit transformation to correct for heteroscedasticity (**Fig 1C**, **Fig 2B**, **Fig 3A**, **Fig 3B**, **Fig 3D**, **Fig 4B, Fig S1A**, **Fig S2B**, **Fig S2D**, and **Fig S2F**). For analyses of siRNA counts or locus counts, a linear model was fitted using log(RPM + 1) transformation to correct for heteroscedasticity (**Fig 2E**, **Fig S2G**, **Fig S3A**). For analyses of distances to nearest genes, a generalized linear model was fitted using log link function to correct for heteroscedasticity (**Fig 5A**). For analyses of fraction of locus length covered by genes, a generalized linear model of quasibinomial family with logit link function was fitted to accommodate the mean-error relationship of fractional data (**Fig 2D**, **FigS3B**, **Fig 5B**). For analyses of DNA methylation levels across different locus categories, a generalized linear model of quasibinomial family with logit link function was fitted to accommodate the mean-error relationship of proportion data (**Fig 5C**, **Fig S5**, **Fig S6**). For analysis of correlations between PC1 (**Fig 4D**), distance to nearest genes, TE overlaps and DNA methylation, Spearman's rank order correlation was used (**Fig S3D-E**, **Fig S4**). P values < 2.2e-16, which is the smallest positive floating point number R can display (*R Core Team*, 2020), were treated as 0 by R, and reported as such in this study.

Supplemental information

Additional file 1: Table S1. General mapping statistics

Additional file 2: Table S2. Pollination success rates

Additional file 3: Supplemental figures. Fig S1-S6

Additional file 4: Supplemental Dataset1. miRNA read counts

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 partial support 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).

Availability of data and materials

All small RNA data have been deposited in the Sequence Read Archive, BioProject PRJNA533115. All R codes regarding data visualization and statistical analyses were deposited in https://github.com/cxli233/zygote_smRNA/

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.

Competing interests

The authors declared that they have no conflicts of interests.

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

Axtell, M.J., 2013b. 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

Bouyer, D., Kramdi, A., Kassam, M., Heese, M., Schnittger, A., Roudier, F., Colot, V., 2017. DNA methylation dynamics during early plant life. Genome Biol. 18, 179. https://doi.org/10.1186/s13059-017-1313-0

Calarco, J.P., Borges, F., Donoghue, M.T.A., Ex, F. Van, Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijo, A., Becker, D., Martienssen, R.A., 2011. Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA. Cell. https://doi.org/10.1016/j.cell.2012.09.001

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

Chakraborty, T., Kendall, T., Grover, J.W., Mosher, R.A., 2021. Embryo CHH hypermethylation is mediated by RdDM and is autonomously directed in Brassica rapa. Genome Biol. 22, 140. https://doi.org/10.1186/s13059-021-02358-3

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

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.

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

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, Jonathan 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 628–637. https://doi.org/10.1101/gr.146985.112.as
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
Ibáñez, S., Carneros, E., Testillano, P.S., Pérez-Pérez, J.M., 2020. Advances in Plant Regeneration: Shake, Rattle and Roll. Plants 9, 897. https://doi.org/10.3390/plants9070897
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

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., 2013b. 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

Kawakatsu, T., Nery, J.R., Castanon, R., Ecker, J.R., 2017. Dynamic DNA methylation reconfiguration during seed development and germination. Genome Biol. 18, 171. https://doi.org/10.1186/s13059-017-1251-x

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

Kozomara, A., Birgaoanu, M., Griffiths-Jones, S., 2019b. 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

Law, J.A., Jacobsen, S.E., 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. 11, 204–220. https://doi.org/10.1038/nrg2719

Li, C., Xu, H., Fu, F.-F., Russell, S.D., Sundaresan, V., Gent, J.I., 2020a. 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., Fu, F.-F., Russell, S.D., Sundaresan, V., Gent, J.I., 2020b. 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

Li, X., Chen, L., Zhang, Q., Sun, Y., Li, Q., Yan, J., 2019. BRIF-Seq: Bisulfite-Converted Randomly Integrated Fragments Sequencing at the Single-Cell Level. Mol. Plant 12, 438–446. https://doi.org/10.1016/j.molp.2019.01.004

Lin, J.-Y., Le, B.H., Chen, M., Henry, K.F., Hur, J., Hsieh, T.-F., Chen, P.-Y., Pelletier, J.M., Pellegrini, M., Fischer, R.L., Harada, J.J., Goldberg, R.B., 2017. Similarity between soybean and *Arabidopsis* seed methylomes and loss of non-CG methylation does not affect seed development. Proc. Natl. Acad. Sci. 114, E9730–E9739. https://doi.org/10.1073/pnas.1716758114

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.

Martinez, 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
Martinez, 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
McClintock, B., 1983. THE SIGNIFICANCE OF RESPONSES OFTHE GENOME TO CHALLENGE. Nobel Lect.
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
Parent, J.-S., Cahn, J., Herridge, R.P., Grimanelli, D., Martienssen, R.A., 2021. Small RNAs guide histone methylation in *Arabidopsis* embryos. Genes Dev. 35, 841–846. https://doi.org/10.1101/gad.343871.120
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

R Core Team, 2020.
Rajkumar, M.S., Gupta, K., Khemka, N.K., Garg, R., Jain, M., 2020. DNA methylation reprogramming during seed development and its functional relevance in seed size/weight determination in chickpea. Commun. Biol. 3, 340. https://doi.org/10.1038/s42003-020-1059-1
Rodrigues, J.A., Ruan, R., Nishimura, T., Sharma, M.K., Sharma, R., Ronald, P.C., Fischer, R.L., Zilberman, D., 2013a. 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
Rodrigues, J.A., Ruan, R., Nishimura, T., Sharma, M.K., Sharma, R., Ronald, P.C., Fischer, R.L., Zilberman, D., 2013b. 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
Slotkin, R. Keith, Vaughn, M., Borges, F., Tanurdžić, M., Becker, J.D., Feijó, J.A., Martienssen, R.A., 2009. Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. Cell 136, 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

Xu, Q., Xie, W., 2018. Epigenome in Early Mammalian Development: Inheritance, Reprogramming and Establishment. Trends Cell Biol. 28, 237–253. https://doi.org/10.1016/j.tcb.2017.10.008

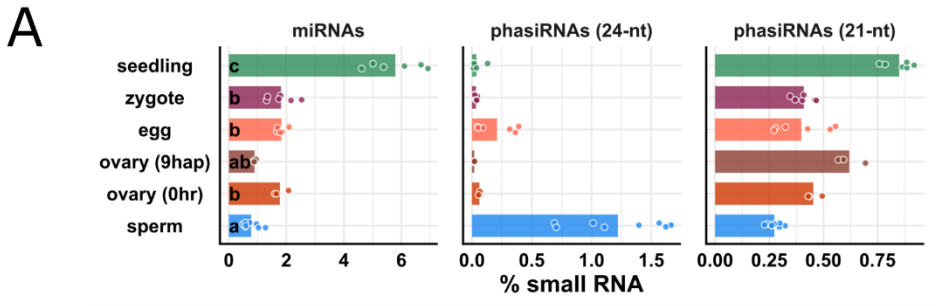
Ye, R., Wang, W., Iki, T., Liu, C., Wu, Y., Ishikawa, M., Zhou, X., Qi, Y., 2012. Cytoplasmic Assembly and Selective Nuclear Import of Arabidopsis ARGONAUTE4/siRNA Complexes. Mol. Cell 46, 859–870. https://doi.org/10.1016/j.molcel.2012.04.013

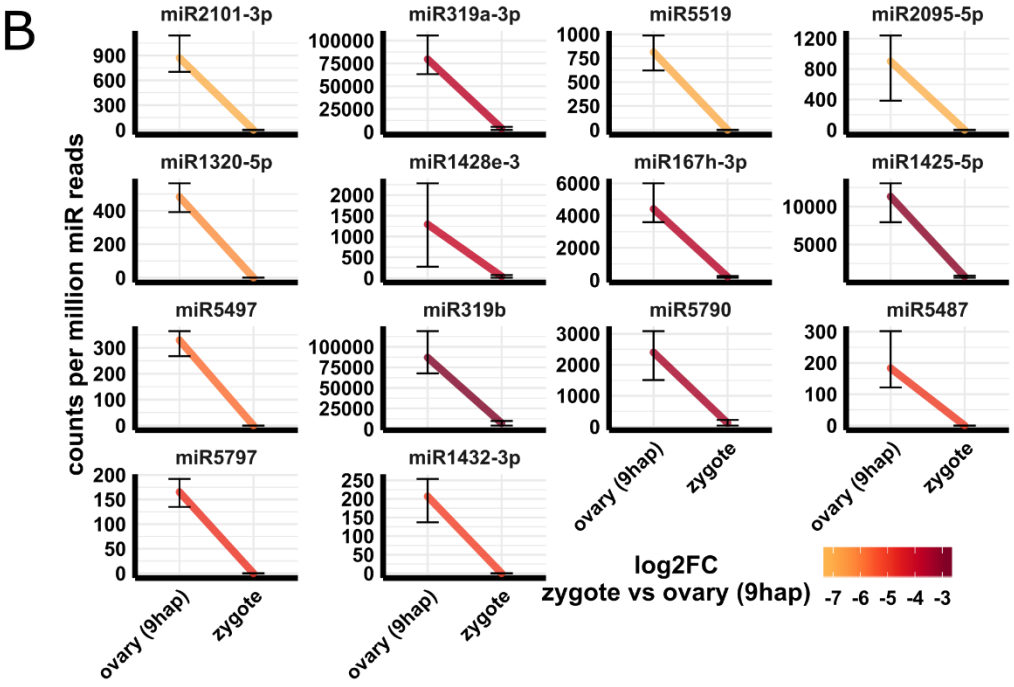
Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., Zilberman, D., 2013. The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. Cell 153, 193–205. https://doi.org/10.1016/j.cell.2013.02.033

