Supplemental Figures and Legends

Table of Contents

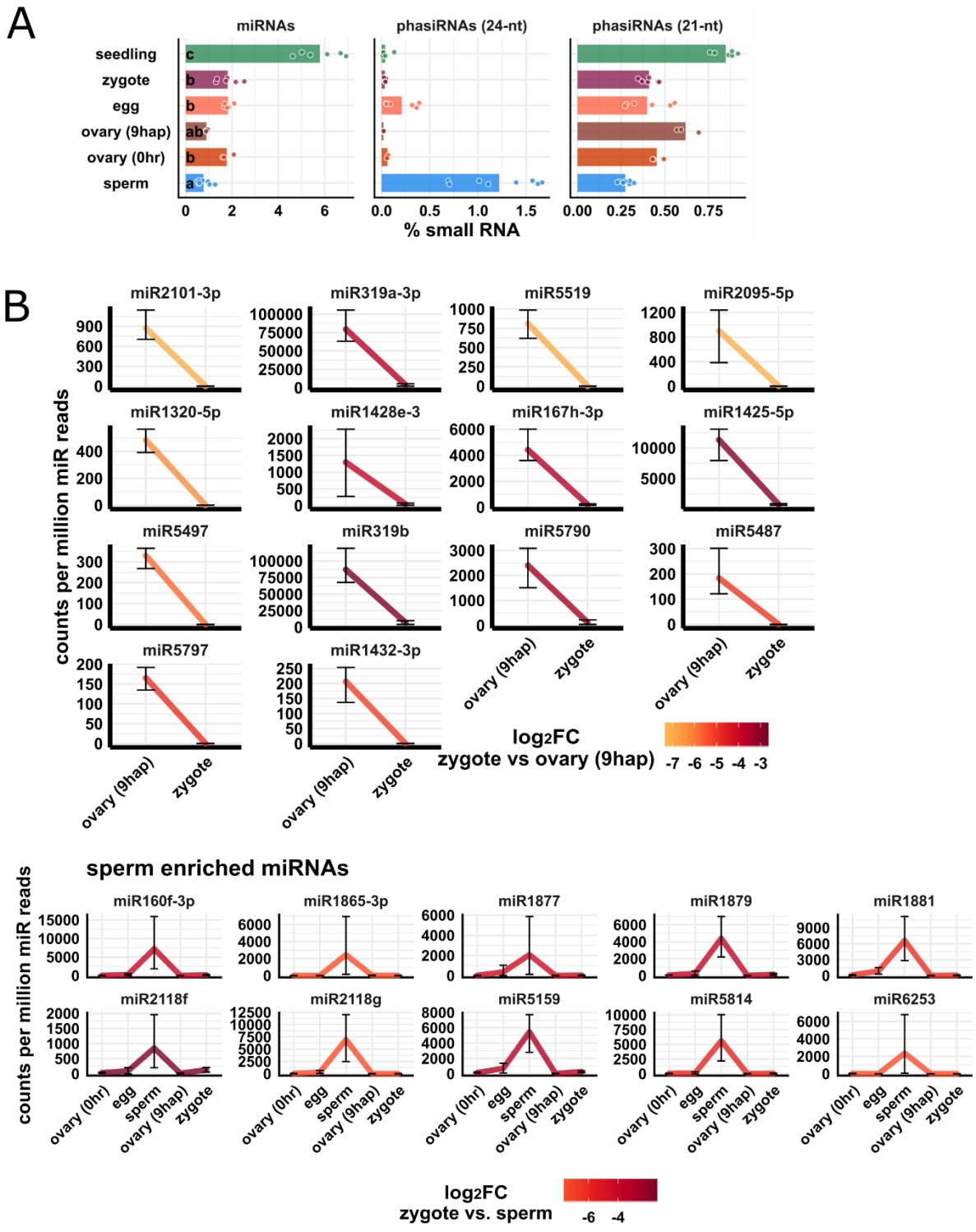

A
miRNAs
phasiRNAs (24-nt)
phasiRNAs (21-nt)
seedling
zygote
egg
ovary (9hap)
ovary (0hr)
sperm
c
b
b
ab
b
a
% small RNA
B
miR2101-3p
miR319a-3p
miR5519
miR2095-5p
miR1320-5p
miR1428e-3
miR167h-3p
miR1425-5p
miR5497
miR319b
miR5790
miR5487
miR5797
miR1432-3p
counts per million miR reads
ovary (9hap)
zygote
log2FC
zygote vs ovary (9hap)
C
sperm enriched miRNAs
miR160f-3p
miR1865-3p
miR1877
miR1879
miR1881
miR2118f
miR2118g
miR5159
miR5814
miR6253
counts per million miR reads
ovary (0hr)
egg
sperm
ovary (9hap)
zygote
log2FC
zygote vs. sperm

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. Color code reflects $\log_2$FC values for zygote - ovary. 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 $\log_2$FC 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).

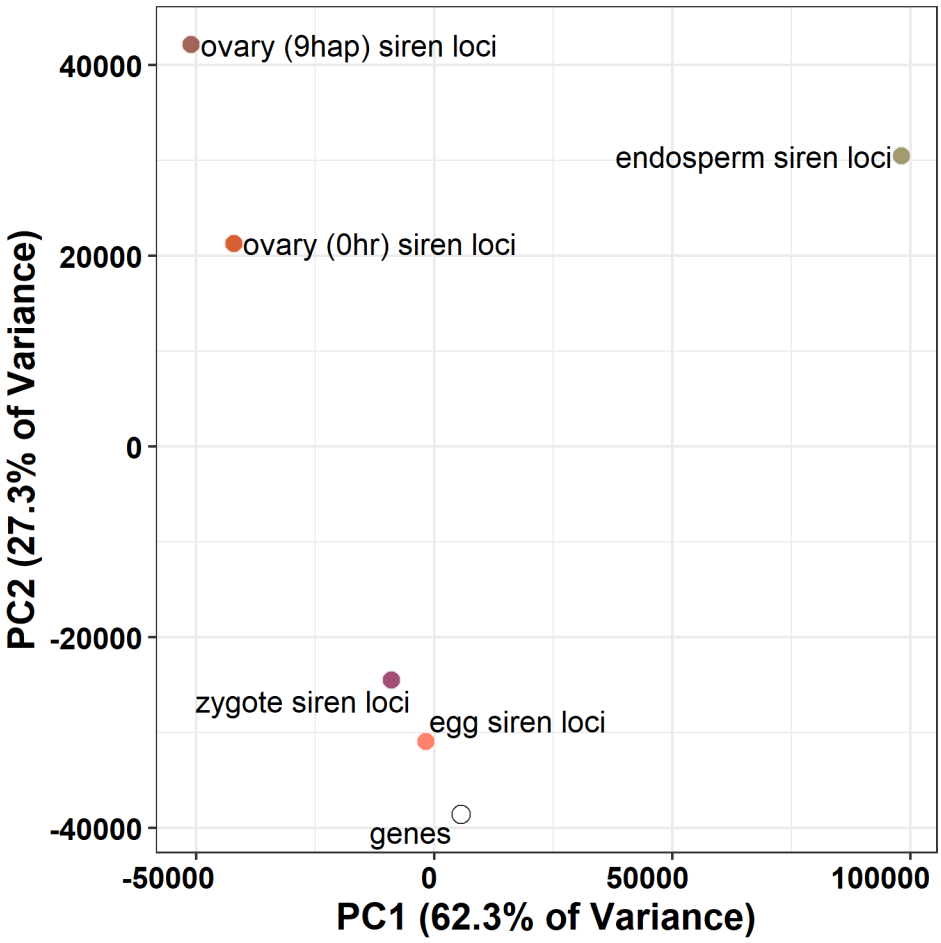


Fig S2, supporting Fig 2
Principal component plot for siren loci distribution across the genome. Distributions are evaluated at 50-kb resolution across the genome. Each data point is the distribution of a siren loci category.

