Supplemental Materials for Burgess D, *et al.* 2021

Contents:

Additional files:

Supplemental Methods

Transposon annotation in the *B. rapa* R-o-18 genome

To produce full-length exemplars a set of best matching sequences was retrieved and aligned using Muscle (https://www.ebi.ac.uk/Tools/msa/muscle/). These sequences were obtained through a combination of retrieving the five best RepeatMasker-generated matches (v. 4-0.5) based on the Smith-Waterman score or the top ten blastn matches using CoGe BLAST (https://genomevolution.org/CoGe/CoGeBlast.pl). Up to 1 kb of sequence was added at both ends to the RepeatMasker-generated matches using the Bedtools slop command. The top 10 CoGe blast matches were examined on CoGe GEvo panels (https://genomevolution.org/coge/GEvo.pl), and the largest high scoring segment pair (HSP) for each sequence was retrieved and extended by 20-50 nt. From the multiple sequence alignment full-length sequences were obtained and classified to superfamily based on characteristic features (Wicker, 2012) and on matches to reference repeats in Repbase (https://www.girinst.org/censor/). LTR retrotransposons were subdivided into LTR and internal region based on the presence of LTR direct repeats (5'-TG…3'CA) flanked by a 5 bp target site duplication (TSD) and the presence of a polypurine tract at the end of the internal region. LTR direct repeats and DNA transposon terminal inverted repeats (TIRs) were detected by BLASTing sequences against themselves. Non-autonomous TRIM and LARD retrotransposons were classified based on the size of their non-coding internal region (<600 bp and >2000 bp, respectively). LINE and SINE elements were defined based on a 3' poly(A) sequence and a flanking 7-21 bp TSD. DNA transposons were classified based on the TSD and characteristic TIR sequences. The ends of Helitrons were defined based on 5'(A)TC and 3'CTRR(T), and the presence of a small G/C-rich hairpin sequence near the 3' end. Captured host sequences, identified as blastn hits to Arabidopsis CDS sequences with an e-value of 1E-05 or less, were masked with N's to avoid the erroneous identification of genes as TEs by RepeatMasker. RepeatModeler sequences that were not manually annotated were sorted to superfamily based on blastn hits with an e-value of 1E-05 or less to classified TE sequences.

Wicker, T. (2012). So many repeats and so little time: How to classify transposable elements. In Plant Transposable Elements, Topics in current genetics. (Springer Berlin Heidelberg: Berlin, Heidelberg), pp. 1–15.