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

Zhou, S., Li, X., Liu, Q., Zhao, Y., Jiang, W., Wu, A., Zhou, D.-X., 2021. DNA demethylases remodel DNA methylation in rice gametes and zygote and are required for reproduction. Mol. Plant S1674205221002185. https://doi.org/10.1016/j.molp.2021.06.006

Supplemental Figures





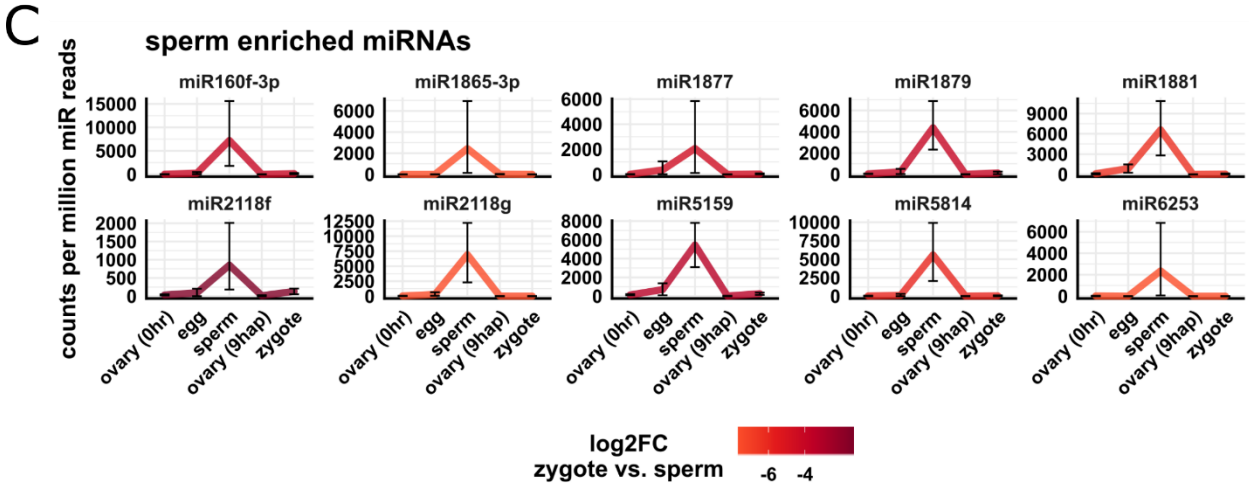


Fig S1, supporting Fig 1

(A) Relative abundances of miRNAs, 21- and 22-nt phasiRNAs in zygote small RNA transcriptomes. x-axis values are relative to total 20 – 25-nt small RNA reads. Each data point is a small RNA transcriptome. Letter grouping ($\alpha = 0.05$) is based on linear models with logit transformation followed by Tukey tests.

(B) Significantly downregulated miRNAs in zygote relative to post-fertilization ovary. Differential abundance is determined by a 2-fold decrease and FDR < 0.05 cutoffs. y-axis values are relative to per million miRNA reads in each sample. Error bars are 95% confidence intervals.

(C) Additional sperm-enriched miRNAs are downregulated in the zygote (see also **Fig 1E**). Sperm-enriched is determined by >1000 reads per million miRNA reads in sperm, but < 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 - sperm. Error bars are 95% confidence intervals. Zygote and 9 hap ovary data are from this study, all other data from Li et al., (2020).

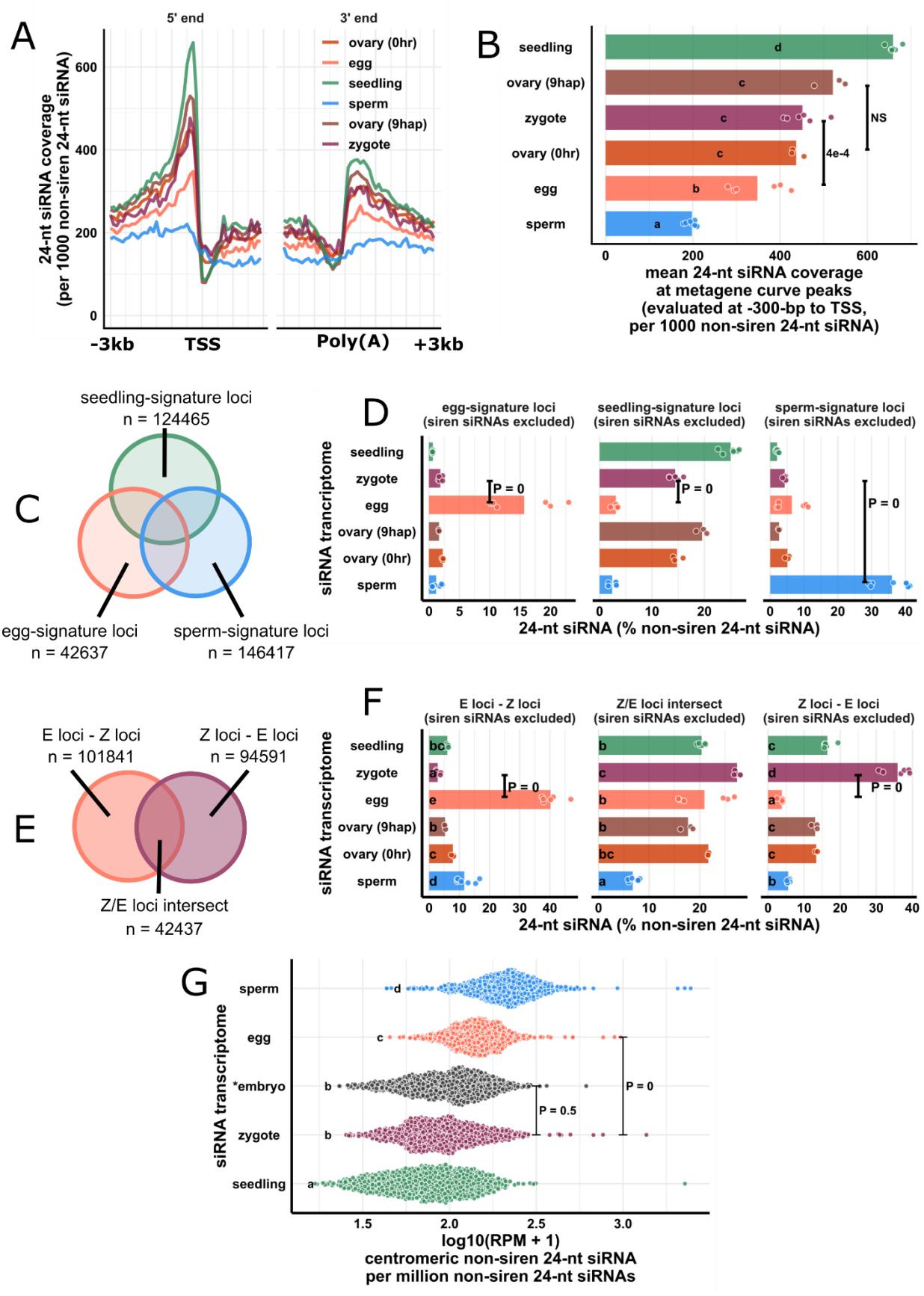

A
5' end
3' end
ovary (0hr)
egg
seedling
sperm
ovary (9hap)
zygote
24-nt siRNA coverage
(per 1000 non-siren 24-nt siRNA)
600
400
200
0
-3kb
TSS
Poly(A)
+3kb
B
seedling
ovary (9hap)
zygote
ovary (0hr)
egg
sperm
d
c
c
c
b
a
NS
4e-4
0
200
400
600
mean 24-nt siRNA coverage
at metagene curve peaks
(evaluated at -300-bp to TSS,
per 1000 non-siren 24-nt siRNA)
C
seedling-signature loci
n = 124465
egg-signature loci
n = 42637
sperm-signature loci
n = 146417
D
egg-signature loci
(siren siRNAs excluded)
seedling-signature loci
(siren siRNAs excluded)
sperm-signature loci
(siren siRNAs excluded)
siRNA trancriptome
seedling
zygote
egg
ovary (9hap)
ovary (0hr)
sperm
P = 0
P = 0
P = 0
0 5 10 15 20
0 10 20
0 10 20 30 40
24-nt siRNA (% non-siren 24-nt siRNA)
E
E loci - Z loci
n = 101841
Z loci - E loci
n = 94591
Z/E loci intersect
n = 42437
F
E loci - Z loci
(siren siRNAs excluded)
Z/E loci intersect
(siren siRNAs excluded)
Z loci - E loci
(siren siRNAs excluded)
siRNA transcriptome
seedling
zygote
egg
ovary (9hap)
ovary (0hr)
sperm
bc
a
e
b
c
d
b
c
b
b
bc
a
c
d
a
c
c
b
P = 0
P = 0
0 10 20 30 40
0 10 20
0 10 20 30 40
24-nt siRNA (% non-siren 24-nt siRNA)
G
siRNA transcriptome
sperm
egg
*embryo
zygote
seedling
d
c
b
b
a
P = 0
P = 0.5
1.5
2.0
2.5
3.0
log10(RPM + 1)
centromeric non-siren 24-nt siRNA
per million non-siren 24-nt siRNAs

Fig S2, supporting Fig 3 & Fig 4

(A) Metagene coverage plot for 24-nt non-egg-siren siRNAs. Coverage is measured over 100-bp intervals and normalized per 1000 24-nt non-egg-siren 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 non-egg-siren siRNAs.

