Several phased siRNA annotation methods can produce frequent false-positives for 24 nucleotide RNA-dominated loci in plants

Seth Polydore[1,2], Alice Lunardon[2], and Michael J. Axtell[1,2*]

1. Genetics Ph.D. Program, Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802 USA.
2. Department of Biology, The Pennsylvania State University, University Park, PA 16802 USA.

[*] Corresponding author: mja18@psu.edu

Abstract

Small RNAs regulate key physiological functions in land plants. Small RNAs can be divided into two categories: microRNAs (miRNAs) and short interfering RNAs (siRNAs); siRNAs are further sub-divided into the imprecisely processed heterochromatic siRNAs and more precisely processed phased siRNAs (phasiRNAs). 21-22 nucleotide (nt)-dominated phasiRNA-producing loci (*PHAS*) are well represented in all land plants to date, but male reproductive tissue-associated 24 nt-dominated *PHAS* loci are known only in monocots. Currently, only one miRNA (miR2275) is known to trigger the production of these 24 nt-dominated *PHAS* loci. In this study, we use stringent methodologies in order to examine whether or not 24 nt-dominated *PHAS* loci also exist in *Arabidopsis thaliana*. We find that highly expressed 24 nt heterochromatic siRNA loci can consistently be mis-identified as *PHAS* loci in different libraries and using different *PHAS*-detecting algorithms. We also find that *MIR2275* is not found in *A. thaliana*, and it seems to have been lost in the last common ancestor

23 of Brassicales. Altogether, our research highlights the potential issues with widely used siRNA
24 phasing-detecting algorithms which may lead to false positives when trying to annotate new
25 *PHAS* loci, especially 24 nt-dominated loci.

26

27 **Introduction**

28 Small RNAs regulate key physiological functions in land plants, ranging from
29 organogenesis (Boualem et al., 2008, 2008; Kutter et al., 2007; Laufs et al., 2004; Williams et
30 al., 2005) to gametogenesis (Grant-Downton et al., 2009). Three major protein families are
31 involved in the biogenesis of small RNAs. The first family is the DICER-LIKE (DCL) protein
32 family. Consisting of four paralogs in the *Arabidopsis thaliana* genome (Baulcombe, 2004;
33 Chapman & Carrington, 2007), DCL proteins hydrolyze RNA precursors into 20-24 nt double-
34 stranded RNA fragments (Millar & Waterhouse, 2005). These double-stranded RNA fragments
35 are then loaded into ARGONAUTE (AGO) proteins, the second protein family, and one strand
36 of the RNA is discarded (Ender & Meister, 2010). Upon Watson-Crick binding to other RNA
37 transcripts in the cell, the AGO/single-stranded RNA complex represses other RNA transcripts
38 (Baumberger & Baulcombe, 2005; Qi et al., 2005). Overall, 10 AGOs are encoded in the *A.*
39 *thaliana* genome (Tolia & Joshua-Tor, 2007). RNA DEPENDENT RNA POLYMERASES
40 (RDRs) are the third family of proteins involved in the biogenesis of many small RNAs. RDRs
41 convert single-stranded RNAs into double-stranded RNAs by synthesizing the complementary
42 strand of the RNA molecule (Willmann, et al., 2011). Six *RDRs* are encoded in the *A. thaliana*
43 genome (Willmann, et al., 2011).

44 Small RNAs can be divided into two major categories: microRNAs (miRNAs) which are
45 precisely processed from single-stranded, stem-and-loop RNA precursors (Millar & Waterhouse,
46 2005), and short interfering RNAs (siRNAs), which are derived from double-stranded RNA
47 precursors (Axtell, 2013). siRNAs are further divided into several different groups, including
48 phased siRNAs (phasiRNAs) and heterochromatic siRNAs (hc-siRNAs). Predominantly 24 nts
49 in length, hc-siRNAs function to repress transcription of deleterious genomic elements by
50 encouraging the formation of heterochromatin in targeted areas (Baulcombe, 2004; Sugiyama et
51 al., 2005). Biogenesis of hc-siRNAs begins transcription of transposable elements and repeat-
52 rich regions via the plant-specific holoenzyme, DNA DEPENDENT RNA POLYMERASE IV
53 (Pol IV) (Onodera et al., 2005). The resulting transcript is then converted into double-stranded
54 RNA by RDR2 and this double-stranded transcript is hydrolyzed by DCL3 (Matzke et al., 2009).
55 phasiRNAs are derived from DNA DEPENDENT RNA POLYMERASE II (Pol II) transcripts
56 that have been targeted by miRNAs (Fei et al., 2013). Upon miRNA-mediated hydrolysis, the
57 RNA transcript is converted into double-stranded RNA by RDR6 (Cuperus et al., 2010). The
58 resulting double-stranded RNA is then cleaved into 21nt double-stranded RNA fragments by
59 DCL4 (and less frequently DCL2) (Axtell et al., 2006).

60 21-22 nt phasiRNA-producing loci (*PHAS*) are clearly represented in all land plants that
61 have been sequenced thus far. However, 24 nt dominated *PHAS* loci are only currently
62 described in rice (Song et al., 2011), maize (Zhai et al., 2015), and other non-grass monocots
63 (Kakrana et al., 2018). Much like 21 nt-dominated *PHAS*, the biogenesis of these 24 nt-
64 dominated *PHAS* loci begins with the Pol II-dependent transcription of a single-stranded RNA
65 precursor which is then targeted by miR2275 and hydrolyzed. To date, miR2275 is the only
66 miRNA known to trigger the production of 24 nt-dominated phasiRNAs (Fei et al., 2013). The

67 resulting RNA transcript is then converted into a double-stranded RNA molecule by RDR6 (Zhai
68 et al., 2015). However, these phasiRNA precursors are then hydrolyzed by DCL5 (a DCL3
69 homolog and is thus sometimes called DCL3b) to produce 24 nt phasiRNAs (Fei et al., 2013).

70 Aside from the combination of their size and biogenesis patterns, 24 nt-dominated *PHAS*
71 loci are distinct in various ways. These loci as well as their triggering miRNA, miR2275, are
72 very specifically expressed in the tapetum during early meiosis and quickly recede in expression
73 in other stages of male gametogenesis in rice and maize (Tamim et al., 2018). The AGO protein
74 that loads these phasiRNAs is unknown; however, in maize, *AGO18b*'s expression levels match
75 those of the 24 nt-dominated *PHAS* loci quite closely and is therefore the most likely candidate
76 to load 24 nt phasiRNAs (Komiya et al., 2014; Zhang et al., 2015). The targets of these 24 nt
77 phasiRNAs are unknown, but they are apparently necessary for proper male gametogenesis (Ono
78 et al., 2018). 24 nt-dominated *PHAS* loci were also described in the non-grass monocots
79 asparagus, lily, and daylily (Kakrana et al., 2018). These phasiRNAs are produced from
80 processing of inverted repeat (IR) RNAs, instead of the double-stranded RNA precursors
81 observed in rice and maize (Kakrana et al., 2018). Although the 24 nt-dominated *PHAS* loci
82 from non-grass monocots are still expressed most greatly in reproductive tissue, in asparagus
83 they are also expressed in female reproductive tissue (Kakrana et al., 2018).