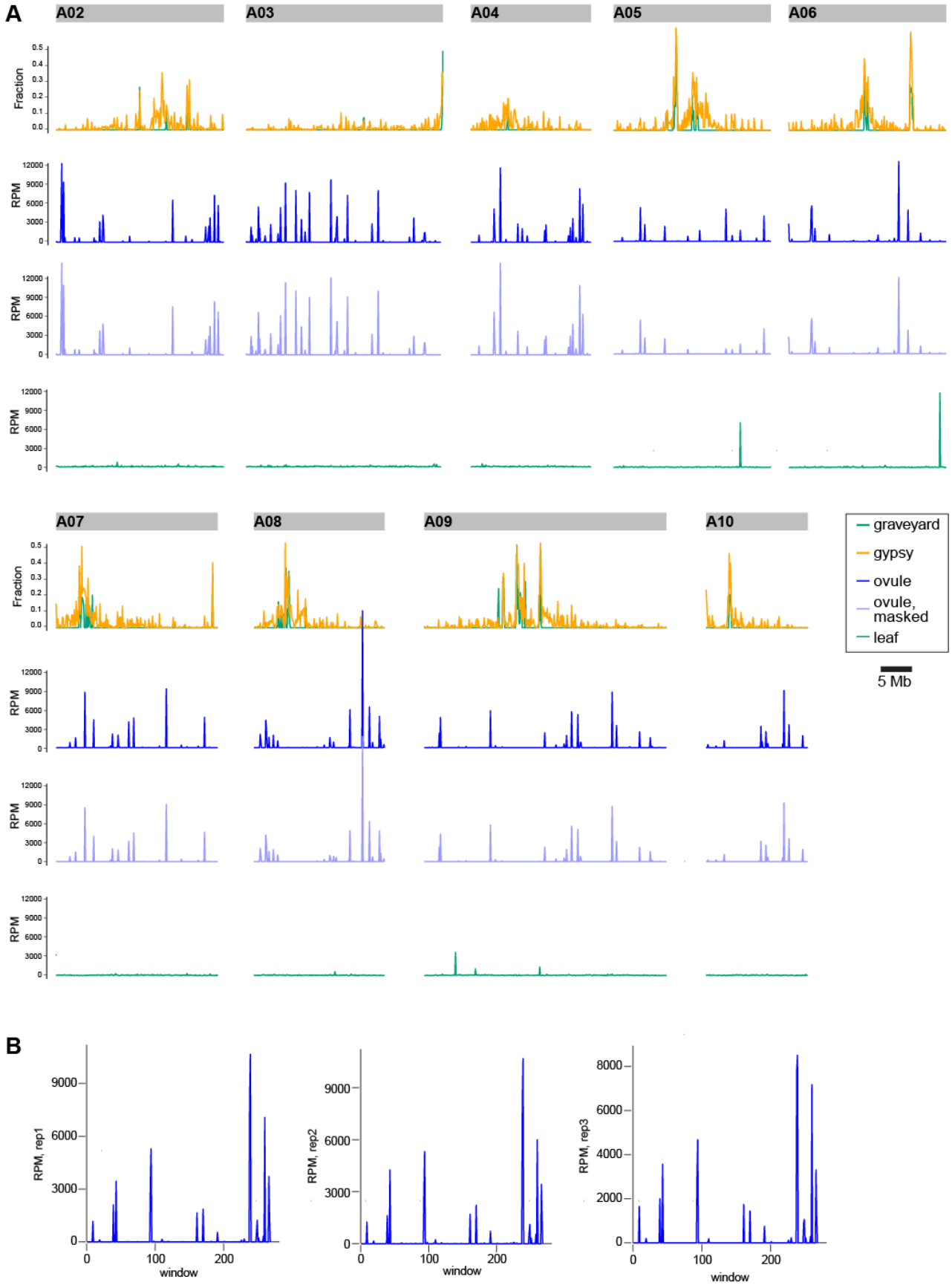
Supplemental Table 4. qRT-PCR results with p-values from 10 dpf seeds

Gene name	Fold change (*rdr2*/WT)	p-value	Fold change (*nrpd1*/WT)	p-value
A08p022510.1_BraROA	13.53	**1.03E-03**	7.67	**2.27E-04**
A08g510460.1_BraROA	6.94	5.98E-02	11.57	**4.90E-03**
A06p049610.1_BraROA	0.43	**8.17E-03**	1.25	4.09E-01
A10p034180.1_BraROA	0.96	8.41E-01	1.21	4.15E-01
A03p033390.1_BraROA	0.82	5.38E-01	2.20	3.61E-01
A03p024580.1_BraROA	3.52	9.28E-02	3.83	2.16E-02
A03p004630.1_BraROA	1.18	4.54E-01	1.18	3.59E-01
A03p003990.1_BraROA	348.14	4.68E-02	587.86	1.84E-02
A09p052220.1_BraROA	0.92	8.37E-01	1.27	3.10E-01
A07p005970.1_BraROA	1.06	8.14E-01	1.07	7.06E-01
A08p003700.1_BraROA	3.01	5.08E-02	5.02	**5.22E-03**
A08p003870.1_BraROA	7.18	**2.89E-03**	2.76	6.89E-02
A05g501190.1_BraROA	0.64	2.45E-01	0.72	9.20E-02
A07p017390.1_BraROA	0.39	3.48E-01	0.44	3.69E-01
A01p015570.1_BraROA	327.78	**1.02E-02**	241.3	**8.86E-03**
A03p020360.1_BraROA	2.55	2.85E-02	2.44	4.62E-02
A02p011510.1_BraROA	1.73	3.78E-01	1.01	9.89E-01
A03:20399808..20402498	1.29	5.77E-01	1.33	3.77E-01
A09p078220.1_BraROA	1.45	3.12E-01	2.19	2.00E-02
A01p059080.1_BraROA	10.26	**1.26E-03**	5.85	**1.69E-03**
A01p003160.1_BraROA	0.5	1.27E-01	0.37	6.80E-02
A02p057100.1_BraROA	1.59	6.03E-01	0.5	3.37E-01
A02p058640.1_BraROA	7.84	**7.49E-03**	9.42	**1.33E-04**