(C) Venn diagram illustrating egg-signature loci (egg – seedling – sperm siRNA loci), seedling-signature loci (seedling – egg – sperm siRNA loci), and sperm-signature loci (sperm – egg – seedling siRNA loci), as in **Fig 3C**.

(D) Bar plot showing relative abundances of 24-nt siRNA across siRNA loci categories defined in **D**. 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 total 24-nt non-egg-siren siRNAs.

(E) Venn diagram illustrating E loci – Z loci, Z loci – E loci, and Z/E loci intersect, as in **Fig 4A**.

(F) Quantification of 24-nt siRNA relative abundances across siRNA loci categories defined in **F**. Each data point is a siRNA transcriptome. Bar heights are averages. x-axis-values are relative to total 24-nt non-egg-siren siRNAs.

(G) Quantification of non-siren 24-nt siRNA at centromeric regions. Each data point is a 50-kb window at centromeric regions across 12 rice chromosomes. x-axis values normalized to per million total non-siren 24-nt siRNAs and log10 transformed. Biological replicates were averaged prior to the analysis.

Letter grouping ($\alpha = 0.05$), and P values are based on Tukey tests. Sizes of overlap in Venn diagrams are not to scale. Zygote and 9 hap ovary data are from this study; Embryo data from Rodrigues et al. (2013); all other data from Li et al. (2020).

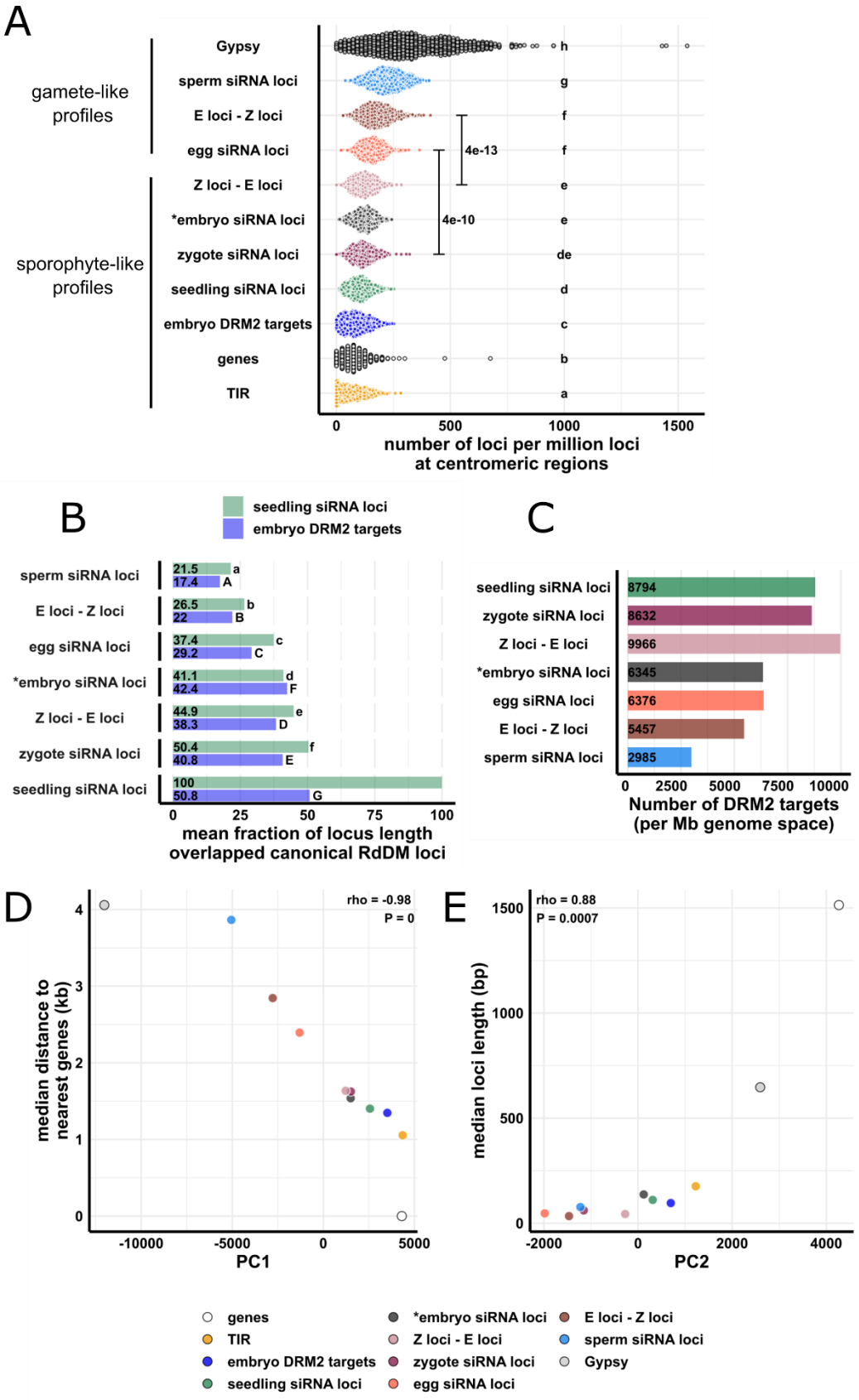

A
gamete-like profiles
sporophyte-like profiles
Gypsy
sperm siRNA loci
E loci - Z loci
egg siRNA loci
Z loci - E loci
*embryo siRNA loci
zygote siRNA loci
seedling siRNA loci
embryo DRM2 targets
genes
TIR
4e-13
4e-10
h
g
f
f
e
e
de
d
c
b
a
0
500
1000
1500
number of loci per million loci at centromeric regions
B
seedling siRNA loci
embryo DRM2 targets
sperm siRNA loci
E loci - Z loci
egg siRNA loci
*embryo siRNA loci
Z loci - E loci
zygote siRNA loci
seedling siRNA loci
21.5 a
17.4 A
26.5 b
22 B
37.4 c
29.2 C
41.1 d
42.4 F
44.9 e
38.3 D
50.4 f
40.8 E
100
50.8 G
0
25
50
75
100
mean fraction of locus length overlapped canonical RdDM loci
C
seedling siRNA loci
zygote siRNA loci
Z loci - E loci
*embryo siRNA loci
egg siRNA loci
E loci - Z loci
sperm siRNA loci
8794
8632
9966
6345
6376
5457
2985
0
2500
5000
7500
10000
Number of DRM2 targets (per Mb genome space)
D
rho = -0.98
P = 0
median distance to nearest genes (kb)
0
1
2
3
4
-10000
-5000
0
5000
PC1
E
rho = 0.88
P = 0.0007
median loci length (bp)
0
500
1000
1500
-2000
0
2000
4000
PC2
genes
TIR
embryo DRM2 targets
seedling siRNA loci
*embryo siRNA loci
Z loci - E loci
zygote siRNA loci
egg siRNA loci
E loci - Z loci
sperm siRNA loci
Gypsy

Fig S3: supporting Fig 4 and Fig 5.
(A) Quantification of **Fig 4C**. Each data point is a 50-kb genomic window inside centromeric regions for all 12 rice chromosomes. x-axis values are number of loci per million total siRNA loci for each loci category.
(B) Bar plots showing mean locus length overlapped by seedling siRNA or embryo DRM2 targets across siRNA loci categories. Statistical comparisons are made across siRNA loci categories within a locus category.
(C) Bar plots showing number of embryo DRM2 targets overlapped by different siRNA loci categories. x-axis values normalized to per million basepairs occupied by siRNA loci categories.
(D) Scatter plot showing correlation of PC1 (**Fig 4D**) and median distance to nearest genes (**Fig 5A**). Median distance to nearest genes for genes is set to 0.
(E) Scatter plot showing correlation of PC2 (**Fig 4D**) and median length of locus.
Letter grouping ($\alpha = 0.05$) and P values are based on Tukey tests. Rho, Spearman's rank order correlation coefficient. *Embryo siRNA data from Rodrigues et al (2013), which was based on a single replicate. Except zygote, all other data from Li et al. (2020).