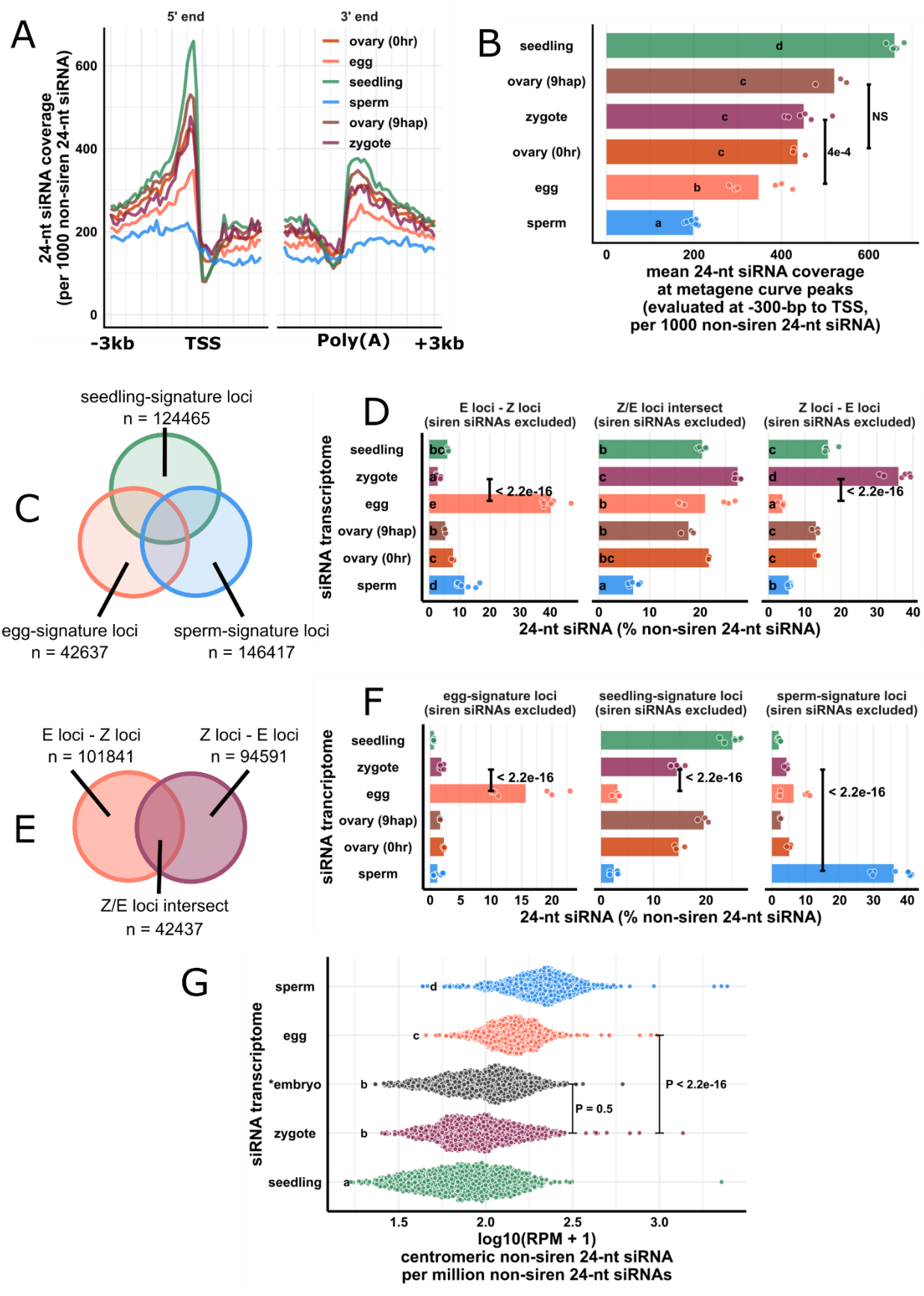

A
5' end
3' end
ovary (0hr)
egg
seedling
sperm
ovary (9hap)
zygote
24-nt siRNA coverage
(per 1000 non-siren 24-nt siRNA)
600
400
200
0
-3kb
TSS
Poly(A)
+3kb
B
seedling
ovary (9hap)
zygote
ovary (0hr)
egg
sperm
d
c
c
c
b
a
NS
4e-4
0
200
400
600
mean 24-nt siRNA coverage
at metagene curve peaks
(evaluated 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
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
< 2.2e-16
< 2.2e-16
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
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
< 2.2e-16
< 2.2e-16
< 2.2e-16
0
5
10
15
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 < 2.2e-16
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 S3, supporting Fig 3 and 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 siRNAs 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).