[1] p-values in **bold** are significant at 0.05 following FDR correction

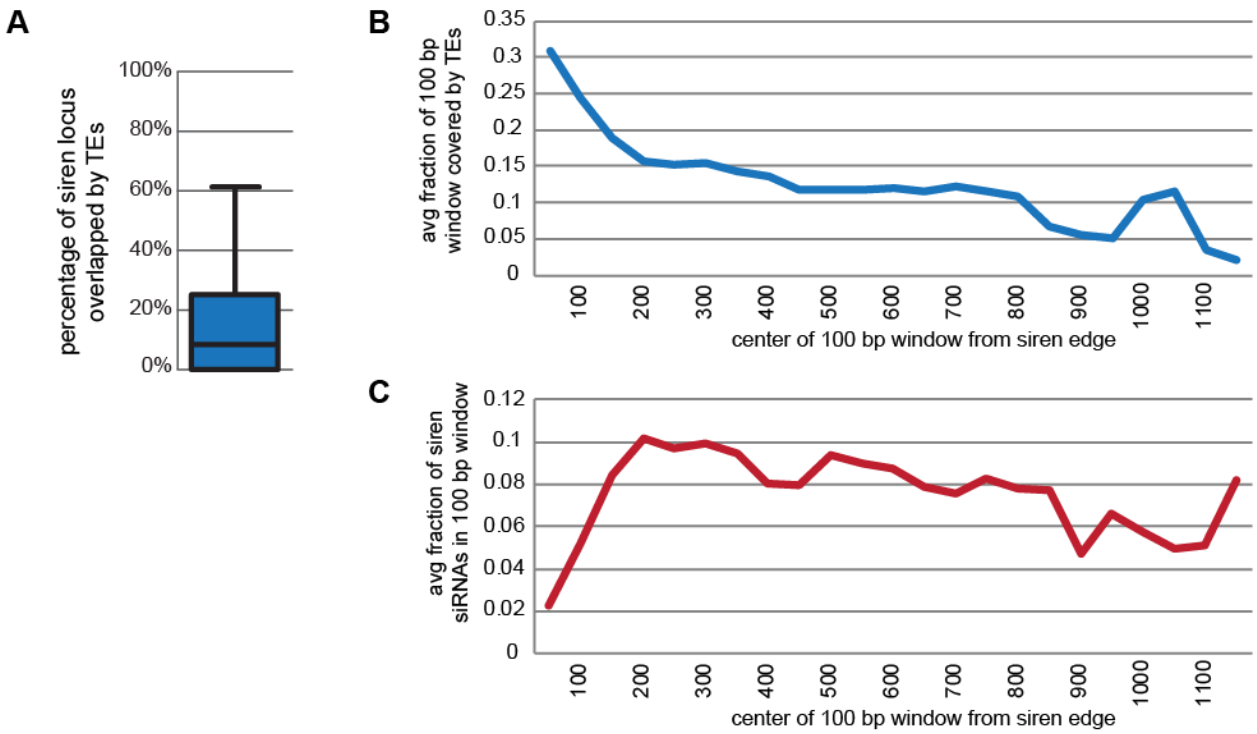
Supplemental Table 5. qRT-PCR primers

Gene name	F primer (5' to 3')	R primer (5' to 3')
A03p043420.1_BraROA (ACT2)	CTTGCACCAAGCAGCATGAA	CCGATCCAGACACTGTACTTCCTT
A08p022510.1_BraROA	CGGCGACTCCAGATCTAGATCAG	GGGTCAATGACGTCATCTATGAGC
A08g510460.1_BraROA	GCTCTATGCCCACATTGAGCGA	TGTTCTGTTGCCCCTGGAGT
A06p049610.1_BraROA	CACAGCGCCTCTCTACACAG	TCGTTGTTAGCTCTGTGAGGCT
A10p034180.1_BraROA	CAACGTCGCGAATCCGACA	CGAGCTTCTGGTCGTCATCG
A03p033390.1_BraROA	CGGTGAGTGTAACTATAGATACCCAGT	ACGCGTTCTTGCAGTCGTC
A03p024580.1_BraROA	GCATCTTGCCAACTTCTGAAGCG	GCCGTGTTGGGGAGTAAACGTAG
A03p004630.1_BraROA	GAGAAGCCAAGGATGAGAGGTC	GCCTGACATCATCATCCTTTCCAAG
A03p003990.1_BraROA	CGCTTCAAATGTGCTTCATGTGAC	GCTCCACATGAACATTGGAAGACTC
A09p052220.1_BraROA	CCCGGTTGCCTTTTTACATGTGG	TGGCGCATACCACAAGCTACC
A07p005970.1_BraROA	CTAATCCACCGGACCCTGT	GAGGGGTTGAAGATGGTTCGATG
A08p003700.1_BraROA	CGTCTCAACAGGAAGAACACTGC	CGCATCAGTAACCATGGACTTGC
A08p003870.1_BraROA	CCTTCATCGGCGTACTTGTGG	GCTGTTGAAAAACTTGAGCCCAGT