84 We set out to search for evidence of 24 nt-*PHAS* loci in plants besides monocots. We
85 searched for 24 nt *PHAS* loci in the *A. thaliana* genome using small RNA-seq data. Currently,
86 several distinct algorithms are available to calculate the "phasing" of a sRNA-producing locus
87 (Dotto et al., 2014; Guo et al., 2015; Zheng et al., 2014). In general, these algorithms calculate
88 the number of reads that are "in-phase" against those that are "out-of-phase" in order to
89 determine the likelihood that a particular locus truly produces phasiRNAs (Axtell, 2010).

90 However, 24 nt-dominated siRNA loci are very numerous in angiosperms, and therefore are a
91 potential source of false-positives during searches for *PHAS* loci. We therefore carefully
92 examined *A. thaliana* 24 nt-dominated loci that consistently passed *PHAS*-locus detecting
93 algorithms using multiple methods and found that they are likely just heterochromatic siRNAs
94 (hc-siRNAs). We also use two other methods to examine the presence of 24 nt-dominated *PHAS*
95 loci in the *A. thaliana* genome. We searched for *rdr6*-dependent, 24 nt-dominated loci and found
96 18 such loci, but none had detectable phasing. We also examined miR2275, which triggers 24 nt
97 phasiRNA biogenesis in rice and maize, but found that the Brassicales clade contains no
98 potential homologs for this miRNA. Overall, our results suggest that there are no true 24 nt
99 *PHAS* loci in *A. thaliana*. Furthermore, our analysis shows that existing phasing score
100 algorithms to detect novel *PHAS* loci can lead to false positives.

101

102 **Results**

103 **miR2275 is not found in the *Brassicales* clade**

104 Currently known 24 nt dominated phasiRNA precursors are known to be targeted only by
105 a single miRNA family, miR2275 (Song et al., 2008; Zhai et al., 2015). We examined all
106 available angiosperm genomes on Phytozome (ver 12.1) for potential homologs of *MIR2275*. In
107 monocots, all but *Spirodela polyrhiza* and *Zostera marina* had potential miR2275 homologs
108 (Figure 1; Figure S1). Ten eudicots had potential miR2275 homologs based on sequence
109 similarity (Figure 1). We therefore interrogated small RNA libraries for these species to see if a
110 mature miR2275 homolog was expressed. Five eudicots had evidence of mature miR2275
111 accumulation (Figure 1). All of the monocots for which we obtained small RNA-seq data

112 expressed mature miR2275, except *Musa acuminata* (Figure 1). Importantly, because of the high
113 specificity of miR2275 expression in developing anthers (Tamim et al., 2018), it's possible that
114 our analysis includes false negatives, especially in situations where no reproductive tissue small
115 RNA libraries were available. Altogether, our data suggest that miR2275 is not found in *A.*
116 *thaliana* and that this loss apparently occurred before the last common ancestor for Brassicales.

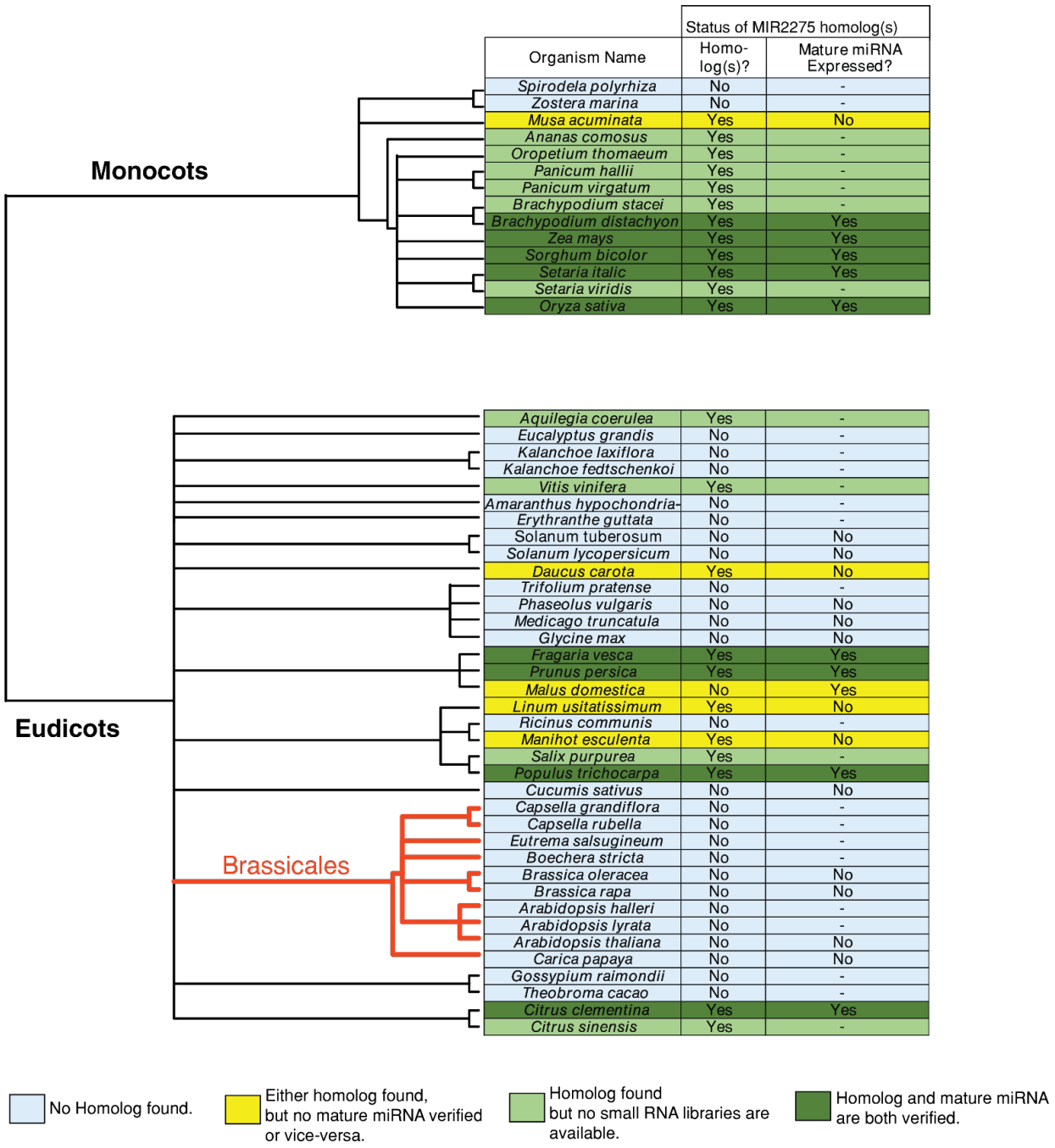


Figure 1. Conservation of *MIR2275* homologs in Phytozome v. 21-curated species. Green cells represent species that have *MIR2275* homologs based on sequence homology. Phylogenetic tree branches are not to scale and are adapted from NCBI's species tree representation. "-" indicates that small RNA-seq data for a given species were not available.