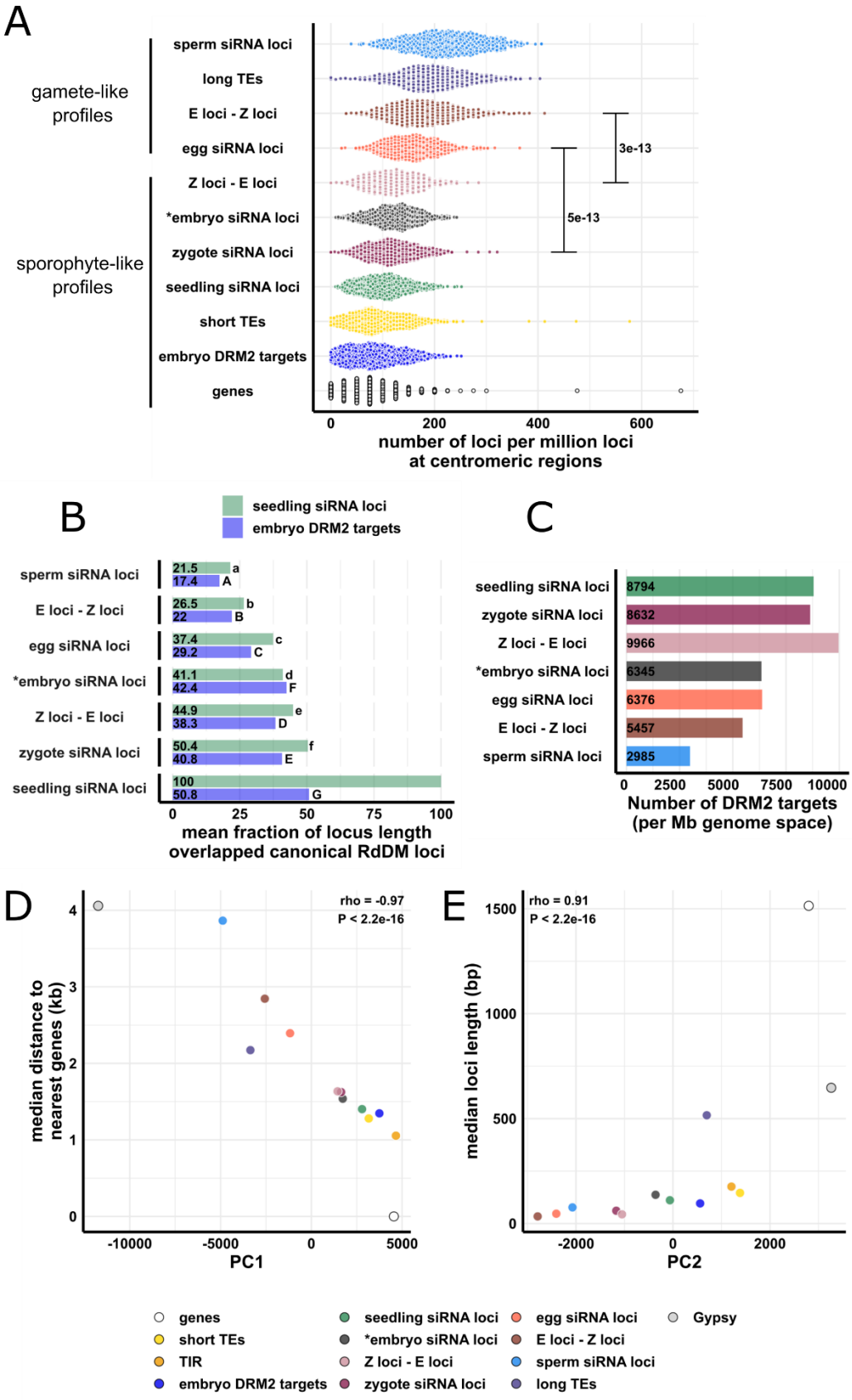

A
gamete-like profiles
sporophyte-like profiles
sperm siRNA loci
long TEs
E loci - Z loci
egg siRNA loci
Z loci - E loci
*embryo siRNA loci
zygote siRNA loci
seedling siRNA loci
short TEs
embryo DRM2 targets
genes
3e-13
5e-13
0
200
400
600
number of loci per million loci at centromeric regions
B
seedling 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
17.4
26.5
22
37.4
29.2
41.1
42.4
44.9
38.3
50.4
40.8
100
50.8
a
A
b
B
c
C
d
F
e
D
f
E
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 = -0.97
P < 2.2e-16
median distance to nearest genes (kb)
0
1
2
3
4
-10000
-5000
0
5000
PC1
E
rho = 0.91
P < 2.2e-16
median loci length (bp)
0
500
1000
1500
-2000
0
2000
PC2
genes
short TEs
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
long TEs
Gypsy

Fig S4: 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 base pairs 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. Short TEs are transposable elements of length 50-bp to 250-bp. Long TEs are transposable elements of length greater than 250-bp. *Embryo (7 days after flowering) siRNA data from Rodrigues et al (2013), which was based on a single replicate. Except zygote, all other data from Li et al. (2020).

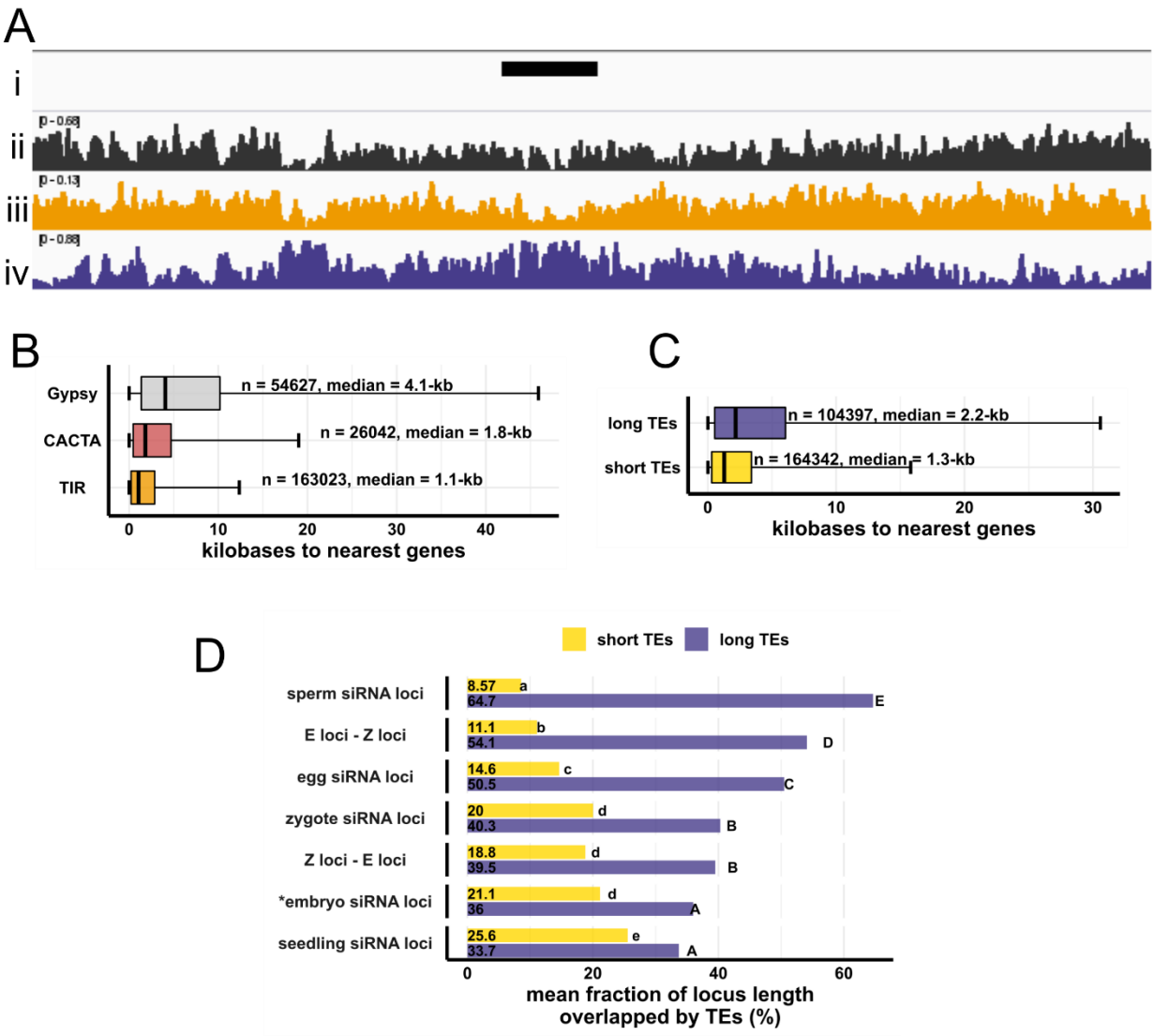


Fig S5: supporting Fig 4 and Fig 5

(A) Distribution of short TEs (50-250-bp) and long TEs (> 250-bp) along chromosome 8. i: centromeric region; ii: non-TE genes; iii: short TEs; iv: long TEs. Chromosome 8 is chosen because it is one of the chromosomes with a completed sequenced centromeric region (Mizuno et al., 2018).

(B) and (C) box plots showing distance of TEs to nearest genes. Middle lines are median. Boxes span interquartile range. Whiskers span 2.5th and 97.5th percentiles. TIR: non-CACTA TIR elements, CACTA: CACTA DNA transposons; *Gypsy*: *Gypsy* retrotransposons; short TEs: TEs of lengths 50-250-bp; long TEs: TEs with length > 250-bp.

(D) Bar plots showing mean locus length overlapped by short TEs or long TEs across siRNA loci categories. Statistical comparisons are made across siRNA loci categories within a TE length category.