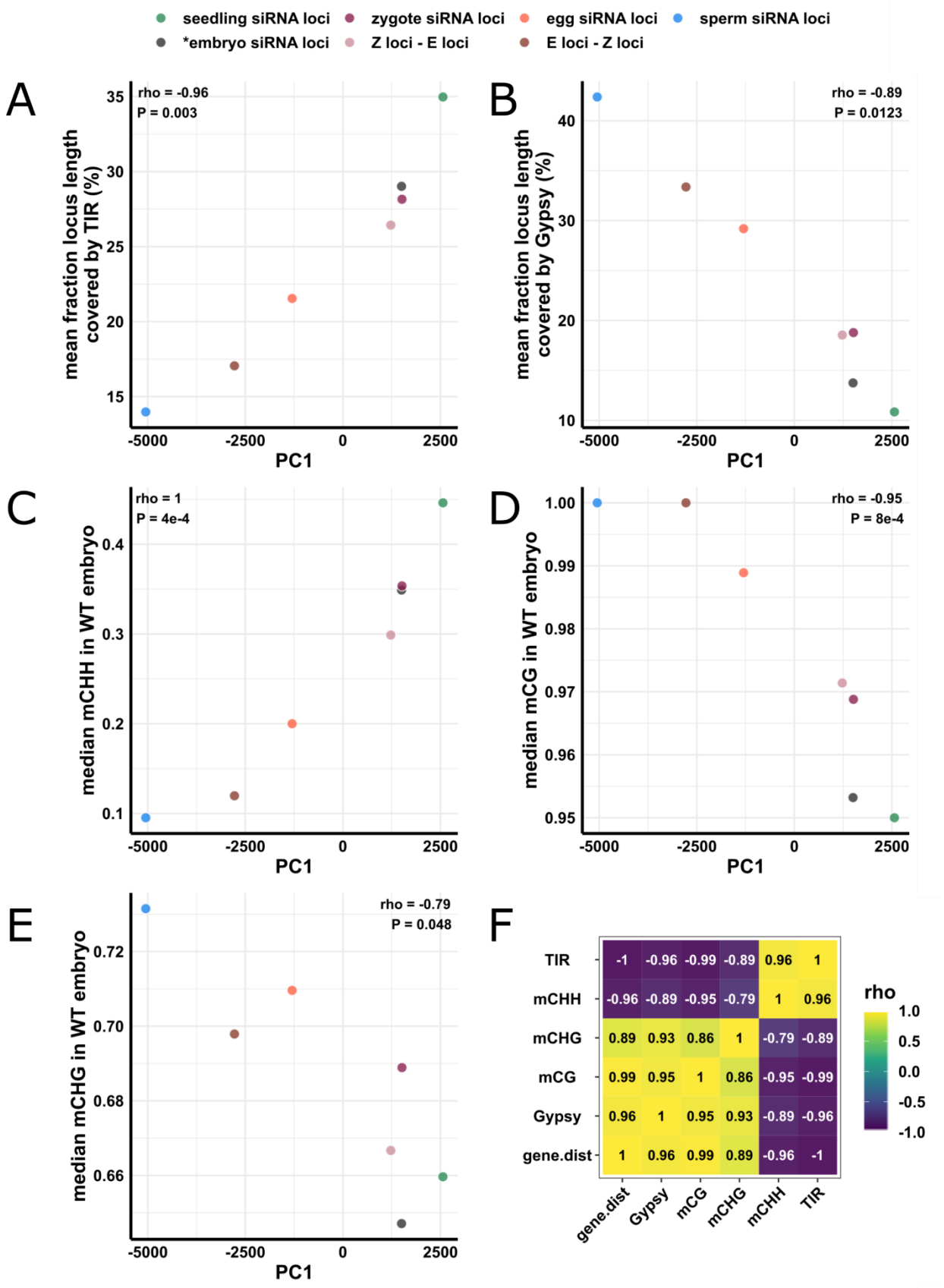

seedling siRNA loci
zygote siRNA loci
egg siRNA loci
sperm siRNA loci
*embryo siRNA loci
Z loci - E loci
E loci - Z loci
A
rho = -0.96
P = 0.003
mean fraction locus length covered by TIR (%)
PC1
B
rho = -0.89
P = 0.0123
mean fraction locus length covered by Gypsy (%)
PC1
C
rho = 1
P = 4e-4
median mCHH in WT embryo
PC1
D
rho = -0.95
P = 8e-4
median mCG in WT embryo
PC1
E
rho = -0.79
P = 0.048
median mCHG in WT embryo
PC1
F
TIR
mCHH
mCHG
mCG
Gypsy
gene.dist
rho

Fig S4, supporting Fig 4 and Fig 5

(A) Scatter plot showing correlation of PC1 (**Fig 4C**) and mean fraction of locus length covered by TIR transposon (**Fig 5B**).

(B) Scatter plot showing correlation of PC1 (**Fig 4C**) and mean fraction of locus length covered by Gypsy retrotransposon (**Fig 5B**).

(C), (D) and (E) Scatter plot showing correlation of PC1 (**Fig 4C**) and median methylation at mCHH, mCG and mCHG contexts, respectively (**Fig 5C,** see also **Fig S5**).

(F) Correlation heat map showing pairwise rank order correlation between distance to nearest genes (gene.dist, **Fig 5A**), *Gypsy* retrotransponson overlaps (**Fig 5B**), DNA methylation (**Fig 5C**, see also **Fig S5**), and TIR transposon overlaps (**Fig 5B**).

Rho, Spearman's rank order correlation coefficient. *Embryo siRNA data from Rodrigues et al (2013), which was based on a single replicate. Except zygote, all other data from Li et al. (2020).

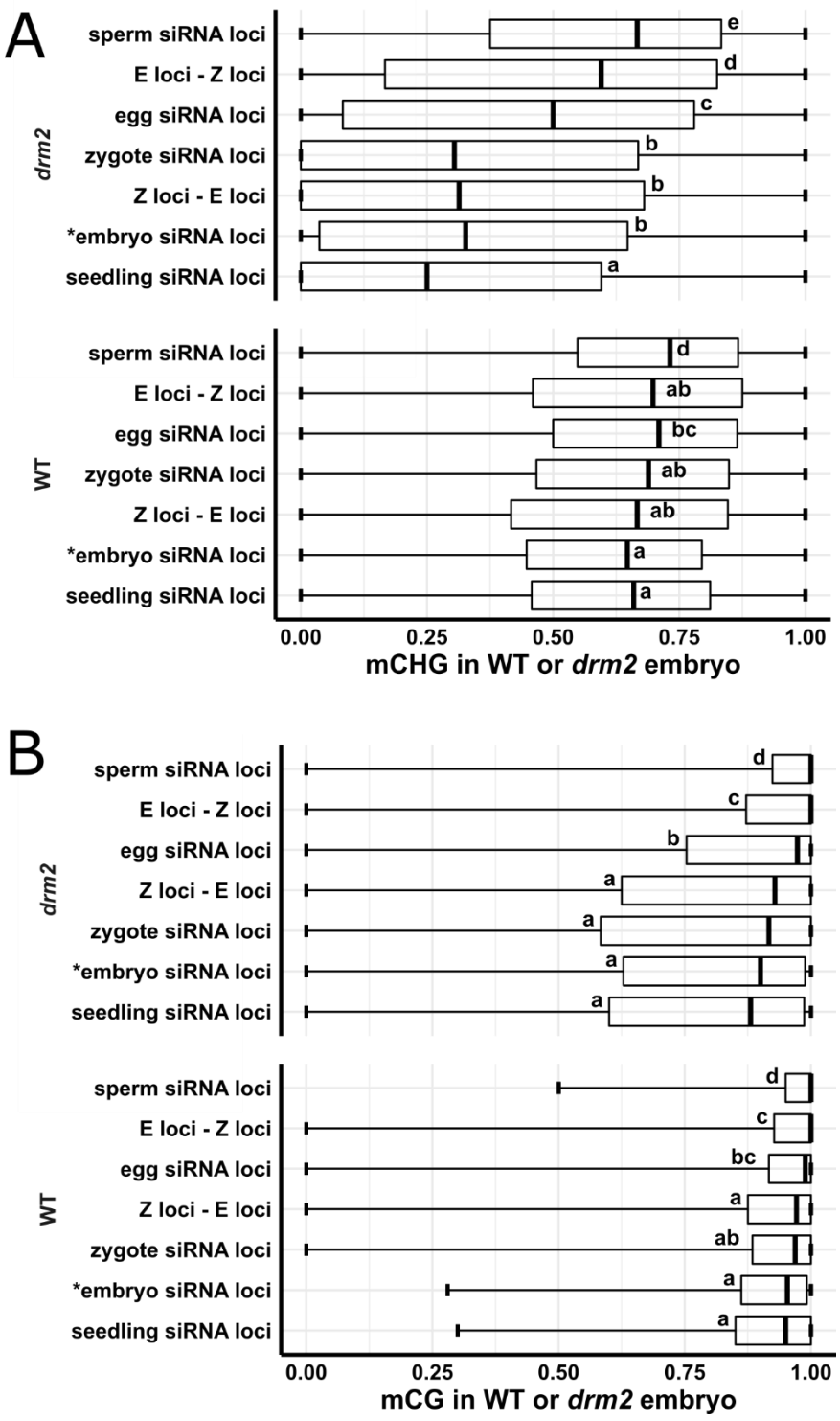


Fig S5, supporting Fig 5

Boxplots showing mCG (A) and mCHG (B) methylation level in wildtype or *drm2* embryo. Middle lines are median. Boxes span interquartile range. Whiskers span 2.5th and 97.5th percentiles. Letter groupings are based on Tukey tests. *Embryo siRNA data from Rodrigues et al (2013), which was based on a single replicate. Except zygote, all other data from Li et al. (2020).