A05g501190.1_BraROA	GCATGTCGATGAAAGTTGTATCCAG	GCATTGAGAGACAGAACGGTGAC
A07p017390.1_BraROA	GGAGGAAGTGACAGTGGCAGT	GGGCTCTGAACTTTAGCAGGTCTC
A01p015570.1_BraROA	GAAAGCCTAATCCAGGACAGAGC	TGACAGGAGAAAGATCCGCACT
A03p020360.1_BraROA	CTTCTTGGAGTGCAAGGCTGATG	GAAGCCTTGAGGCTCGCTATG
A02p011510.1_BraROA	CATAGTGGCCGAACTCTTGGAC	AAAGGGAAGCCCTGATGGTC
A03:20399808..20402498	CAGGTCATGGTACGACGACTC	GCAATGTATGTAGGTGTGTCCGT
A09p078220.1_BraROA	GCGTTGAGCTTGGAGGAGCA	CAAACAAATCAGCAACGGGACCA
A01p059080.1_BraROA	CGCTACTACGGTTTTGTCACGTTC	GGCACCAAAACCGACAACGTAG
A01p003160.1_BraROA	GACAAACACTAGCGTCGGACA	GTCGCCGTAACTAAATCCGCAA
A02p057100.1_BraROA	GCAAGACCTTGTTGTGATGCTC	GGAATCCGACAGAGTTCAGGC
A02p058640.1_BraROA	CACCACCAGAGGTGATAGAGC	GTTACTGGAGCCGTTCTGTCAC


A
A02
A03
A04
A05
A06
A07
A08
A09
A10
Fraction
RPM
graveyard
gypsy
ovule
ovule, masked
leaf
5 Mb
B
RPM, rep1
RPM, rep2
RPM, rep3
window

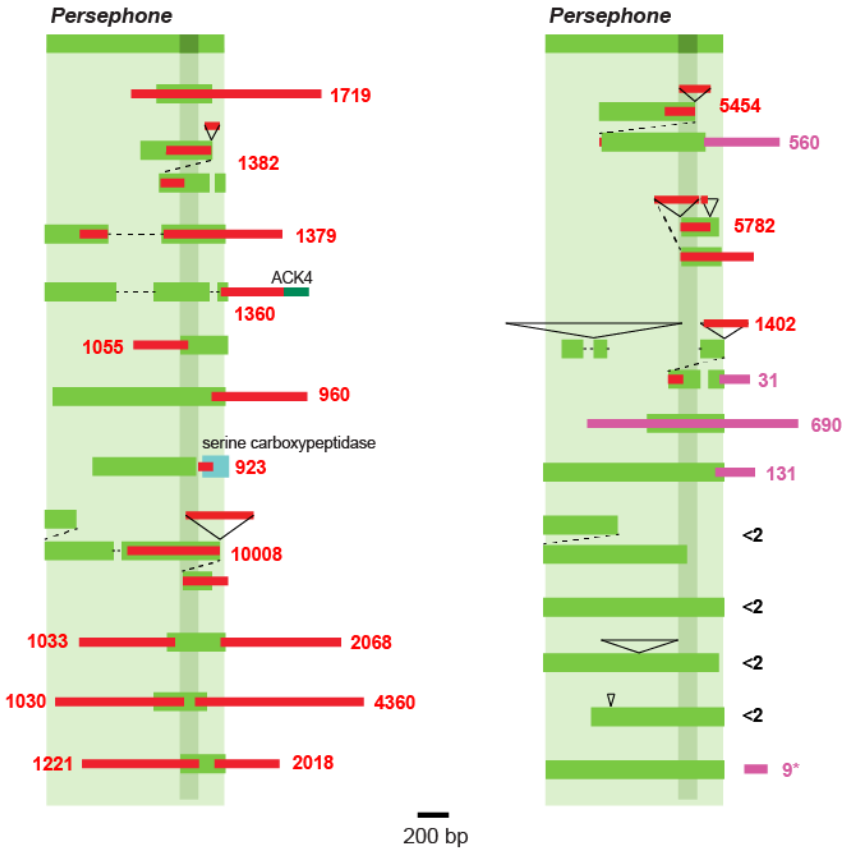
Supplemental Figure 1. *B. rapa* ovule 24-nt siRNAs predominantly map to non-repetitive hotspots on all chromosomes.