Three *A. thaliana* hc-siRNA loci consistently pass *PHAS*-detecting algorithms

124 Although *A. thaliana* lacks miR2275, it's possible that 24 nt *PHAS* loci exist in *A.*
125 *thaliana* and are triggered by a different small RNA. We reasoned that true 24 nt *PHAS* loci
126 would be dependent on one or more of the well-described *A. thaliana RDR* genes: *RDR1, RDR2,*
127 or *RDR6*. We thus examined a previously described differential expression analysis that
128 identified *A. thaliana* small RNA loci that were down-regulated in an *rdr1/rdr2/rdr6* triple
129 mutant (Polydore & Axtell, 2018). The phase scores of *rdr1/rdr2/rdr6*-dependent, 24 nt-
130 dominated loci were calculated in eight independent small RNA libraries using three different
131 algorithms (Figure 2a). Reasonable cut-offs for *PHAS* loci detection were determined by
132 examining the phase score distributions in the three merged wild-type libraries when well-known
133 21 nt *PHAS* loci were analyzed (Figure S2). Of the 31,750 loci examined, only three (Figure S3;
134 Dataset S1) passed the *PHAS* loci algorithms consistently in all 8 libraries examined (Figure 2a).

135 We were interested in determining why these three 24 nt dominated loci consistently pass
136 the *PHAS*-detection algorithms. As phasiRNA precursors are known to be targeted by miRNAs,
137 we predicted whether or not any known *A. thaliana* miRNAs could target these three loci. We
138 were unable to find any obvious miRNA target sites at these loci. Although miRNA target sites
139 were not apparent at these loci, it's possible that other siRNAs might target them and initiate
140 siRNA phasing. We therefore attempted to determine the most common phase register at each of
141 these loci in the 8 different wild-type libraries (Figure S4). If these loci are truly phased, then we
142 would expect the same phase register to predominate in each library examined. While this was
143 the case for *TAS2* (a positive control), none of the three 24 nt clusters passing the *PHAS*-
144 detection algorithms had consistent phase registers (Figure S4). We also note that these three
145 clusters didn't have the majority of their mapped reads falling into a single phase register in any
146 of the libraries examined, as one would expect for a true *PHAS* locus (Figure S4).

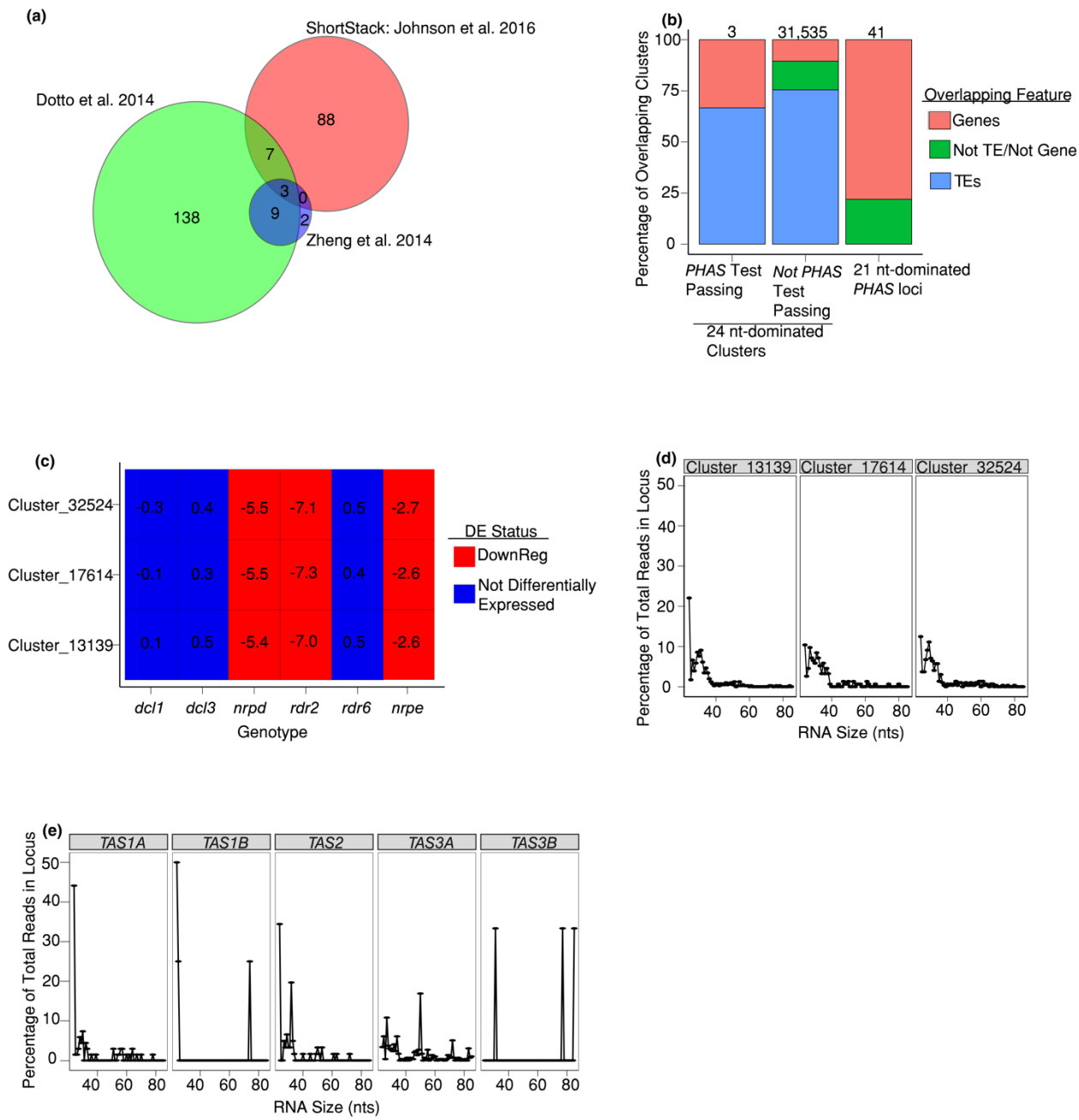


Figure 2. Properties of three *A. thaliana* 24 nt-dominated small RNA clusters that were called 'phased' by three different methods.

(a) Venn diagram shows numbers of 24 nt-dominated loci that were called 'phased' by the indicated algorithms.

(b) Percentage of the three *PHAS*-test passing clusters overlapping either genes or transposable elements. The percentage is calculated as: (number of clusters intersecting a feature/total number of clusters in category)*100.

155 (c) Accumulation of the three *PHAS*-test passing clusters in different genetic backgrounds.
156 Numbers represent the ratio of small RNA accumulation in the indicated genotypes over that in
157 corresponding wild-type library as computed by DESeq2. The differential expression status was
158 determined via DESeq2 at an FDR of 0.1.

159 (d) Percentage of short RNAs from *dcl2/dcl3/dcl4* triple mutant libraries by read length for the
160 three *PHAS*-test passing loci.

161 (e) Same as in Panel d, except for five known *A. thaliana TAS* loci.

162

163 We then determined where in the genome these loci were encoded. Similar to hc-siRNA
164 loci, these three loci primarily overlap with repeat- and transposable-elements (Figure 2b). In
165 comparison, many known *A. thaliana PHAS* loci are primarily derived from genes (Figure 2b).
166 Rice and maize annotated 24 nt *PHAS* loci are derived from long non-coding RNAs that are
167 encoded in regions of the genome that are devoid of protein-coding genes, transposable
168 elements, or repeats (Song et al., 2008; Zhai et al., 2015). We also note that sRNA accumulation
169 from these three loci is down-regulated in *nrpd1-3* (NRPD is the largest sub-unit of Pol IV), *nrpe*
170 (NRPE is the largest subunit of Pol V), and *rdr2* (Figure 2c). hc-siRNAs have the same genetic
171 requirements (Matzke et al., 2009), which strongly suggests that these three clusters produce hc-
172 siRNAs.