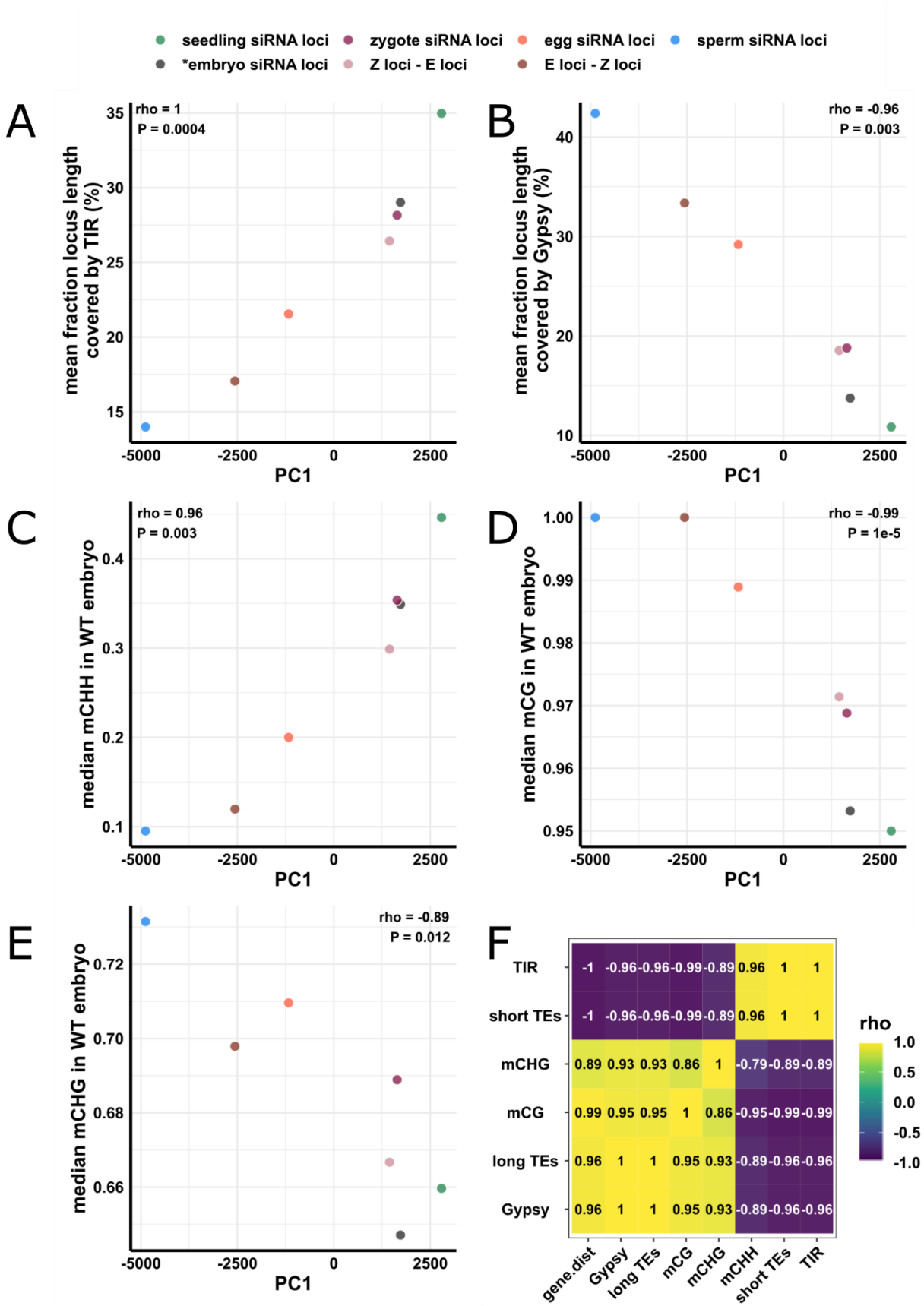

seedling siRNA loci
zygote siRNA loci
egg siRNA loci
sperm siRNA loci
*embryo siRNA loci
Z loci - E loci
E loci - Z loci
A
rho = 1
P = 0.0004
mean fraction locus length covered by TIR (%)
PC1
B
rho = -0.96
P = 0.003
mean fraction locus length covered by Gypsy (%)
PC1
C
rho = 0.96
P = 0.003
median mCHH in WT embryo
PC1
D
rho = -0.99
P = 1e-5
median mCG in WT embryo
PC1
E
rho = -0.89
P = 0.012
median mCHG in WT embryo
PC1
F
TIR
short TEs
mCHG
mCG
long TEs
Gypsy
gene.dist
rho

Fig S6, supporting Fig 4 and Fig 5

(A) Scatter plot showing correlation of PC1 (**Fig 4C**) and mean fraction of locus length covered by short transposons (50-bp – 250-bp) (**Fig 5B**).

(B) Scatter plot showing correlation of PC1 (**Fig 4C**) and mean fraction of locus length covered by long retrotransposons (> 250-bp) (**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**), Long TEs overlaps (**Fig 5B**), DNA methylation (**Fig 5C**, see also **Fig S5**), and short TEs overlaps (**Fig 5B**).

Rho, Spearman's rank order correlation coefficient. *Embryo (7 days after flowering) siRNA data from Rodrigues et al (2013), which was based on a single replicate. Except zygote, all other data from Li et al. (2020).

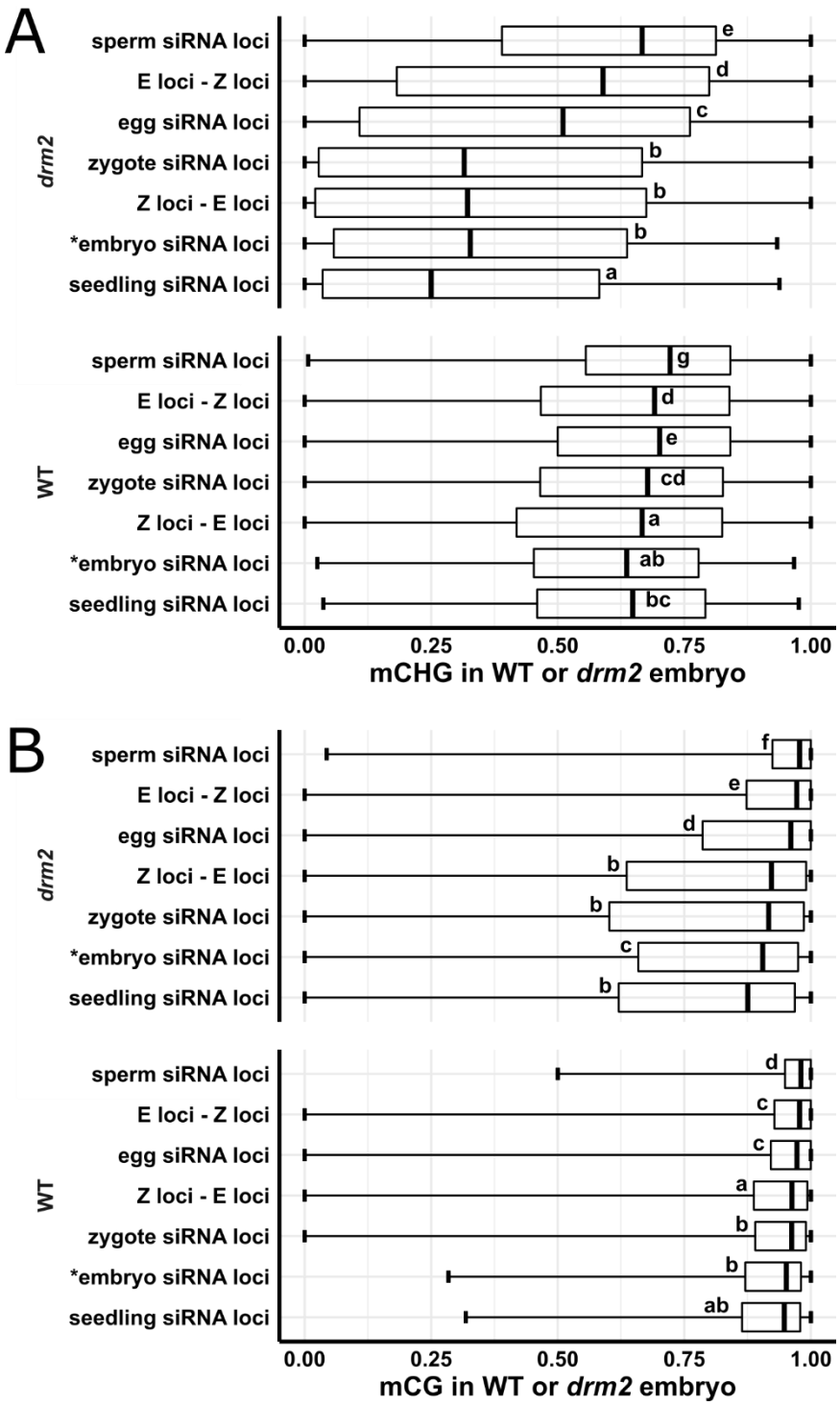


Fig S7, supporting Fig 5

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 (7 days after flowering) siRNA data from Rodrigues et al (2013), which was based on a single replicate. Except zygote, all other data from Li et al. (2020).