(**A**) Distribution of pooled, uniquely-aligning ovule and leaf siRNA on *B. rapa* chromosomes 2-9 (chromosome 1 shown in Figure 1C). RPM in 100 kb windows is plotted. Discrete, highly-expressed peaks in leaf data correspond to the intersection of sense/antisense transcripts or to transcribed regions with secondary structure. SiRNA production from these regions does not require RDR2 and is therefore not canonical RdDM. (**B-C**) Replicates of uniquely-aligning ovule siRNA abundance over chromosome 1 demonstrate the consistency of biological replicates (B, unmasked; C, masked).



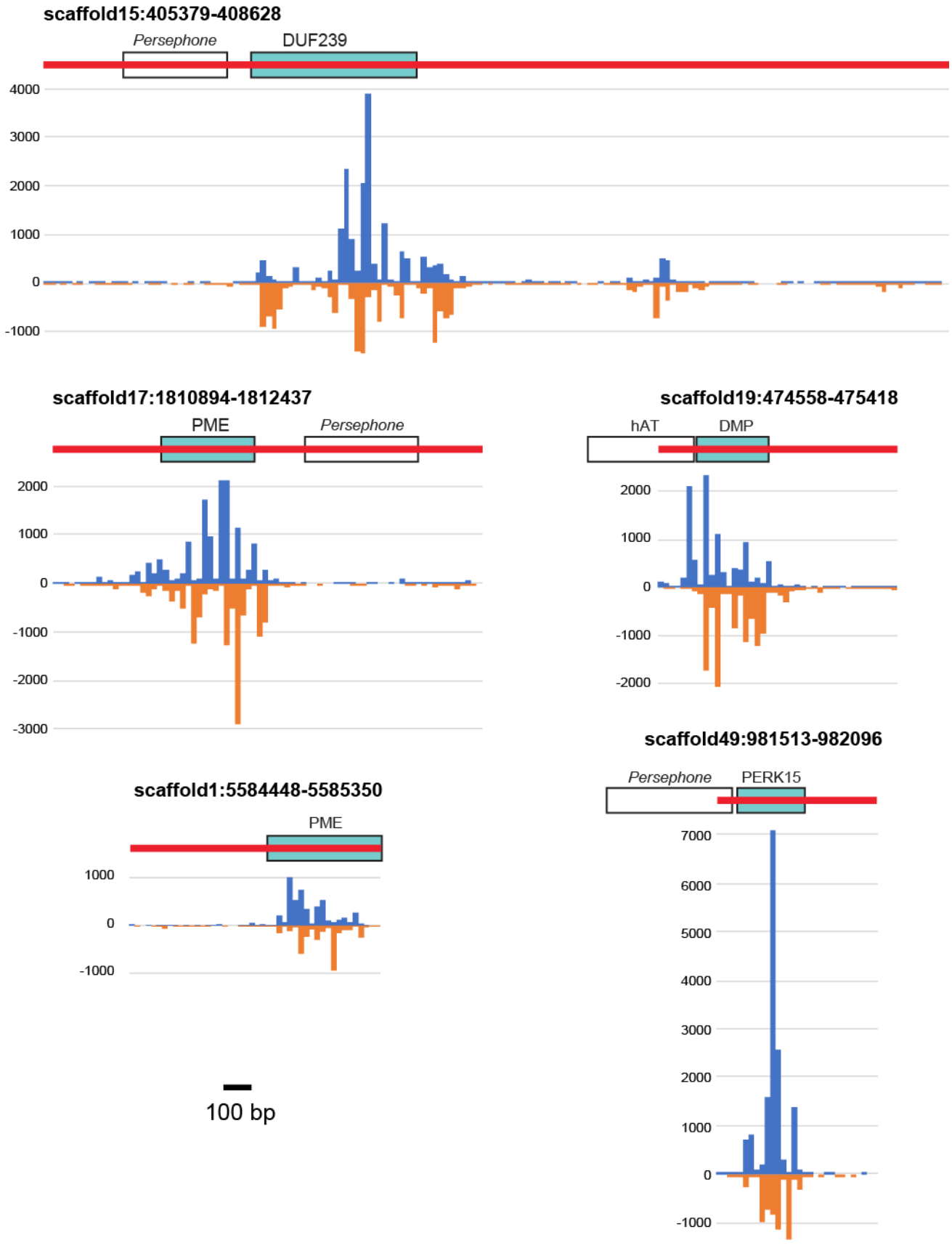
Supplemental Figure 2. Siren loci overlap TEs primarily at their edges.