173 The Pol IV transcripts from which hc-siRNAs are derived are around 26-50 nts in length,
174 and accumulate to detectable levels in the *dcl2-1/3-1/4-2t* (*dcl234*) triple mutant (Ye et al., 2016;
175 Zhai et al., 2015). We therefore analyzed the lengths of reads mapping to our three putative 24
176 nt *PHAS* loci in *dcl234* triple mutants. We found most reads were less than 40 nts long,
177 indicating that these putative 24 nt *PHAS* loci are associated with short precursors similar to hc-
178 siRNA loci (Figure 2d). In comparison, reads mapping to the known *TAS* loci had a wider range
179 of sizes, ranging from 24-80 nts (Figure 2e). Overall, the size profile of the precursor RNAs
180 further delineate these three loci from known *PHAS* loci.

181 Finally, we examined the AGO-enrichments of small RNAs from these loci. Canonical
182 hc-siRNAs are loaded into AGO4 in order to repress other loci transcriptionally (Mi et al., 2008).
183 While sRNAs from our three putative 24 nt *PHAS* loci are not particularly enriched in AGO4-
184 immunoprecipitation libraries, known 21 nt *PHAS* loci are quite depleted in the same dataset
185 (Figure S5). We also note that sRNAs from the three putative 24 nt *PHAS* loci are depleted in
186 AGO1 immunoprecipitation libraries (Figure S5), probably owing to the lack of 21 nt sRNAs,
187 which AGO1 primarily loads, produced at these loci (Mi et al., 2008).

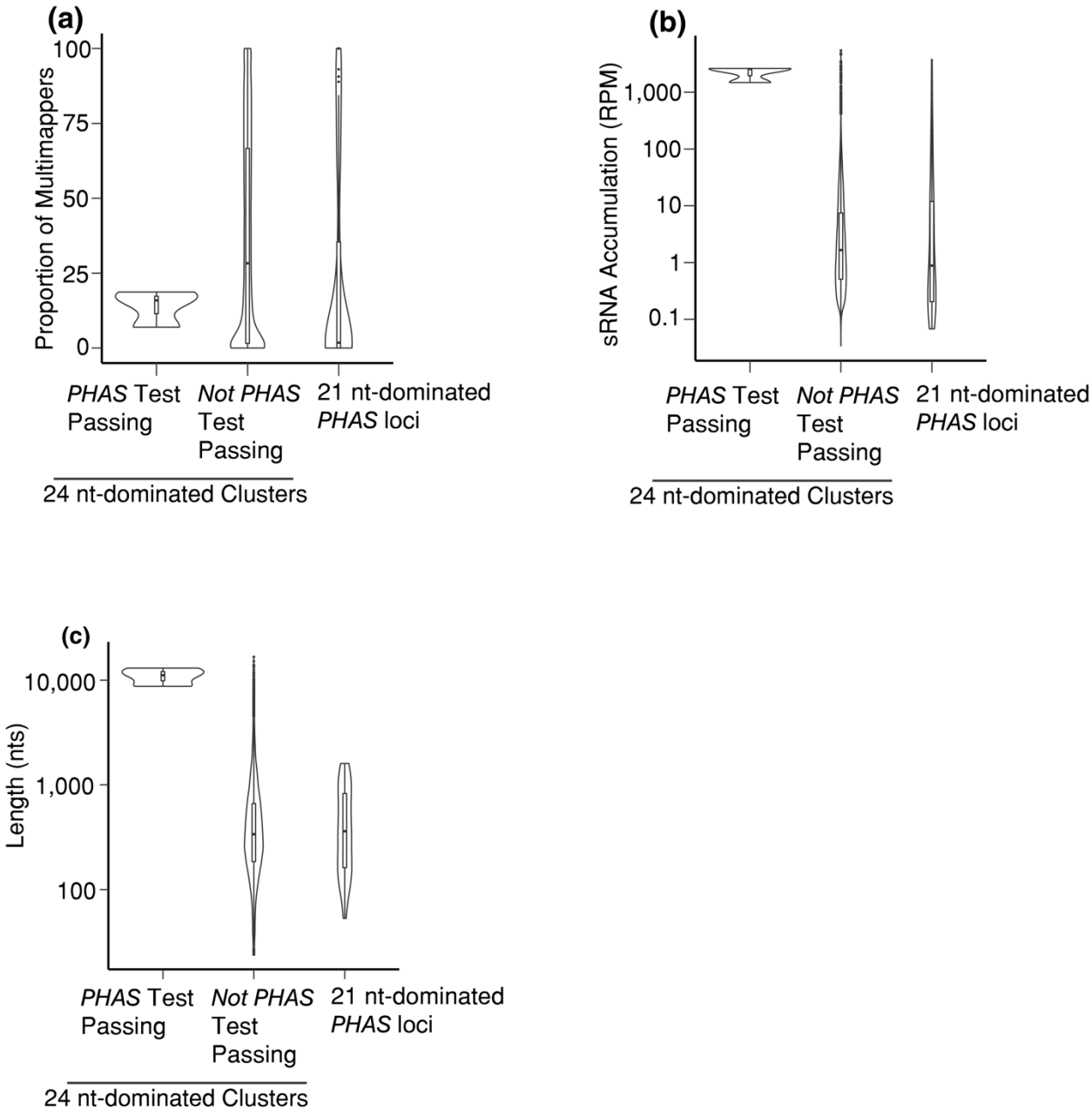
188

189 **The three *PHAS*-Test passing loci have distinct characteristics**

190 hc-siRNAs are known to produce small RNAs in a very imprecise manner, unlike the
191 largely precise processing of phasiRNAs (Axtell, 2013). We were interested in how three hc-
192 siRNAs could consistently pass three different *PHAS*-detecting algorithms. One simple
193 explanation is possible erroneous placement of ambiguously mapped reads. These multi-
194 mapping reads could cause the *PHAS*-detecting algorithms to overestimate the number of "in
195 phase" reads. We therefore determined the proportion of multi-mapping reads in each locus, but
196 these loci have a low proportion of multi-mapping reads compared to other 24 nt-dominated
197 clusters (Figure 3a). We note that these three loci are more highly expressed than most other 24
198 nt-dominated clusters and even known *PHAS* loci in *A. thaliana* (Figure 3b). These three
199 clusters accumulate to nearly 1,000 RPM, while most other 24 nt-dominated clusters and known
200 *PHAS* loci only produce around 1 RPM (Figure 3b). Another interesting feature of these loci is
201 their length. All three of these loci are around 10,000 nts in length, far greater than the 200-800
202 nt length of canonical *PHAS* loci and most other 24 nt-dominated loci (Figure 3c). These results

203 indicate that these loci are simply highly expressed and particularly long hc-siRNA producing

204 clusters (Figure S3).