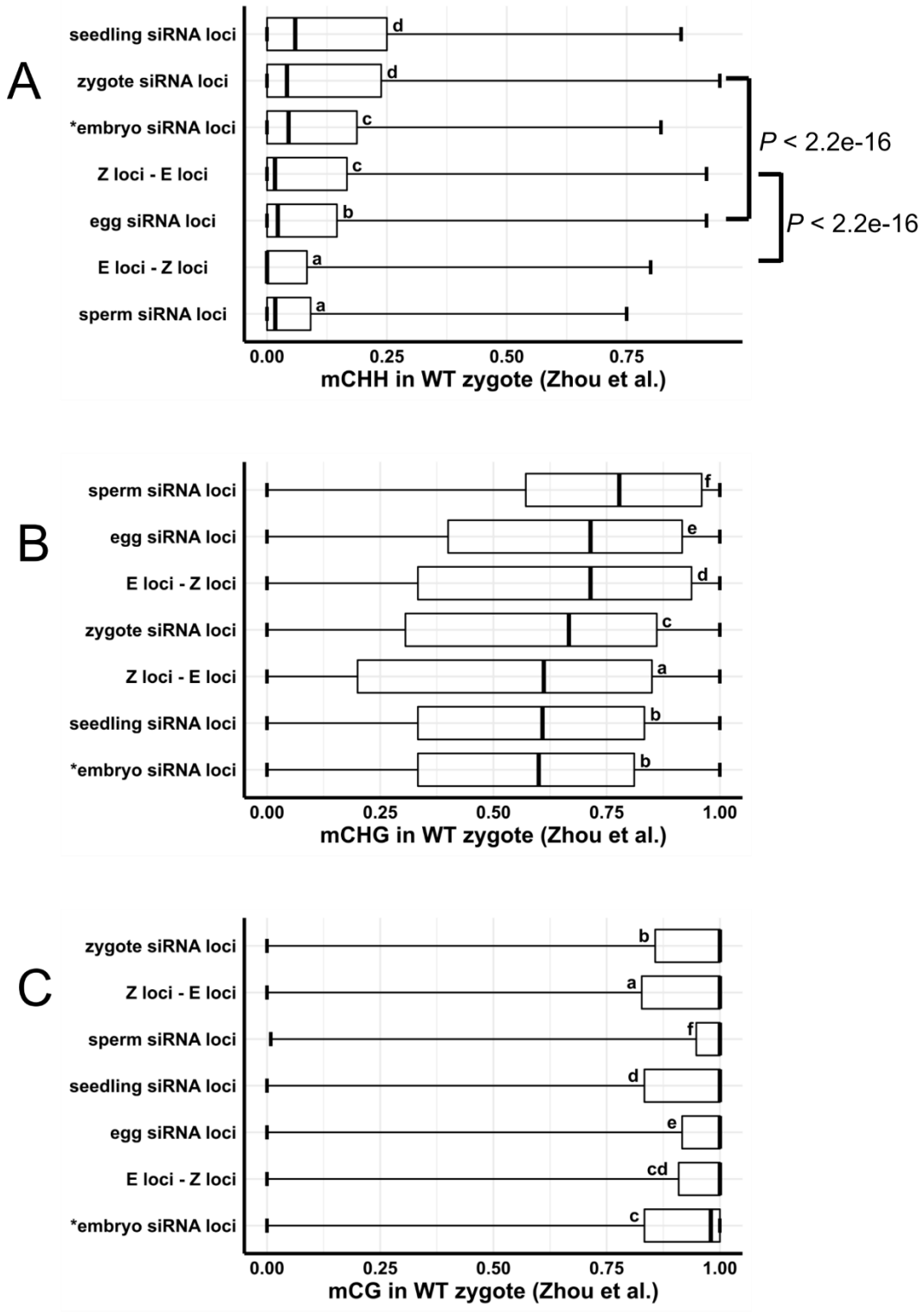


Fig S8, supporting Fig 5

Boxplots showing DNA methylation level in wildtype zygote at CHH (A), CHG (B) and CG (C) contexts. Middle lines are median. Boxes span interquartile range. Whiskers span 2.5th and 97.5th percentiles. *Embryo (7 days after flowering) siRNA data from Rodrigues et al (2013), without replicates. Zygote methylation data from Zhou et al. (2021).