(A) Distribution of TE coverage over *B. rapa* siren loci. **(B)** Change in TE coverage in 100-bp windows across siren loci. **(C)** siRNA accumulation in 100-bp windows across siren loci.



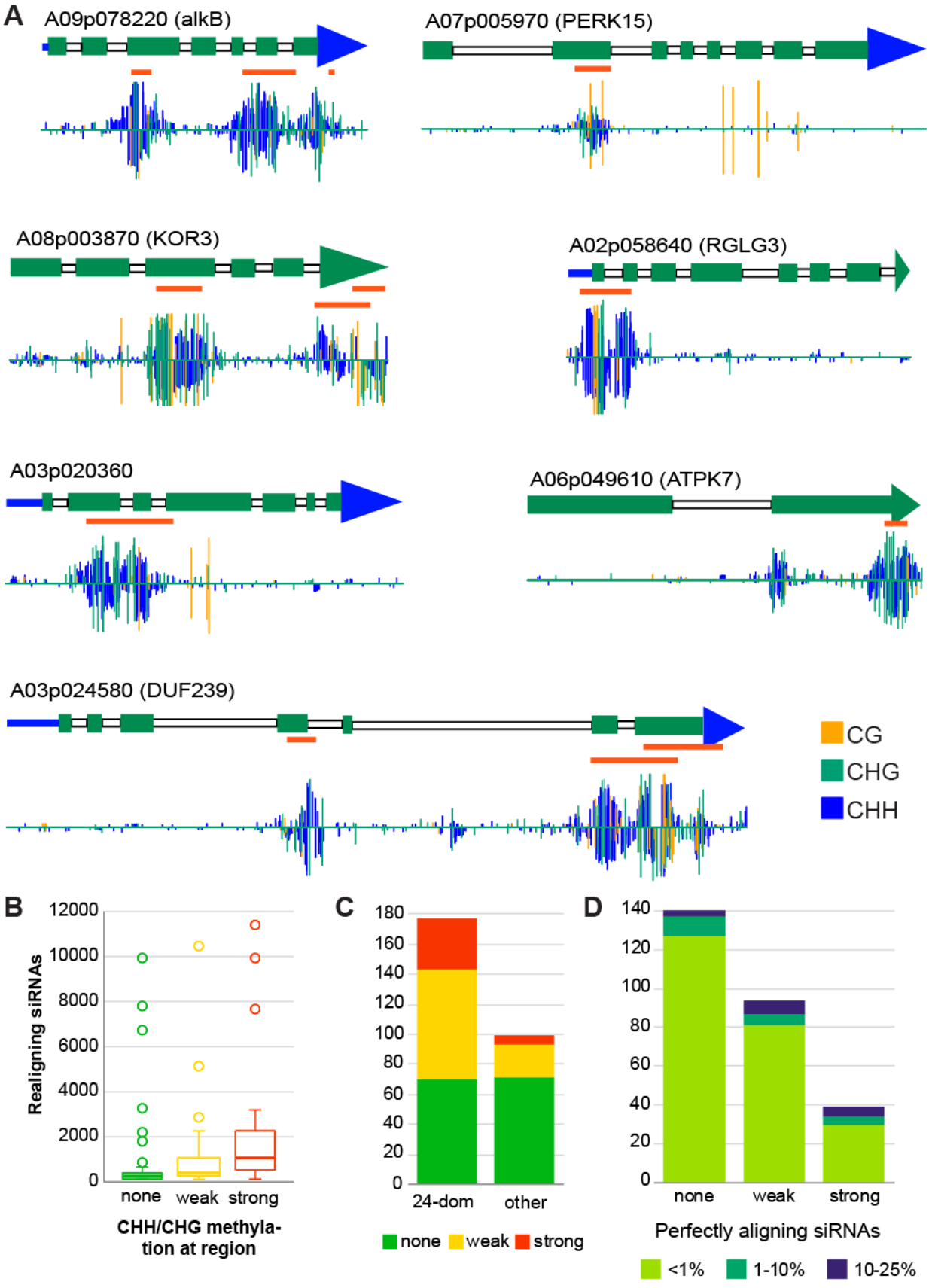
Supplemental Figure 3. Additional examples of *Persephone* elements (as in Figure 3).