205

206

207 **Figure 3. The three *PHAS*-Test passing small RNA clusters have distinct properties**
208 **compared to other 24 nt-dominated Clusters.**

209 (a) The proportion of ambiguously mapped reads produced by the three *PHAS*-passing clusters
210 are compared to other 24 nt-dominated clusters and clusters intersecting known *A. thaliana* 21
211 nt-dominated *PHAS* loci. The width of the density plot shows the frequency. The inset boxes
212 show medians (horizontal lines), the 1st-3rd quartile range (boxes), the 95% confidence of
213 medians (notches), other data out to 1.5 times the interquartile range (whiskers) and outliers
214 (dots).

215 (b) Same as panel a except showing small RNA accumulation.

216 (c) Same as panel a except showing small RNA cluster length.

217

218 **A small number of *A. thaliana rdr6*-dependent, 24 nt-dominated siRNA loci exist, but are**

219 **not phased**

220 Canonical monocot 24 nt-dominated *PHAS* loci are produced in part through the
221 biochemical activity of RDR6 (Zhai et al., 2015). We therefore used an alternative strategy to
222 identify possible 24 nt *PHAS* loci by first identifying *rdr6*-dependent, 24 nt-dominated siRNA
223 loci. We found 18 such loci (Figure S6; Dataset S1). None of these loci were consistently
224 phased in the eight wild-type *A. thaliana* libraries (Table S2) tested according to any of the three
225 algorithms we used (Dataset S3). We determined the genetic dependencies of these loci and
226 found that these loci are down-regulated in *nrpd, nrpe,* and *rdr2* backgrounds (Figure 4).
227 Furthermore, these loci either overlap transposable elements or are mostly found in otherwise
228 intergenic regions (Figure S7a). As these are typical features of hc-siRNAs, it's possible that hc-
229 siRNAs were erroneously placed at these loci. Again, we examined the proportion of
230 ambiguously mapped reads in these *rdr6*-dependent, 24 nt-dominated clusters compared to other
231 24 nt-dominated clusters and found that these loci had low proportions of multi-mapping reads
232 (Figure S7b). Our result argues against this hypothesis.

233 *rdr6*-dependent, 24 nt-dominated clusters are strongly expressed relative to other 24 nt-
234 dominated clusters (Figure S7c). These *rdr6*-dependent clusters have median expression level of
235 nearly 100 RPM compared to the nearly 1 RPM median expression level of the other 24 nt-
236 dominated clusters (Figure S7c). This result is similar to the three *A. thaliana* clusters that
237 passed our *PHAS*-detection algorithms (Figure 3b). However, the lengths of the *rdr6*-dependent,
238 24 nt-dominated clusters are similar to other 24 nt-dominated clusters (Figure S7d). Overall
239 while we do find evidence for a small number of *rdr6*-dependent, 24 nt dominated siRNAs in *A.*
240 *thaliana*, they are not phased, nor do they appear readily discernable from more typical hc-
241 siRNA loci. Thus overall we have found no evidence of 24 nt dominated *PHAS* loci in *A.*
242 *thaliana*.

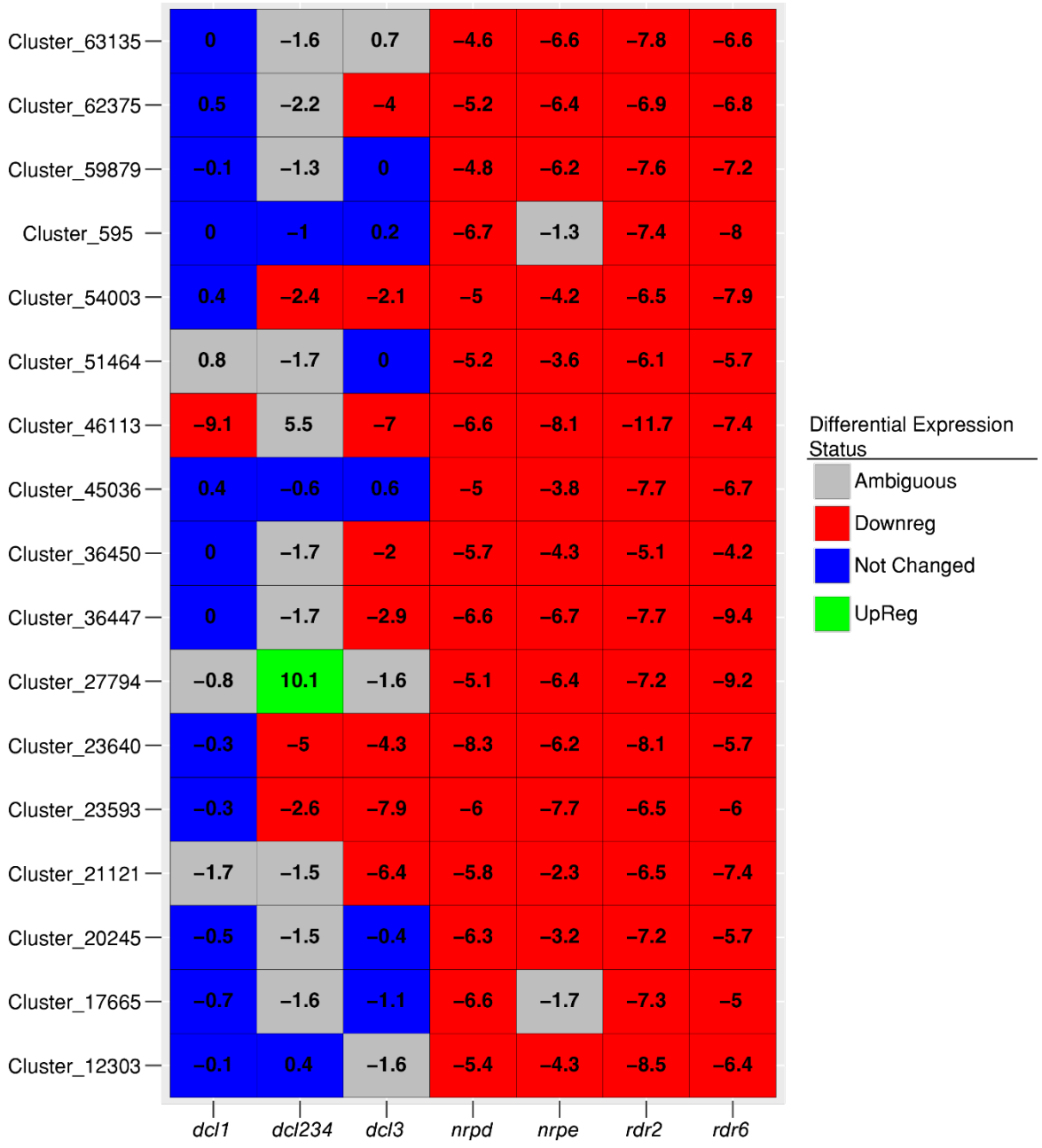


Figure 4. Accumulation of 24 nt-dominated, *rdr6*-dependent small RNA clusters in different mutant backgrounds.

Numbers represent the log2-transformed ratios of small RNA accumulation in the indicated genotypes over that in the corresponding wild-type libraries as computed by DESeq2. The differential expression status was determined via DESeq2 at an FDR of 0.1.