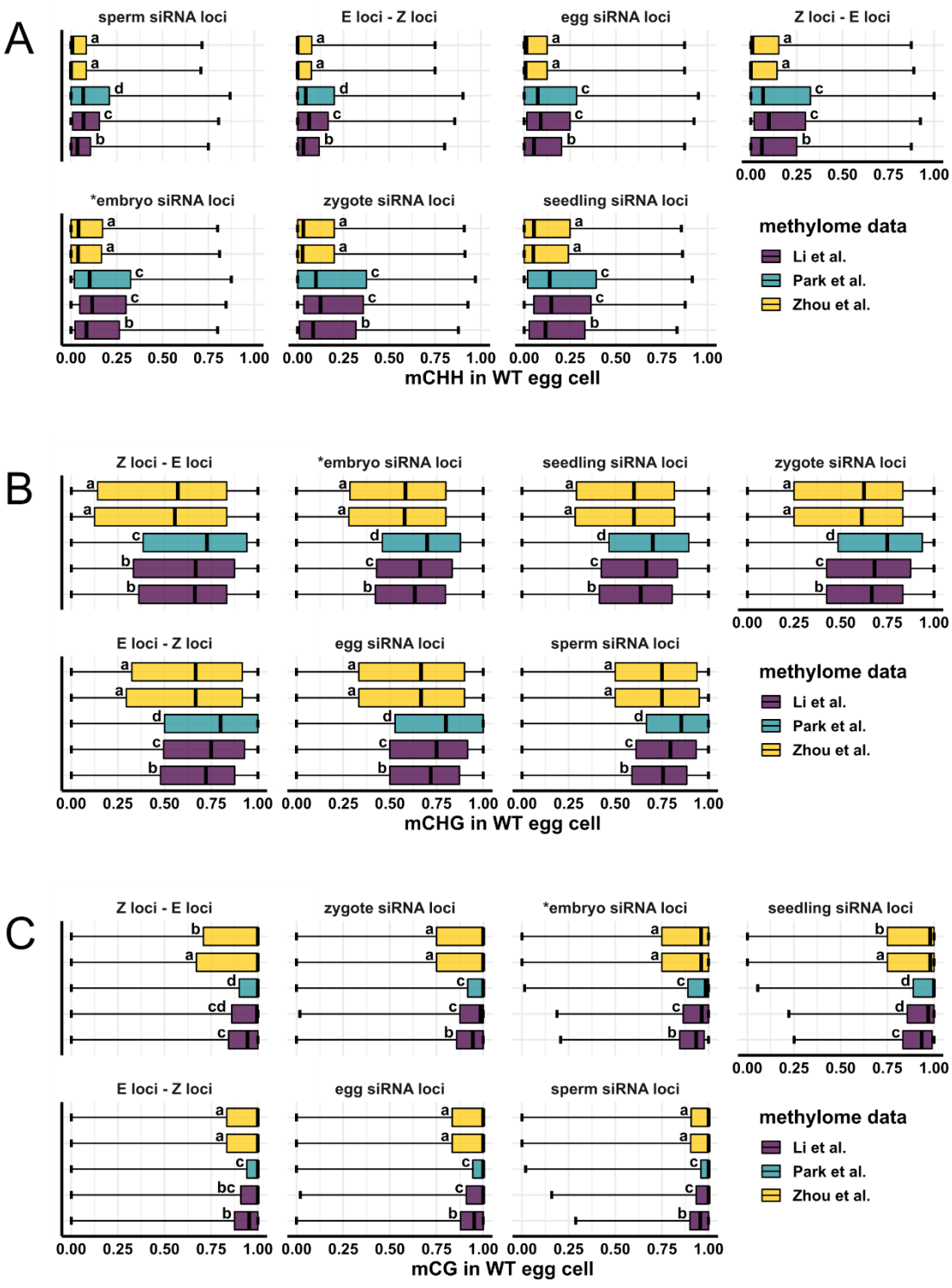


Fig S9, supporting Fig 5

Boxplots showing DNA methylation level in wildtype egg cell at CHH (A), CHG (B) and CG (C) contexts. Middle lines are median. Boxes span interquartile range. Whiskers span 2.5th and 97.5th percentiles. *Embryo (7 days after flowering) siRNA data from Rodrigues et al (2013), which was based on a single replicate. Data sources: Park et al. (2016), Li et al. (2020), and Zhou et al. (2021).

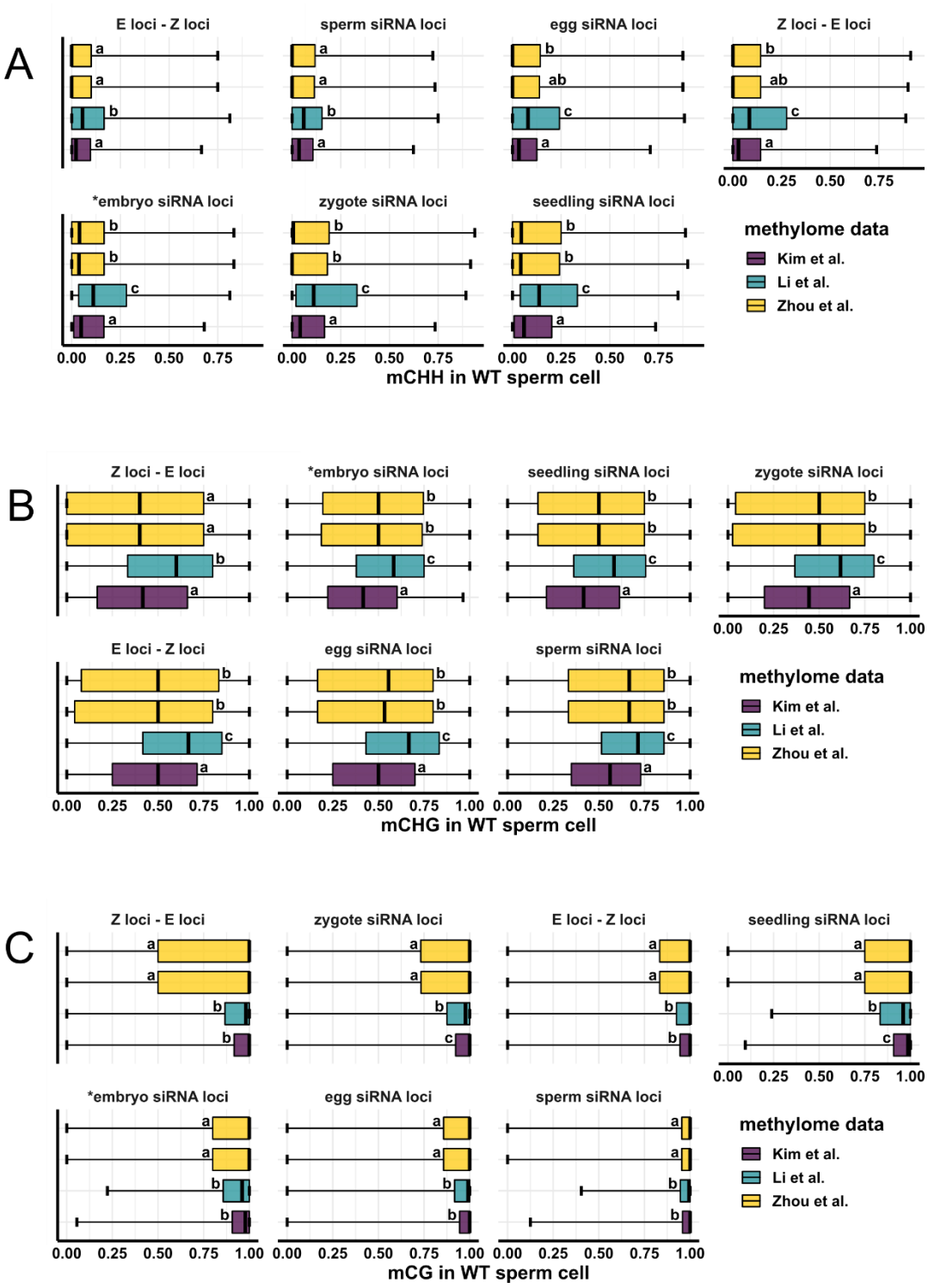


Fig S10, supporting Fig 5

Boxplots showing DNA methylation level in wildtype sperm cell at CHH (A), CHG (B) and CG (C) contexts. Middle lines are median. Boxes span interquartile range. Whiskers span 2.5th and 97.5th percentiles. *Embryo (7 days after flowering) siRNA data from Rodrigues et al (2013), which was based on a single replicate. Data sources: Kim et al. (2019), Li et al. (2020), and Zhou et al. (2021).