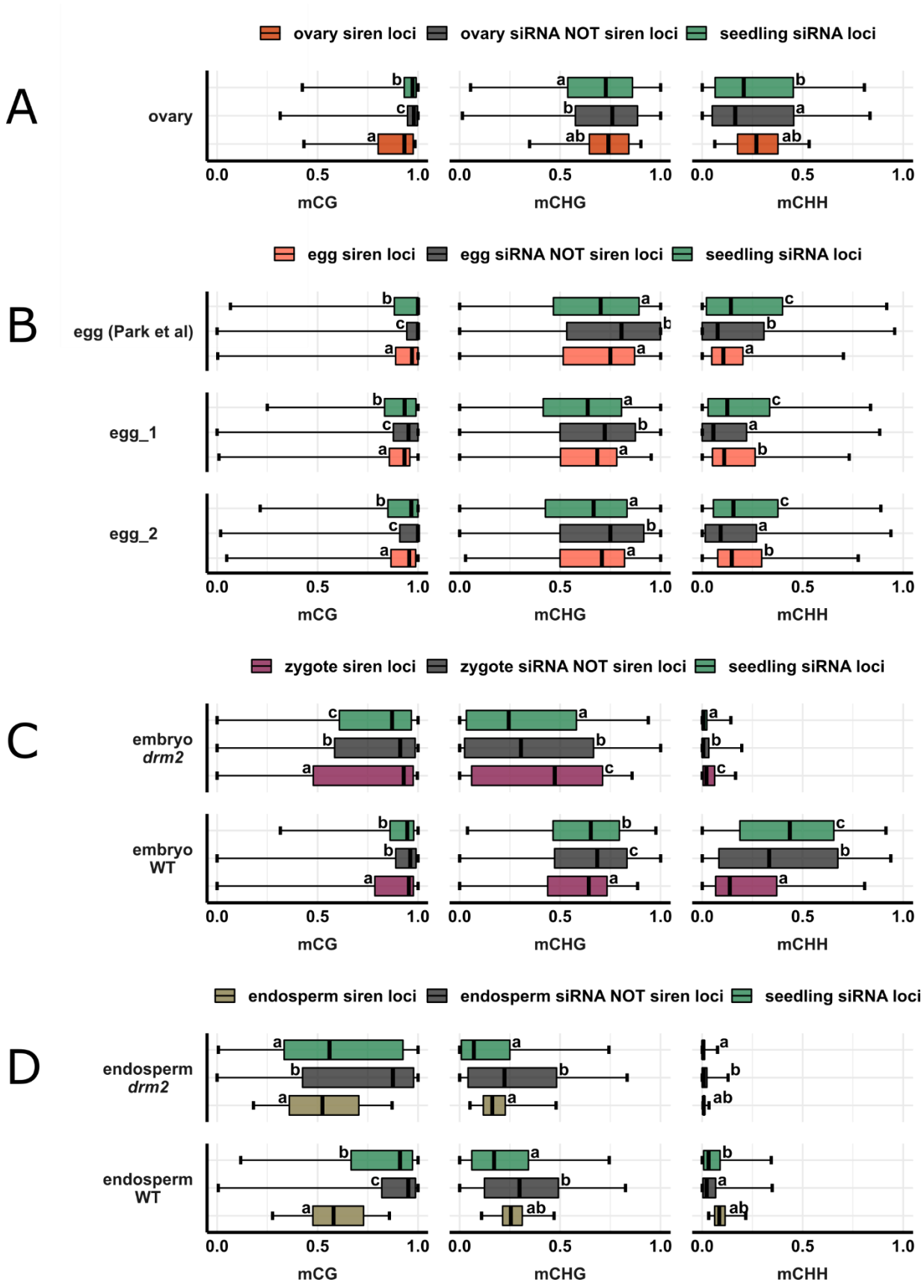

A
ovary siren loci
ovary siRNA NOT siren loci
seedling siRNA loci
ovary
mCG
mCHG
mCHH
B
egg siren loci
egg siRNA NOT siren loci
seedling siRNA loci
egg (Park et al)
egg_1
egg_2
C
zygote siren loci
zygote siRNA NOT siren loci
seedling siRNA loci
embryo drm2
embryo WT
D
endosperm siren loci
endosperm siRNA NOT siren loci
seedling siRNA loci
endosperm drm2
endosperm WT
0.0
0.5
1.0

Fig S6, supporting Fig 2 and Fig 5

Boxplots showing DNA methylation level in ovary (A), egg cell (B), wildtype or *drm2* embryo (C) and wildtype or *drm2* endosperm (D). Middle lines are median. Boxes span interquartile range. Whiskers span 2.5^{th} and 97.5^{th} percentiles. siRNA NOT siren loci refer to siRNA loci that are not siren loci. Letter groupings are based on Tukey tests. Egg (Park et al) data from Park et al., (2016). Except zygote, all other data from Li et al. (2020).

Chapter 5

Conclusion and future directions

5.1 The small RNA landscapes in gametes and their implications

We have generated the first small RNA transcriptome of egg cell from plants, as well as the first non-*Arabidopsis* sperm cell small RNA transcriptome. These data provide insights into the epigenomes of gametes.

Cereal species, including rice, have a well-studied genome organization for transposable elements (TE). Small terminal inverted repeat (TIR) transposons are enriched at gene edges, while larger transposons, such as *Gypsy* elements, are gene distal (Han, Qin, and Wessler 2013). Along with this genome organization, there is a well-characterized DNA methylation pattern, where DNA methylation in different contexts is distributed differently. CG and CHG methylation are highest at gene distal and heterochromatic regions of the genome, while CHH methylation is highest at gene-TE boundaries. Regions with high mCHH levels are termed 'mCHH islands', which is a hallmark of RNA-directed DNA methylation (RdDM) in cereals (Gent et al. 2013). In vegetative tissues, such as seedling, 24-nt siRNAs follow a particular pattern: enriched at gene-TE boundary, coincide with high mCHH/RdDM activity, and produced from TIR transposable elements (**Fig 1**). However, in gametes, 24-nt siRNAs are further away from genes and tend to be produced from *Gypsy* retrotransposons instead (**Fig 1**). Regions that gained siRNAs in gametes did not correspond to mCHH islands found in seedling, nor do they direct *de novo* mCHH in gametes (Li et al. 2020). The distinct siRNA patterns in egg *vs.* sperm suggest distinct epigenomic features in gametes.

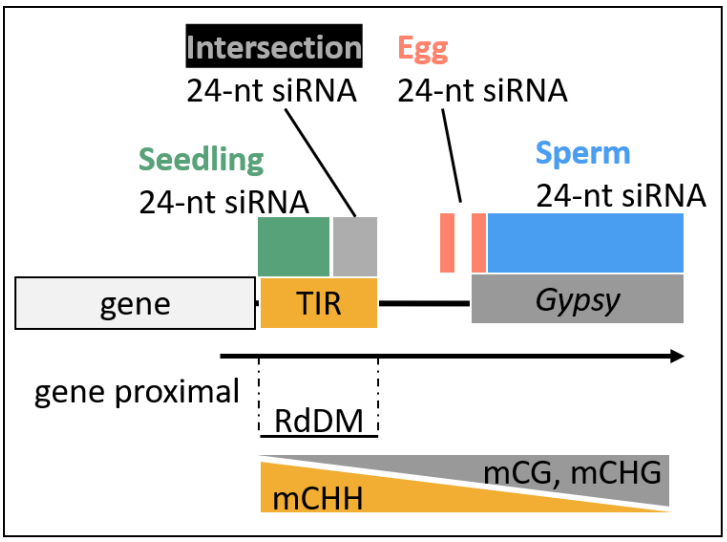


Fig 1: Graphical, non-quantitative summary of 24-nt siRNA patterns in rice seedling, egg cell, and sperm cell. Data: Genome organization (Han, Qin, and Wessler 2013); DNA methylation (Gent et al. 2013); small RNA transcriptome (Li et al. 2020).

In *Arabidopsis*, heterochromatin decondensation occurs in pollen vegetive cell (VC), which promotes the production of heterochromatic siRNAs. These pollen-derived siRNAs were reported to traffic into the sperm cells to reinforce posttranscriptional silencing (Slotkin et al. 2009; Martínez et al. 2016). An analogous pathway may be present in the rice male germline, which could explain the unique siRNA accumulation pattern in rice sperm cells. However, in *Arabidopsis*, the trafficked siRNAs are mostly 21 – 22-nt, whereas in rice most of the siRNAs are 24-nt. If trafficking were to occur in the rice male germline, it would imply that 24-nt siRNAs were trafficked. Our limited data on rice pollen VC (two biological replicates) suggest that pollen VCs also accumulate 24-nt siRNAs at centromeric regions (**Fig 2A**), much like sperm cells, which implies trafficking from VC is at least theoretically possible. However, we also detected 12 miRNAs (out of ~300 expressed miRNA genes) that were more highly expressed in pollen VC than in sperm cells (**Fig 2B**, FDR < 0.05, log2FC > 1), suggesting either trafficking

does not occur across all small RNAs, or siRNAs are produced independently in sperm cells and in pollen VC. For unknown technical reasons, I was never able to isolate high quality RNA from pollen VC. Since companion cells of gametes (VC and central cell) are not the focus of this dissertation, we decided not to pursue this further.