Erroneous detection of 24nt-dominated *PHAS* loci occurs in other land plant species

251 We were interested in determining if erroneous annotation of *PHAS* loci was unique to *A.*
252 *thaliana* small RNAs or if these results could be replicated in other species. We therefore
253 interrogated publicly available *Brassica rapa*, *Cucumis sativus*, *Phaseolus vulgaris*, and *Solanum*
254 *tuberosum* small RNA libraries (all eudicots) using the three algorithms, searching for putative
255 24 nt-dominated *PHAS* loci. We specifically chose these four species because they all lacked a
256 potential *MIR2275* homolog (Figure 1) and thus any 24 nt-dominated *PHAS* cluster is likely a
257 false positive. Each species had 24 nt-dominated small RNAs that were misannotated as *PHAS*
258 loci (Figure S8). We first determined where in the genome the 24 nt-dominated loci are
259 encoded. Like in *A. thaliana,* 24 nt-dominated loci that passed the *PHAS*-detection algorithm
260 seem to come predominantly from transposable elements. This was true in all the species
261 examined except for *P. vulgaris* (Figure 5a). Curiously, the *PHAS*-test passing clusters had a
262 slightly higher proportion of ambiguously mapped reads compared to other 24 nt-dominated loci
263 in *B. rapa*, *P. vulgaris*, and *S, tuberosum* (Figure 5b). For *C. sativus*, the proportion of multi-
264 mapping reads in *PHAS*-test passing 24 nt-dominated clusters was substantially higher than other
265 24 nt-dominated clusters (Figure 5b). It's still unlikely that multi-mapping reads contribute
266 significantly to phasing at these loci because their proportions are similar in three of the four
267 species tested (Figure 5b).

268 In the four species tested, the *PHAS*-test passing 24 nt-dominated loci had both
269 significantly greater lengths (Figure 5c) and expression on average (Figure 5d). We observed
270 this same trend in *A. thaliana* (Figure 3b-c), which suggests that long lengths and high
271 expression levels of 24 nt-dominated loci are conducive to *PHAS* loci mis-annotations.

272

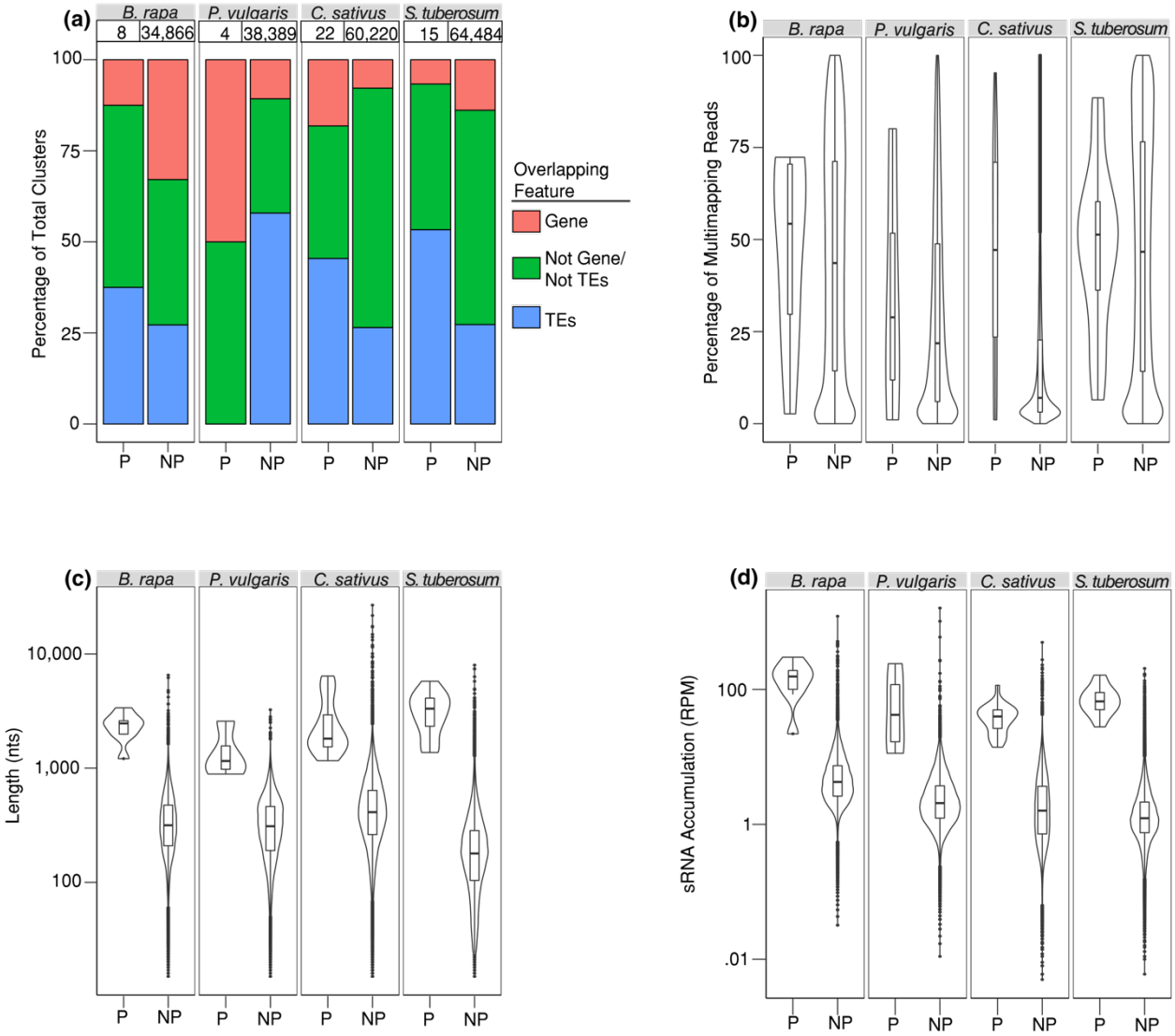


Figure 5. Four divergent species also contain 24 nt-dominated clusters that passed the three *PHAS*-detecting algorithms.

(a) Percentage of 24 nt-dominated Clusters overlapping genes, transposable elements, or neither. The species name is shown in the top grey boxes. P: *PHAS*-test passing locus; NP: Not *PHAS*-test passing locus.

(b) Same as panel a except showing the proportion of multi-mapping reads.

(c) Same as panel a except showing length of the small RNA cluster.

(d) Same as panel a except showing small RNA accumulation.

Discussion

284 ***rdr6*-dependent, 24 nt-dominated clusters in the *A. thaliana* genome**

285 We found a handful of *rdr6*-dependent, 24 nt-dominated clusters to be encoded in the *A.*
286 *thaliana* (TAIR 10) genome. However, these clusters have the same genetic dependencies as hc-
287 siRNAs (Figure 4), are not phased, and are derived from transposable elements (Figure S7a).
288 The only other known *rdr6*-dependent, transposon-overlapping small RNA clusters encoded in
289 *A. thaliana* are Epigenetically Activated Short Interfering RNAs (easiRNAs) (Creasey et al.,
290 2014). easiRNAs are derived from transcriptionally active transposable elements that are
291 targeted and cleaved by miRNAs (Creasey et al., 2014). Through the biochemical actions of
292 RDR6 and DCL4, the miRNA cleavage product is converted into 21-22 nt double-stranded RNA
293 molecules (Creasey et al., 2014). Furthermore, these easiRNAs are thought to direct initial
294 repression of transposons (He et al., 2015). It's possible that these *rdr6*-dependent, 24 nt-
295 dominated clusters could represent a transitory stage of hc-siRNA targeting, in which the genetic
296 machinery of hc-siRNAs is used but with a remaining initial dependency on RDR6.