Siren sequences are shown in red and 24-nt clusters with RPM falling below siren designation are shown in pink. Deletions are marked by dashed lines, insertions by triangles designating the length of the insertion. A dotted line connects tandem *Persephone* elements. The number of uniquely-aligning 24-nt siRNAs in each 24-nt cluster is indicated in RPM (reads per million alignable 19-26mer siRNAs in the pooled ovule libraries). Siren regions containing gene fragments, shown in blue, are labeled with the best hit *A. thaliana* gene. Shown in green is the 3'-UTR and part of the coding region of a gene related to Arabidopsis peroxisomal acyl-CoA oxidase *ACX4* (AT3G51840). Not all gene fragments/overlapping genes associated are shown. The final *Persephone* element in this figure, delineated with an asterisk, has a low number of 24-nt siRNAs in ovule, but a substantially increased number in seed coat.


scaffold15:405379-408628
Persephone
DUF239
4000
3000
2000
1000
0
-1000
scaffold17:1810894-1812437
PME
Persephone
2000
1000
0
-1000
-2000
-3000
scaffold19:474558-475418
hAT
DMP
2000
1000
0
-1000
-2000
scaffold1:5584448-5585350
PME
1000
0
-1000
100 bp
scaffold49:981513-982096
Persephone
PERK15
7000
6000
5000
4000
3000
2000
1000
0
-1000

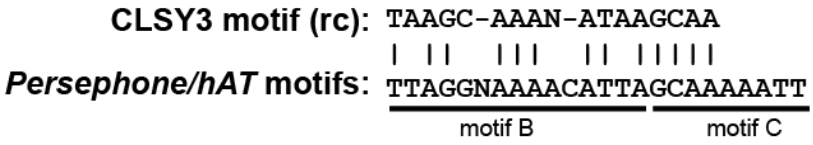
Supplemental Figure 4. siRNAs accumulate at siren gene fragments.

Generally, siRNA accumulation is most abundant from gene fragments (blue boxes) within siren loci (red bars). Transposable elements (white boxes) are not the primary source of siRNAs. Total accumulation of siRNA 5' ends in 20 bp windows tiled across the siren locus are shown below (blue, sense; orange, antisense). The second siren locus containing a gene fragment similar to PME42 (AT4G03930) is scaffold17:1810894-1812437.