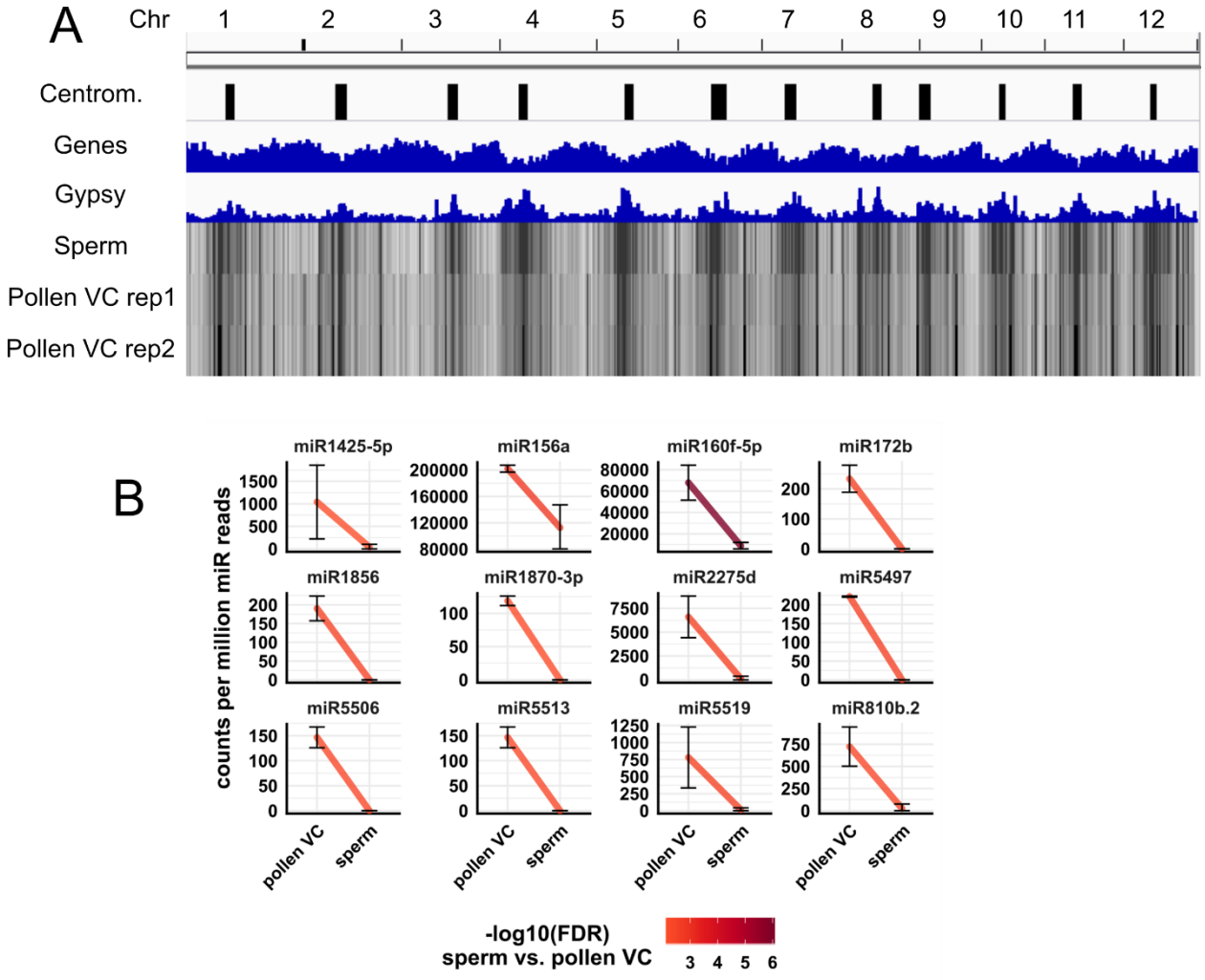


Fig 2: Pollen vegetative cell and sperm cell small RNA transcriptomes.

(A) Heat map showing abundance of 24-nt siRNAs across genome. The first three tracks are centromeric regions [as defined by (Mizuno, Matsumoto, and Wu 2018)], genes, and *Gypsy* retrotransposons. Chr. Chromosomes; Centrom. Centromeric regions.

(B) miRNAs that are significantly more abundant in pollen VC. y-axis values are relative to per million miRNA reads. Color code reflects log10(FDR) values for pollen VC vs. sperm. Error bars are 95% confidence intervals for each cell type. VC: vegetative cell.

The lack of correspondence between 24-nt siRNAs and mCHH in gametes raised two questions: 1) Why does the mCHH pattern not match 24-nt siRNAs in gametes, given our prior understanding that 24-nt siRNAs function in RdDM? 2) What are the functions of siRNAs in gametes if they are not associated with RdDM? I have a few speculations that might serve as potential answers to these questions.

First, it is predicted that rice and maize gametes are arrested at G1 of cell cycle, since in rice and maize, S phase and G2 phase genes are rapidly upregulated after fertilization (Anderson et al. 2017; Chen et al. 2017). These results indicate the absence of DNA synthesis in gametes. Unmethylated cytosines are incorporated into DNA during replication; therefore, loss of mCHH islands due to lack of RdDM requires that the cells undergo S phase. Thus, the absence of S phase in gametes could explain why the lack of gene-proximal siRNAs does not lead to the lack of mCHH islands in gametes.

Second, the gain of siRNAs at gene-distal regions of the genome in gametes does not necessarily invoke RdDM. DRM2 is the DNA methyltransferase required for RdDM [reviewed in (Law and Jacobsen 2010)]. It is likely that, for yet unknown reasons, DRM2 may have no activity at gene-distal regions of the genome. The full RdDM pathway is expressed in both rice gametes (Anderson et al. 2013). The lack of DRM2 activity at gene-distal regions may be due to either the lack of chromatin accessibility for DRM2, or due to the lack of associated factors that facilitate DRM2 activity. The above speculations are not mutually exclusive. DRM2 require a

nascent transcript of RNA polymerase V (Pol V) as a scaffold [reviewed in(Matzke and Mosher 2014)], and transcription generally require a certain level of chromatin accessibility. The lack of accessibility and the lack of Pol V transcription can both contribute to the lack of DRM2 activity. The above speculations could explain why the gain of siRNAs at heterochromatic gamete-specific siRNA loci did not produce high CHH methylation at these loci in gametes.

The potential functions of siRNAs in gametes are largely unexplored. Since 24-nt siRNAs in gametes did not appear to target RdDM in gametes, one could hypothesize that siRNAs in gametes target chromatin modifications rather than DNA methylation. For example, in fission yeast (*S. pombe*), siRNAs can target histone 3 methylation at lysine 9 (H3K9me) via nascent Pol II transcripts as scaffolds and histone methyltransferases [reviewed in (Zentner and Henikoff 2013)]. Such a process does not involve DNA methylation. It is possible that analogous pathways function in plant gametes.

Lastly, since siRNAs are an output of the epigenome, siRNAs may be a by-product of chromatin changes in the gametes or surrounding cells (Wang and Köhler 2017; Borg and Berger 2015; Borg et al. 2020; Ingouff et al. 2010; 2017). There is a possibility that siRNAs themselves may not serve biological functions in gametes.

5.2 The zygote has initiated a reset to the canonical siRNA transcriptome before the first embryonic division

In addition to small RNA landscapes of gametes, we also generated the first zygote small RNA landscape from plants. We were curious how the siRNA transcriptome has changed in the zygote relative to the gametes, since in angiosperms zygotic genome activation (ZGA) is already underway before the zygote divides (Anderson et al. 2017; Chen et al. 2017; Zhao et al. 2019).

The fact that the siRNA patterns in gametes are very distinct from each other makes this question even more interesting. We collected zygotes at 9 hours after pollination (9 hap), corresponding to the completion of S phase and a major wave of ZGA in rice (Anderson et al. 2017).

Our data (Chapter 4) revealed that the zygote inherit maternal small RNAs, while lack paternal siRNA signatures. This observation has important implications for the current models of post-fertilization silencing through the male germline. Any effects of sperm-transmitted siRNAs on embryo are likely to be indirect, likely via chromatin modifications and not via siRNAs themselves.

A large number of 24-nt siRNA loci were newly detected in zygote, and they have interesting properties: gene proximal, tendency to overlap with TIR transposons, and associated with RdDM in embryogenesis (**Fig 3**). These observations suggest the zygote has initiated a reset to the canonical siRNA pattern detected in seedling and other vegetative tissues (e.g., leaves and root). Thus, we conclude that resetting of other aspects of the epigenome (DNA methylation, histone modifications, and chromatin conformation) may be initiated in plant zygotes as well.