297

298 **Possible losses of the *MIR2275*-generated phasing in eudicots**

299 We examined all Phytozome (ver 12.1)-curated angiosperm genomes (Goodstein et al.,
300 2012) for the presence of *MIR2275* homologs. *MIR2275* shows remarkable conservation in
301 monocots (Figure 1), showing loss only in *Spirodela polyrhiza* (Wang et al., 2014) and *Zostera*
302 *marina* (Olsen et al., 2016). This could be because these are aquatic plants with reduced
303 morphology, and thus there is relaxed selective pressure for *MIR2275*. Another explanation is
304 that the genome assemblies for these species could be incomplete. The lack of *MIR2275* in
305 eudicots was more extensive (Figure 1), with no plants in the Brassicales clade containing a

306 verified *MIR2275* homolog (Figure 1). Interestingly, *Aquilegia coerulea*, a basal eudicot
307 (Sharma & Kramer, 2014), contained a *MIR2275* homolog (Figure 1; Figure S1) suggesting that
308 the last common ancestor for eudicots may have contained *MIR2275*, and that the lack of
309 detected putative *MIR2275* homologs in many eudicot plant species could be due to a derived
310 loss of *MIR2275*.

311 We only examined whether or not a possible *MIR2275* homolog could be detected
312 through sequence homology, if its predicted secondary structure is conducive to *MIRNA*
313 processing (Figure S1), and in some species whether or not the putative *MIR2275* homolog was
314 expressed in available small RNA libraries (Figure 1). We did not, however, determine if these
315 species produce true 24 nt-dominated phasiRNAs. It's possible that the potential *MIR2275*
316 homolog in these species is an orphaned *MIRNA* or has taken on a new function. Furthermore,
317 although some putative *MIR2275* homologs could be found, we note that they contain
318 mismatches in the stem of the secondary structure which may hinder miRNA biogenesis (Figure
319 S1). Further research is necessary to determine if these species truly encode 24 nt-dominated
320 *PHAS* loci.

321

322 **The difficulties of annotating 24 nt-dominated *PHAS* loci**

323 The discovery of 24 nt-dominated *PHAS* loci in maize (Zhai et al., 2015), rice (Song et
324 al., 2008), asparagus, daylily, and lily (Kakrana et al., 2018) opened up the possibility that these
325 loci exist in other species, even distantly-related eudicots. However, 24 nt-dominated hc-siRNA
326 clusters are widespread in land plant genomes (Ghildiyal & Zamore, 2009). This is true even in
327 *A. thaliana*, although 24 nt sRNAs are thought to repress transposable elements (Matzke et al.,

328 2009) and only about 20-30% of the *A. thaliana* genome consists of transposon/repetitive
329 elements (Barakat et al., 1998). In contrast, other plant genomes consist of as much as 80%
330 transposons/repetitive elements (Springer et al., 2009). The sheer number of 24 nt-dominated
331 sRNA loci could mean that several of them could meet various annotation criteria simply by
332 chance; this is something we have previously noted to occur in natural antisense siRNA
333 annotation (Polydore & Axtell, 2018).

334 The supposed 24 nt-dominated *PHAS* loci examined in this study consistently showed
335 higher levels of expression than other 24 nt-dominated sRNA loci (Figure 3b). This isn't
336 particularly surprising because the accumulation of reads (in- and out-of-phase) is a factor in
337 most *PHAS*-detection algorithms (Figure S2a-c). However, it seems that particularly highly
338 expressed 24 nt-dominated small RNAs are able to consistently pass *PHAS*-detection algorithms
339 because of this. It's possible that the sheer number of reads produced at these loci means that
340 these loci produce enough "in-phase" reads by chance to score highly in *PHAS*-detecting
341 algorithms.

342

343 **Annotating 24 nt-dominated *PHAS* loci in the future**

344 Our study demonstrates that when annotating novel 24 nt-dominated *PHAS*, more than
345 utilizing *PHAS*-detecting algorithms is necessary for robust annotation. First, examining
346 different available mutants, especially of genes involved in small RNA biogenesis, can be critical
347 in determining the type of small RNA in question (Figure 2c). All types of phasiRNAs are
348 known to be reliant on the biochemical activity of RDR6 and Pol II (Cuperus et al., 2010; Song
349 et al., 2008; Zhai et al., 2015). While it's entirely possible that non-canonical phasiRNAs that

350 are reliant on RDR2 and Pol IV may be described eventually, such a study must verify with
351 robust methodologies that the *PHAS* loci aren't false positives due to the sheer number of Pol
352 IV/RDR2 –dependent reads produced in the land plant genome.

353 We searched for *MIR2275* homologs because miR2275 is the only miRNA known to
354 trigger 24 nt phasiRNAs. If a species lacks a *MIR2275* homolog, then one should be skeptical of
355 any 24 nt-dominated small RNA cluster that is annotated as *PHAS* locus in that species. An
356 organism producing a mature miR2275 small RNA homolog is not sufficient to show that 24 nt-
357 dominated *PHAS* loci are produced in that species. One should also take care to ascertain if any
358 24 nt-dominated small RNA cluster called as a *PHAS* locus also contains a miR2275 target site
359 that is "in phase" with the phasiRNAs produced from the transcript. miR2275 is also very
360 specifically expressed in the tapetum of male floral tissue. In asparagus, 24 nt-dominated
361 phasiRNAS have also been shown to be expressed in female floral tissue. However, 24 nt-
362 dominated phasiRNAs have not been demonstrated to be expressed outside of reproductive tissue
363 as of yet. Therefore, any 24 nt-dominated *PHAS* loci annotated in libraries not produced from
364 reproductive tissues are more likely to be mis-annotations.

365 Despite the fact that several small RNA clusters in *A. thaliana* are able to consistently
366 pass the *PHAS*-detecting algorithms in different libraries, reproducibility among distinct small
367 RNA libraries is of utmost importance. It could also be useful to employ several different
368 *PHAS*-test algorithms as well. In *A. thaliana*, three loci were able to consistently pass our
369 *PHAS*-detecting algorithms (Figure S3), but the number of clusters that each algorithm detected
370 on its own was generally much higher (Figure 2a). Had we not used stringent *PHAS*-detecting
371 methods, it would've been plausible to assume we had found 24 nt-dominated *PHAS* loci in *A.*
372 *thaliana* based on the number of clusters that consistently pass alone. Utilizing multiple small

373 RNA libraries should become easier to accomplish as more and more libraries in different
374 treatments/genotypes become available for diverse species. Certain *PHAS*-detecting programs,
375 such as *PHASIS* (Kakrana et al., 2017), automatically evaluate potential small RNA clusters in
376 different libraries individually before merging the results of each library.