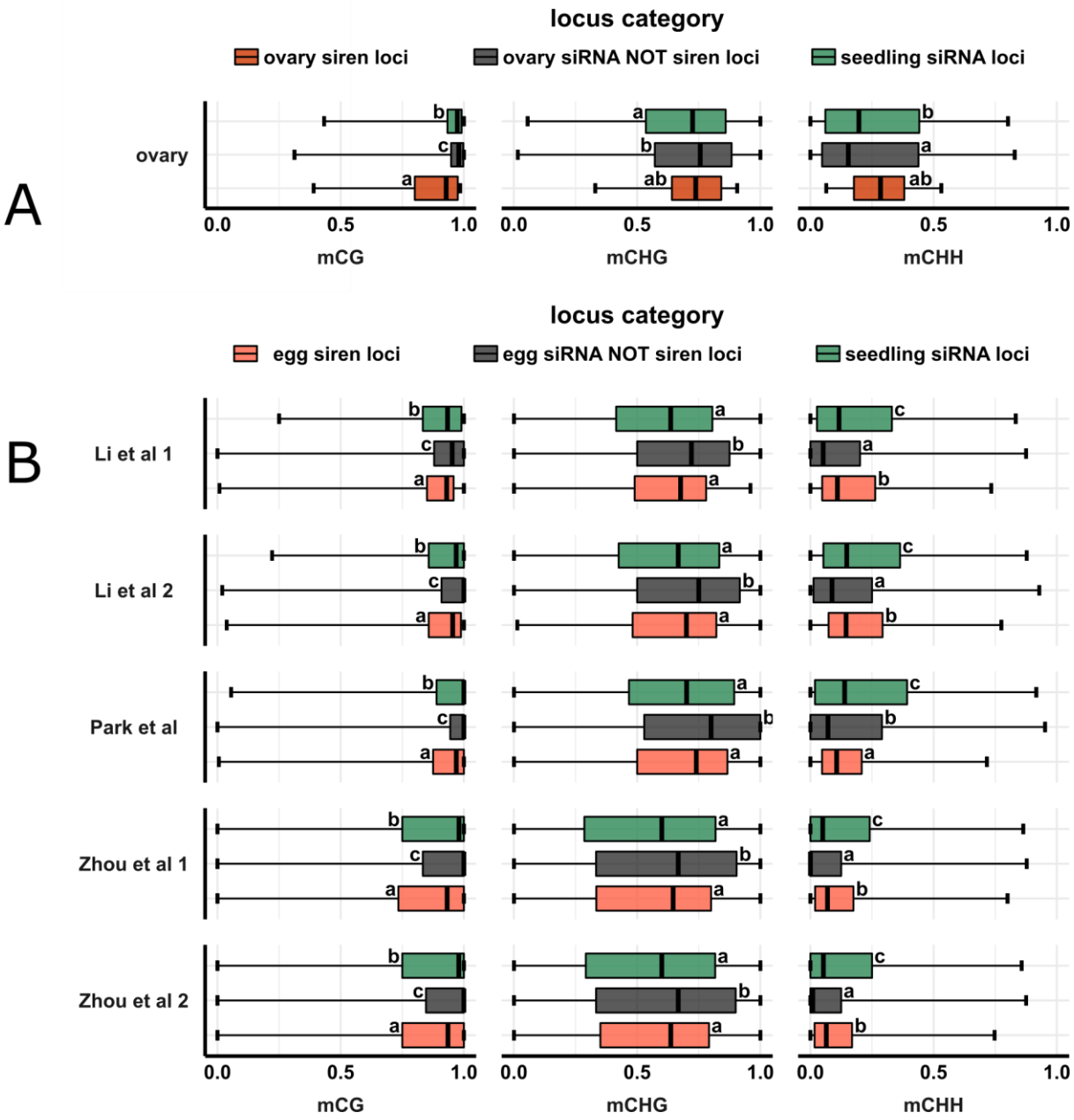


Fig S11, supporting Fig 2 and Fig 5

Boxplots showing DNA methylation level in ovary (A), and egg cell (B). Middle lines are median. Boxes span interquartile range. Whiskers span 2.5th and 97.5th percentiles. siRNA NOT siren loci refer to siRNA loci that are not siren loci. Letter groupings are based on Tukey tests. Data sources: Park et al. (2016), Li et al. (2020), and Zhou et al. (2021).

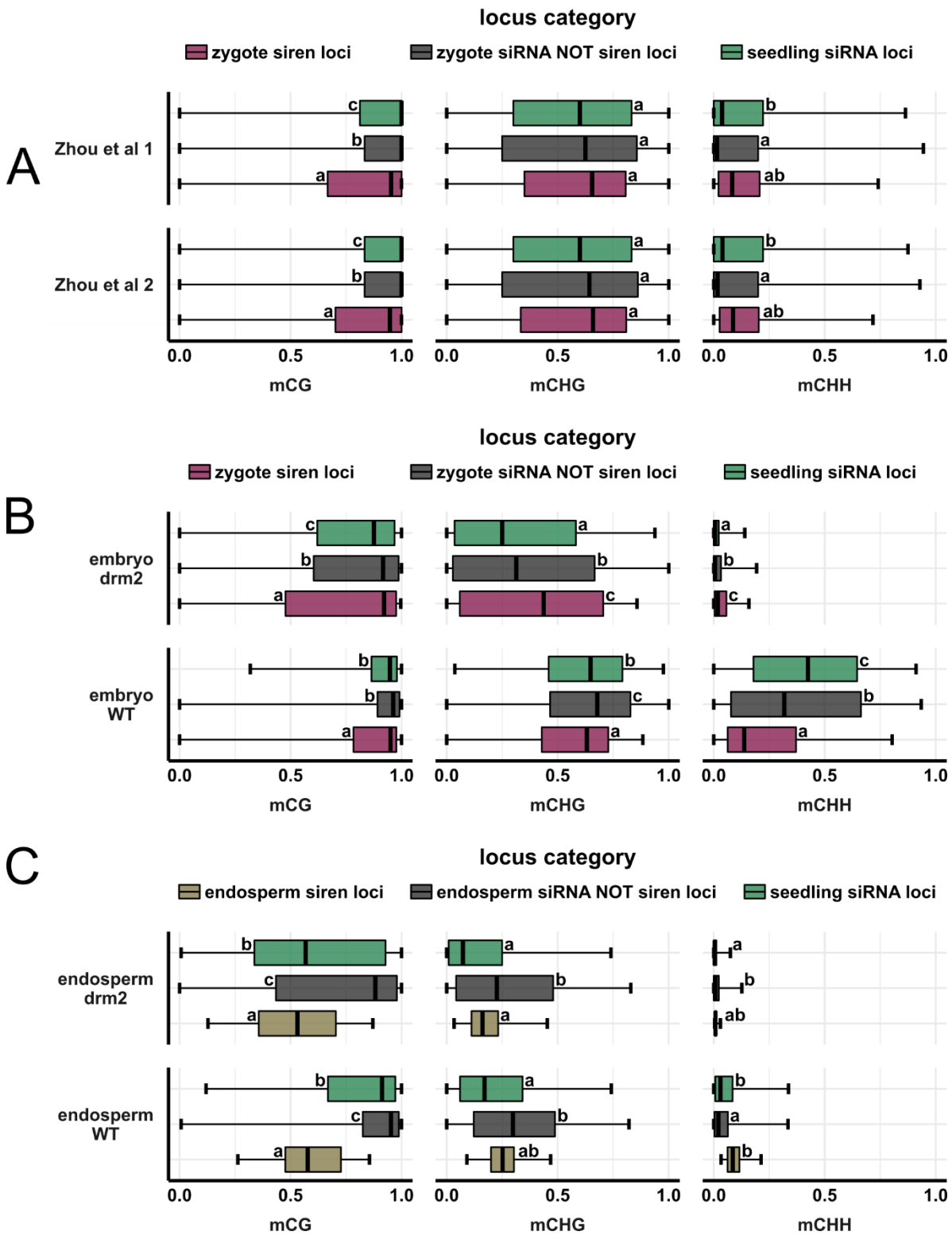

A
locus category
zygote siren loci
zygote siRNA NOT siren loci
seedling siRNA loci
Zhou et al 1
Zhou et al 2
mCG
mCHG
mCHH
B
locus category
zygote siren loci
zygote siRNA NOT siren loci
seedling siRNA loci
embryo drm2
embryo WT
mCG
mCHG
mCHH
C
locus category
endosperm siren loci
endosperm siRNA NOT siren loci
seedling siRNA loci
endosperm drm2
endosperm WT
mCG
mCHG
mCHH

Fig S12, supporting Fig 2 and Fig 5

Boxplots showing DNA methylation level in zygote (A) wildtype or *drm2* embryo (B) and wildtype or *drm2* endosperm (C). 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 endosperm siren loci. Letter groupings are based on Tukey tests. Zygote (Zhou et al.) data from Zhou et al. (2021). All other data from Li et al. (2020) and this study.