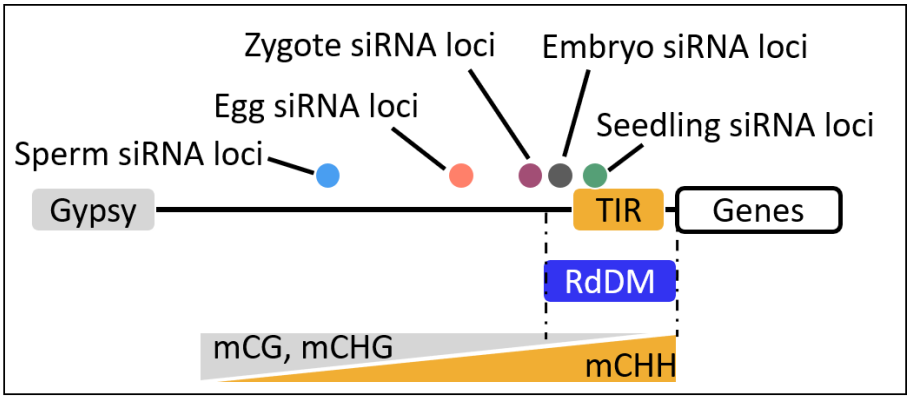


Fig 3: Graphical, non-quantitative summary of 24-nt siRNA patterns in rice egg cell, sperm cell, embryo, and zygote. Data: seedling and gametes (Li et al. 2020); 7-8 days after fertilization embryo (Rodrigues et al. 2013); zygote (Chapter 4).

5.3 Future directions

An important aspect of gametes and zygotes that has not been thoroughly explored is histone modifications. Borg et al. (2020) reported that the *Arabidopsis* sperm cell lack the repressive histone mark H3K27me3, and key genes for sperm differentiation and embryogenesis are marked with an activating histone mark H3K4me. Histone modifications in egg cell and zygote remained poorly characterized due to technical difficulties associated with isolating these exceptional cell types. ChIP-seq is the gold standard for chromatin studies, which typically requires at least 10^5 cells for sequencing library preparation. Recent advances in technology, such as CUT&Tag ('Cleavage Under Targets and Tagmentation'), have reduced the minimal input requirement to 100 – 1000 cells (Kaya-Okur et al. 2019), making profiling histone modifications in egg and zygote not a technical impossibility. However, since egg cell and zygote are collected by manual isolation (Chen et al. 2017; Li et al. 2019; Zhou et al. 2019; Zhao et al. 2019), 100 – 1000 cells is still a daunting number. Efficient, high-throughput methods for egg and zygote isolation must be established before the application of low-input ChIP-seq alternatives. Methods for isolating selected cell types based on cell type specific markers have been developed, e.g., INTACT (Deal and Henikoff 2011, 'Isolation of Nuclei Tagged in Specific Cell Types'), FACS, and FANS (fluorescent activated cell/nuclei sorting). Since egg cell and zygote are lowly represented (< 1/1000 cells in an ovule), the above methods need to be modified to increase sensitive and efficiency.

Recent advances in single cell sequencing technologies opened new opportunities for egg cell and zygote. 'Single cell' methods refer to barcoding individual cells/nuclei in a sample, such that after sequencing and demultiplexing, the data from each cell/nucleus are resolved. Thus,

current ‘single cell’ methods do not produce data from *a single cell*, instead from many barcoded cells or nuclei. Since egg cell and zygote are lowly represented in the ovule, likely 1/1000 or less of the sequencing reads from a single cell sequencing run will be attributed to egg cells or zygotes, making single cell sequencing technologies less useful for egg cell and zygote. However, methods can be developed to enrich for egg cell and zygote (e.g., via INTACT, **Fig 4**). Egg- or zygote-enriched samples might be suitable for single cell sequencing. Egg and zygotes can be bioinformatically distinguished from ovule cells using marker genes and clustering since the data for individual cell/nuclei are resolved (**Fig 4**). These experiments will answer questions regarding whether siRNAs in gametes direct histone modifications, and whether histone modifications initiate a reset to the sporophytic pattern in zygote.

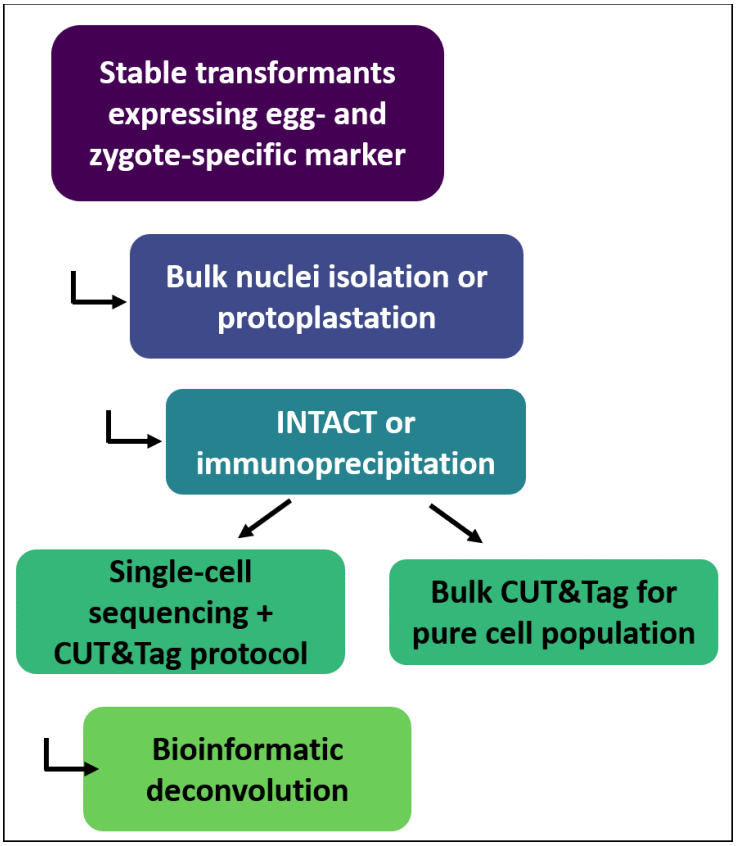


Fig 4: proposed workflow to enable genome-wide profiling of histone modifications in egg cell and zygote.

As more omics data are gathered for gametes and zygotes, the power to discover regulators of ZGA and pluripotency increases. For example, BBM1 is an important pluripotency factor, which is not detected in the egg cell but expressed in the zygote and embryo. When *BBM1* is ectopically expressed in the egg cell, it bypasses the requirement for fertilization and directly converts the egg cell into a haploid embryo (Khanday et al. 2019). Despite its importance in ZGA and pluripotency induction, very little is known about how *BBM1* is regulated, and the cis-regulatory sequences (CRSs) conferring its expression pattern have yet to be characterized. Similar examples include but are not limited to WUS-like homeobox (WOX) transcription factors (Lowe et al. 2016; Anderson et al. 2017; Haecker 2004). Epigenomic and transcriptomic datasets might elucidate the regulation of these key developmental regulators (**Fig 5**).