A
A09p078220 (alkB)
A07p005970 (PERK15)
A08p003870 (KOR3)
A02p058640 (RGLG3)
A03p020360
A06p049610 (ATPK7)
A03p024580 (DUF239)
CG
CHG
CHH
B
Realigning siRNAs
12000
10000
8000
6000
4000
2000
0
none
weak
strong
CHH/CHG methyla-
tion at region
C
180
160
140
120
100
80
60
40
20
0
24-dom
other
none
weak
strong
D
140
120
100
80
60
40
20
0
none
weak
strong
Perfectly aligning siRNAs
<1%
1-10%
10-25%

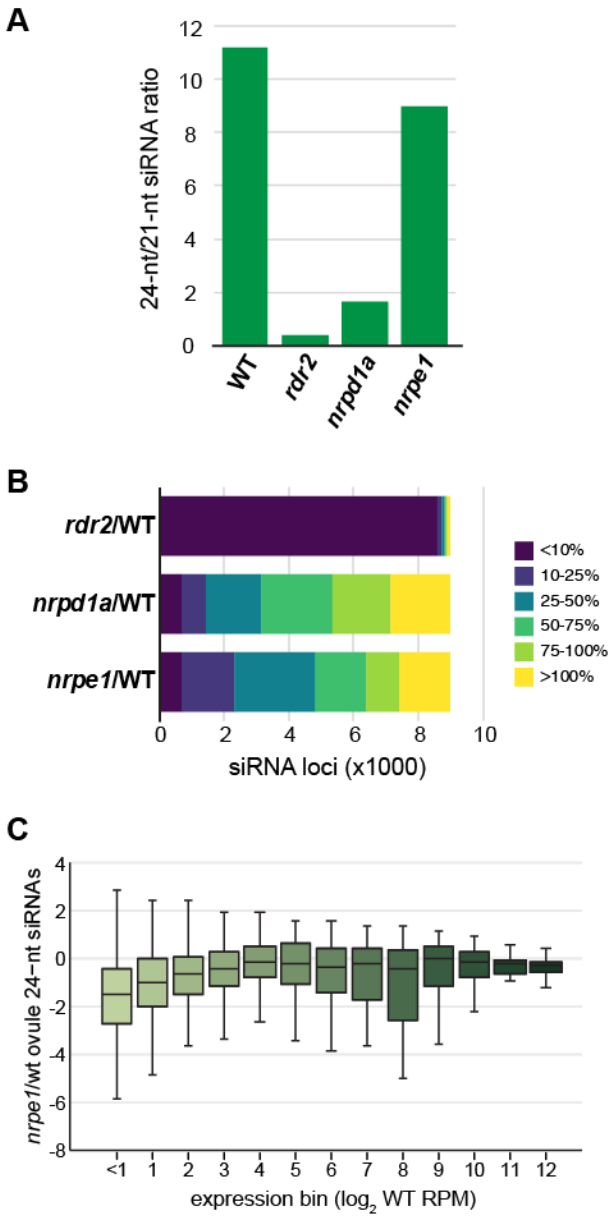
Supplemental Figure 5. Protein coding genes are specifically methylated at regions of homology to siren gene fragments.

(A) Cytosine methylation in ovule for several protein-coding genes with homology to siren gene fragments. The name of each gene is listed above the gene model, with the Arabidopsis homolog listed in parentheses. Coding exons are shown in green, UTRs are shown in blue and introns in white. Regions with similarity to siren gene fragments are depicted with an orange bar beneath the gene model. Below each gene model is a representation of cytosine methylation across the gene, demonstrating that abundant methylation coincides with the region of similarity to the siren gene fragment. Stronger CHH and CHG methylation in ovules is associated with higher numbers of realigning siRNAs **(B)** and a preponderance of 24-nt realigning siRNAs **(C)**. **(D)** Even at genes with strong methylation, most re-aligning siRNAs are not perfectly complementary.



Supplemental Figure 6. Comparison of siren TE motifs and CLSY3 binding motif.

The reported CLSY3-associated "motif 1" is similar to motifs B and C that are shared by siren-associated *Persephone* and *Bra_hAT1* elements.



Supplemental Figure 7. *nrpd1a-2* is required only at the highest-expressing siRNA loci in ovules. Although *nprd1a-1* has a reduction in 24-nt siRNA accumulation comparable to *rdr2-2*, most siRNA loci show only a small reduction in siRNA accumulation. **(A)** The ratio of 24-nt to 21-nt siRNAs in *B. rapa* ovules from WT and three RdDM mutants. **(B)** The percent of siRNA expression remaining in ovules of three *B. rapa* RdDM mutants. **(C)** There is no correlation between expression level in wild type and the loss of siRNAs in *nrpe1*. Figure 7C demonstrates that loci with the highest expression in wild type have the strongest loss of siRNAs in *nrpd1a*.