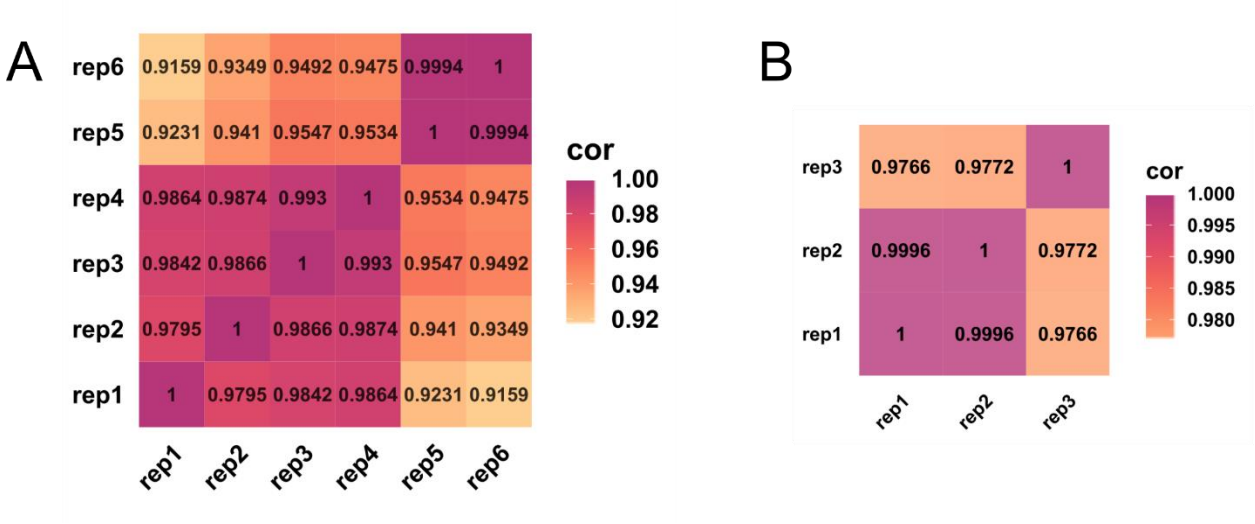
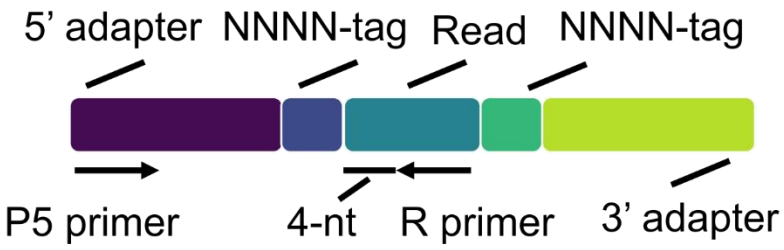


Fig S13, Pearson correlation of 24-nt siRNAs across replicates.

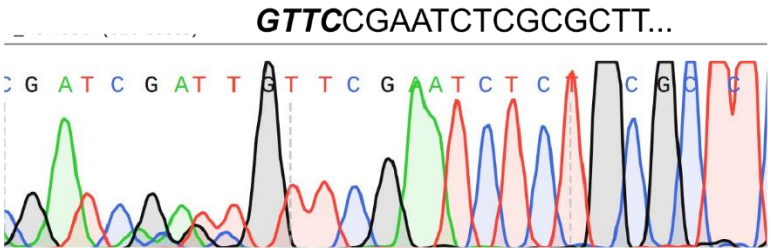
(A) Pairwise Pearson's correlation coefficient between six zygote libraries.

(B) Pairwise Pearson's correlation coefficient between three post-fertilization ovary libraries.

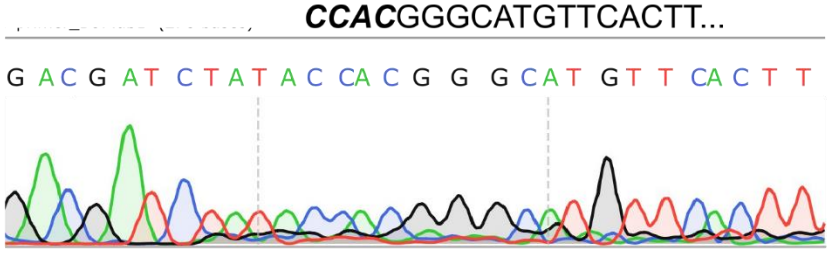
A


5’ adapter
NNNN-tag
Read
NNNN-tag
P5 primer
4-nt
R primer
3’ adapter

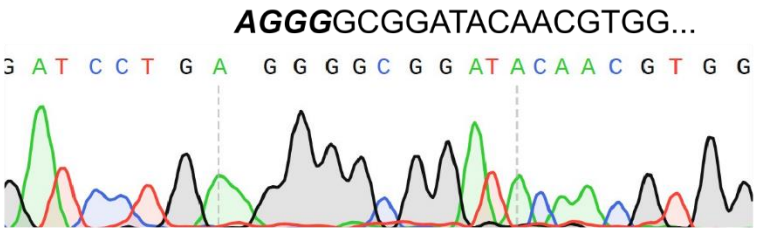
B


GTTCCGAATCTCGCGCTT...

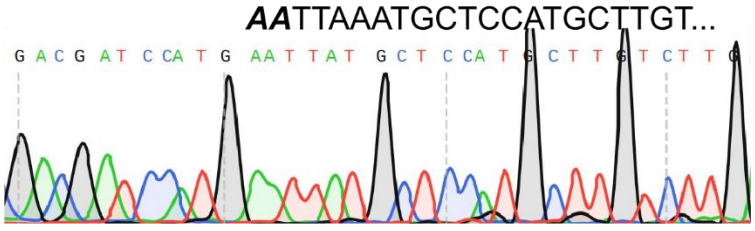
C


CCACGGGCATGTTCACTT...

D


AGGGGCGGATACAACGTGG...

E


AATTAAATGCTCCATGCTTGT...

F

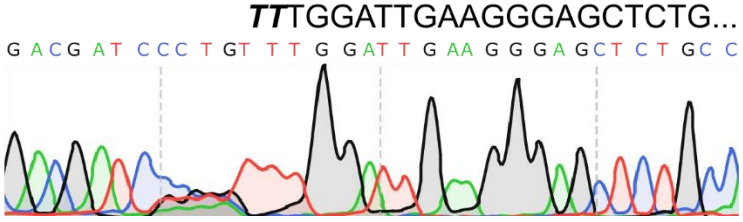

TTTGGATTGAAGGGAGCTCTG...

Fig S14, PCR and Sanger sequencing confirmation of Z-E loci siRNA reads.
(A) Schematics of the library molecule and the design of PCR. The Illumina universal P5 sequence is used as the forward primer. The reverse complement of the 3' portion of the small RNA read is used as the reverse primer. Note that the terminal 2 – 4-nt of the siRNA is omitted out in the reverse primer, such that the identity of the PCR product can be confirmed by Sanger sequencing. Amplifications that yielded PCR products of the correct predicted size were selected for sequencing, and the sequences are presented in (B) through (F).
(B) Sanger sequencing trace for the major siRNA at loci: Chr5:24981248-24981273. Bold italic text represents nucleotides not determined by the primer. Primer sequence: 5'-TGGCGGAAGCGCAGAGATTC-3'
(C) Sanger sequencing trace for the major siRNA at loci: Chr10:15829605-15829629. Bold italic text represents nucleotides not determined by the primer. Primer sequence: 5'-ATATCAAAGTGAACATGCCC-3
(D) Sanger sequence tracing for the major siRNA at loci: Chr3:9303476-9303712. Bold italic text represents nucleotides not determined by the primer. Primer sequence: 5'-TCCACCCACGTTGTATCCGC-3'
(E) Sanger sequence tracing for the major siRNA at loci: Chr3:20476277-20476300. Bold italic text represents nucleotides not determined by the primer. Primer sequence: 5'-TCAAGACAAGCATGGAGCATAA-3'
(F) Sanger sequence tracing for miRNA159b as a positive control. Bold italic text represents nucleotides not determined by the primer. Primer sequence: 5'-GGCAGAGCTCCCTTCAATCCA-3'