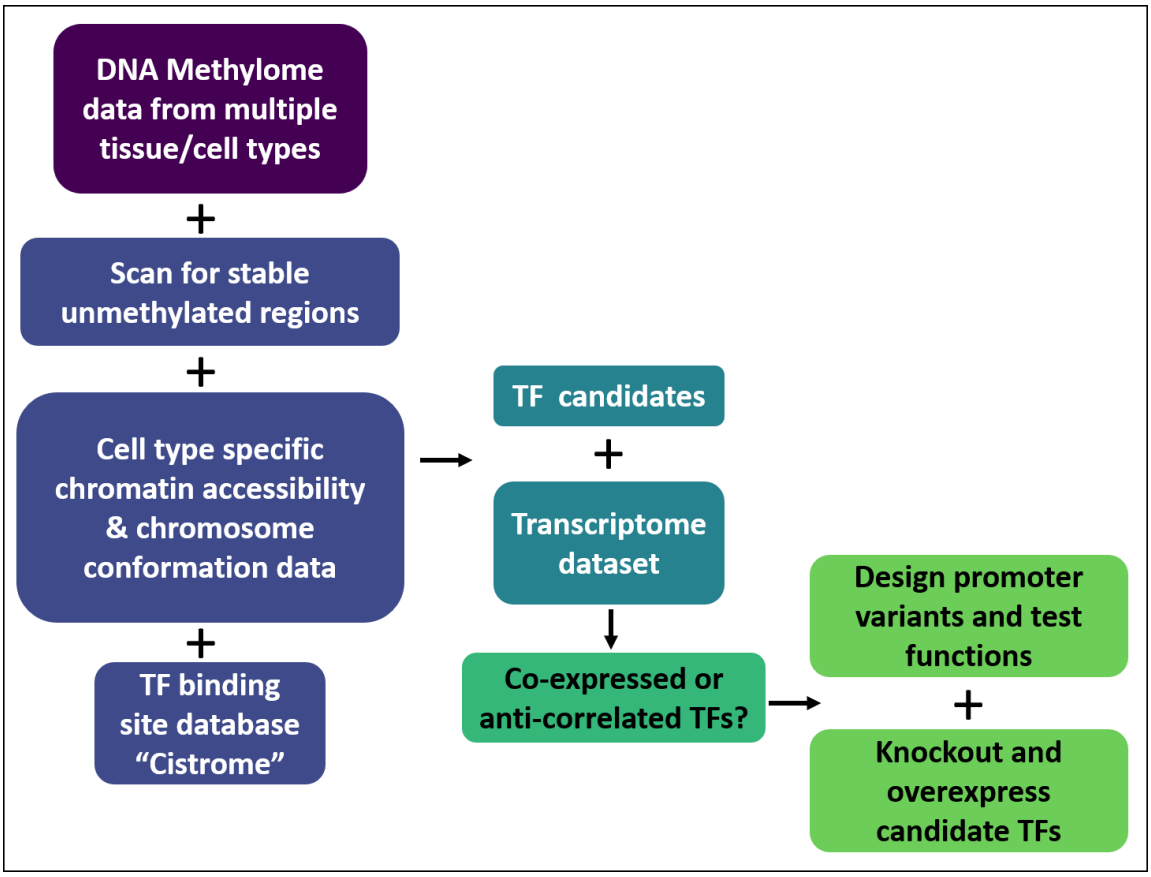


Fig 5: Proposed workflow to probe regulators and cis-regulatory regions for pluripotency factors and developmental regulators using omics datasets.

It has been reported in maize that stable unmethylated DNA marks regulatory sequences (Crisp et al. 2020), and that accessible chromatin regions have low levels of DNA methylation relative to their flanking regions (Marand et al. 2021). Thus, stable unmethylated regions (SUMRs) likely contain CRSs. Cell type specific chromatin accessibility data (from ATAC-seq) and chromosome contact data (from Hi-C) can inform if a putative CRS is accessible in a cell type of interest, and if chromosome contacts are formed between a distal CRS and a core promoter. Transcription factor (TF) binding site datasets [from ChIP-seq and similar experiments, also termed 'cistrome' (O'Malley et al. 2016)] can be used to detect TF binding sites within SUMRs. Gene expression alas (from RNA-seq) can be used to explore if there are co-expressed TFs that also bind within SUMRs, which may indicate a regulatory role. Candidate TFs and enhancers can then be subjected to reverse genetics experiments to test if the disruption of such TFs and enhancers leads to misexpression of genes of interest (**Fig 5**). These experiments may shed light on the regulation of pluripotency factors and developmental regulators (e.g., *BBM1*), which have broad agricultural and biotechnological applications (Khanday et al. 2019).

Lastly, I propose that epigenomic data will enable genome-wide, systematic annotation of CRSs in plants, since CRSs tend to contain certain signatures, such as low DNA methylation, accessible in cell types where they are active, enriched for histone modifications (e.g., lysine acetylation and H3K4me for activation, H3K27me3 for repression), and bound by TFs. All the above features are informative for the discovery and annotation of CRSs. Coding sequences are relatively easy to annotate due the conservation of gene products, whereas CRSs annotation

remains challenging, because they are short, dispersed, and oftentimes function differently between species. Forward and reverse genetics on model plants (mainly *Arabidopsis*, but also maize, rice, and tomato) has characterized many key developmental regulators. However, while their promoters have been characterized in some cases, overall, we understand considerably less about CRSs regulating these developmental regulations. Epigenome-enabled discovery and annotation (**Fig 5**) could enable precise trait tailoring (e.g., flowering time, branching, organ size, etc.), which could enable fine tuning of crops to suit a particular farm site or environment. For example, one may take a multi-stress-resilient wild crop relative, gather genomic, epigenomic and transcriptomic data, identify candidate developmental regulators and their CRSs, and rapidly fine tailor this wild plant into a high yielding crop using genome editing. We do not have such power now, but I suggest intellectually this is not impossible.

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 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.

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.

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.

Crisp, Peter A., Alexandre P. Marand, Jaclyn M. Noshay, Peng Zhou, Zefu Lu, Robert J. Schmitz, and Nathan M. Springer. 2020. "Stable Unmethylated DNA Demarcates Expressed Genes and Their

Cis-Regulatory Space in Plant Genomes." *Proceedings of the National Academy of Sciences* 117 (38): 23991–0. https://doi.org/10.1073/pnas.2010250117.

Deal, Roger B, and Steven Henikoff. 2011. "The INTACT Method for Cell Type–Specific Gene Expression and Chromatin Profiling in Arabidopsis Thaliana." *Nature Protocols* 6 (1): 56–68. https://doi.org/10.1038/nprot.2010.175.

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.

Haecker, A. 2004. "Expression Dynamics of WOX Genes Mark Cell Fate Decisions during Early Embryonic Patterning in Arabidopsis Thaliana." *Development* 131 (3): 657–68. https://doi.org/10.1242/dev.00963.

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.

Ingouff, Mathieu, Svenja Rademacher, Sarah Holec, Lucija Šoljić, Nie Xin, Anne Readshaw, Shi Hui Foo, Benoît Lahouze, Stefanie Sprunck, and Frédéric Berger. 2010. "Zygotic Resetting of the HISTONE 3 Variant Repertoire Participates in Epigenetic Reprogramming in Arabidopsis." *Current Biology* 20 (23): 2137–43. https://doi.org/10.1016/j.cub.2010.11.012.

Ingouff, Mathieu, Benjamin Selles, Caroline Michaud, Thiet M. Vu, Frédéric Berger, Andrea J. Schorn, Daphné Autran, et al. 2017. "Live-Cell Analysis of DNA Methylation during Sexual Reproduction in *Arabidopsis* Reveals Context and Sex-Specific Dynamics Controlled by Noncanonical RdDM." *Genes & Development* 31 (1): 72–83. https://doi.org/10.1101/gad.289397.116.

Kaya-Okur, Hatice S., Steven J. Wu, Christine A. Codomo, Erica S. Pledger, Terri D. Bryson, Jorja G. Henikoff, Kami Ahmad, and Steven Henikoff. 2019. "CUT&Tag for Efficient Epigenomic Profiling of Small Samples and Single Cells." *Nature Communications* 10 (1): 1930. https://doi.org/10.1038/s41467-019-09982-5.

Khanday, Imtiyaz, Debra Skinner, Bing Yang, Raphael Mercier, and Venkatesan Sundaresan. 2019. "A Male-Expressed Rice Embryogenic Trigger Redirected for Asexual Propagation through Seeds." *Nature* 565 (7737): 91–95. https://doi.org/10.1038/s41586-018-0785-8.

Law, Julie A., and Steven E. Jacobsen. 2010. "Establishing, Maintaining and Modifying DNA Methylation Patterns in Plants and Animals." *Nature Reviews Genetics* 11 (3): 204–20. https://doi.org/10.1038/nrg2719.

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.

Lowe, Keith, Emily Wu, Ning Wang, George Hoerster, Craig Hastings, Myeong-Je Cho, Chris Scelonge, et al. 2016. "Morphogenic Regulators *Baby Boom* and *Wuschel* Improve Monocot Transformation." *The Plant Cell* 28 (9): 1998–2015. https://doi.org/10.1105/tpc.16.00124.

Marand, Alexandre P., Zongliang Chen, Andrea Gallavotti, and Robert J. Schmitz. 2021. "A Cis-Regulatory Atlas in Maize at Single-Cell Resolution." *Cell* 184 (11): 3041-3055.e21. https://doi.org/10.1016/j.cell.2021.04.014.

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): 16030. https://doi.org/10.1038/nplants.2016.30.

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.

Mizuno, Hiroshi, Takashi Matsumoto, and Jianzhong Wu. 2018. “Composition and Structure of Rice Centromeres and Telomeres.” In *Rice Genomics, Genetics and Breeding*, edited by Takuji Sasaki and Motoyuki Ashikari, 37–52. Singapore: Springer Singapore. https://doi.org/10.1007/978-981-10-7461-5_3.

O’Malley, Ronan C., Shao-shan Carol Huang, Liang Song, Mathew G. Lewsey, Anna Bartlett, Joseph R. Nery, Mary Galli, Andrea Gallavotti, and Joseph R. Ecker. 2016. “Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape.” *Cell* 165 (5): 1280–92. https://doi.org/10.1016/j.cell.2016.04.038.

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.

Slotkin, R. Keith, Matthew Vaughn, Filipe Borges, Miloš Tanurdžić, Jörg D. Becker, José A. Feijó, and Robert A. Martienssen. 2009. “Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen.” *Cell* 136 (3): 461–72. https://doi.org/10.1016/j.cell.2008.12.038.

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.

Zentner, Gabriel E, and Steven Henikoff. 2013. “Regulation of Nucleosome Dynamics by Histone Modifications.” *Nature Structural & Molecular Biology* 20 (3): 259–66. https://doi.org/10.1038/nsmb.2470.

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, Xuemei, Ce Shi, Peng Zhao, and Mengxiang Sun. 2019. “Isolation of Living Apical and Basal Cell Lineages of Early Proembryos for Transcriptome Analysis.” *Plant Reproduction* 32 (1): 105–11. https://doi.org/10.1007/s00497-018-00353-6.