377 It's not always possible to determine the small RNA that targets the phasiRNA precursor
378 transcript. Indeed, several reads may be predicted to target a certain transcript by chance due to
379 the sheer number of unique reads in a small RNA library. However, due to hydrolysis of the
380 precursor transcript following small RNA-mediated targeting, phasiRNAs have well-defined
381 termini from which they are produced (Axtell et al., 2006; Cuperus et al., 2010). Therefore, a
382 true *PHAS* loci should have a large proportion of its reads reproducibly falling into a particular
383 phase register. This is what we observed with well-characterized *PHAS* locus*, TAS2*, but not
384 with the three loci that pass our rigorous *PHAS*-annotation regime in *A. thaliana* (Figure S4).

385 As 21-22 nt-dominated *PHAS* loci are far more common in land plants, especially outside
386 of the monocots, one can conservatively limit their discovery of new *PHAS* loci to 21-22 nt-
387 dominated small RNAs and not employ as rigorous methodology as the ones used in this study.
388 24 nt-dominated *PHAS* loci on the other hand seem to be have undergone loss in many
389 angiosperms. However, even in the species in which they are conserved, they seem to play very
390 specific, reproductive-associated roles as evidenced by their expression patterns. Great caution
391 should be used for annotating 24 nt-dominated *PHAS* loci in the future.

392

393 **Materials and Methods**

394 **Finding potential *MIR2275* homologs in angiosperms**

395 The Phytozome (ver 12.1)-curated angiosperm genomes sequences were downloaded.
396 The mature *O. sativa* miR2275a sequence was downloaded from miRBase (ver 21.) (Griffiths-
397 Jones et al., 2008) and searched against all the other genomes using Basic Local Alignment
398 Search Tool (BLAST ver. 2.7.1) with options –penalty -3, -gapopen 5, -gapextend 2, and –evalue
399 10 (Altschul et al., 1990). Only results with at least 85 % sequence identity covering at least 18
400 nucleotides of miR2275a were kept for further analysis.

401 In order to determine the predicted secondary structures of the BLAST results of interest,
402 the sequences corresponding to the BLAST result, plus 200 nucleotides upstream and
403 downstream, was extracted from the *A. thaliana* (TAIR10) genome. The secondary structures of
404 the sequences were predicted using the mFOLD web server (Zuker, 2003) and visually inspected
405 to determine if the sequence formed a hairpin structure consistent with accepted norms for
406 miRNA biogenesis (Axtell & Meyers, 2018).

407 We downloaded publicly available small RNA libraries (Dataset S2) for *Brassica rapa*,
408 *Cucumis sativus*, *Phaseolus vulgaris*, and *Solanum tuberosum* and merged and aligned them
409 against their respective genomes using ShortStack (ver 3.8.1) with default options. The genomes
410 used were ver 1.0 for *B. rapa* (Wang et al., 2011), ver 1.0 for *P. vulgaris* (Schmutz et al., 2014),
411 ver. 2 for *C. sativus* (Huang et al., 2009), and dm_v404 for *S. tuberosum* (Hardigan et al., 2016).
412 The merged libraries were collapsed to non-redundant reads and investigated using CD-HIT (ver
413 4.6.8) using the options -n 4, -d 0, and -g 1. The *O. sativa* miR2275a sequence was used as a
414 query.

415

416 **Determination of potential 24 nt-dominated *PHAS* loci in different wild-type, inflorescence**
417 **libraries**

418 Wild-type biological triplicate small RNA libraries (GSE105262) (Polydore & Axtell,
419 2018) were merged and aligned against the *A. thaliana* (TAIR 10) genome using ShortStack (ver
420 3.8) (Johnson et al., 2016) using 25 known *A. thaliana PHAS* loci as a query file (Table S1).
421 Three distinct *PHAS* –detecting algorithms (Dotto et al., 2014; Johnson et al., 2016; Zheng et al.,
422 2014) were used to determine phase scores in the merged run. These scores were used as the
423 basis to determine cutoffs for calling significantly phased loci (Figure S2).

424 Wild-type and *rdr1-1/2-1/6-15* (*rdr1/2/6*) triple mutant small RNA libraries
425 (GSE105262) (Polydore & Axtell, 2018) were aligned against the *A. thaliana* (TAIR 10) genome
426 using ShortStack (ver 3.2) with option –pad 75 (Table S2). Differential expression to determine
427 down-regulated clusters was performed as previously described (Polydore & Axtell, 2018).
428 Down-regulated, 24 nt-dominated small RNAs clusters were catalogued into a list. Eight wild-
429 type, inflorescence small RNA libraries (Table S2) were run individually against the *A. thaliana*
430 (TAIR 10) genome utilizing ShortStack (ver 3.8.1) using the results of our wild-type and
431 *rdr1/2/6* libraries run as a query file. The phase scores of loci corresponding to the down-
432 regulated, 24 nt-dominated small RNAs clusters were evaluated using the binary sequence
433 alignment (BAM)-formatted alignments from each run. ShortStack and an in-house Python
434 script were used to perform the three phase score calculations. For *B. rapa*, *C. sativus*, *P.*
435 *vulgaris*, and *S. tuberosum,* the previous Shortstack merged small RNA alignments were
436 analyzed in the same way.

437

438 **AGO Immunoprecipitation, genetic dependency, and properties of loci analyses**

439 Calculating the lengths, proportion of multi-mapping reads, small RNA expression levels
440 (in Reads Per Million (RPM)), determining the genetic dependencies, and the AGO enrichments
441 of various loci were performed as previously described (Polydore & Axtell, 2018).

442

443 **Examining *PHAS*-test passing loci for potential miRNA targeting**

444 Sequences corresponding to the *A. thaliana* loci that consistently passed the *PHAS*-
445 detection algorithms plus 200 base pairs up-stream and down-stream were extracted from the *A.*
446 *thaliana* genome**.** Mature miRNA sequences were downloaded from miRBase (ver 21) and
447 aligned against these sequences using the Generic Small RNA-Transcriptome Aligner. A
448 miRNA was considered to be potentially targeting a sequence if it aligned with an Allen et. al.
449 (2015) score of 3 or less.

450

451 **Acknowledgements**

452 This research was funded by NSF Award 1339207 to MJA, the NIH Bioinformatics and
453 Statistics (CBIOS) Predoctoral Training Program (1T32GM102057- 0A1), and the Henry W.
454 Popp Graduate Assistantship Fund to SP.

455

456 **Literature cited**

457 **Allen, E., Xie, Z., Gustafson, A. M., & Carrington, J. C.** (2005). microRNA-directed phasing
458 during trans-acting siRNA biogenesis in plants. *Cell*, *121*(2), 207–221.
459 https://doi.org/10.1016/j.cell.2005.04.004

460 **Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J.** (1990). Basic local
461 alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.
462 https://doi.org/10.1016/S0022-2836(05)80360-2
463 **Axtell, M. J.** (2010). A Method to Discover Phased siRNA Loci. In *Plant MicroRNAs* (pp. 59–
464 70). Humana Press. https://doi.org/10.1007/978-1-60327-005-2_5
465 **Axtell, M. J.** (2013). Classification and Comparison of Small RNAs from Plants. *Annual Review*
466 *of Plant Biology*, *64*(1), 137–159. https://doi.org/10.1146/annurev-arplant-050312-
467 120043
468 **Axtell, M. J.** (2015). Non-coding RNAs: The small mysteries of males. *Nature Plants*, *1*(5),
469 15055. https://doi.org/10.1038/nplants.2015.55
470 **Axtell, M. J., Jan, C., Rajagopalan, R., & Bartel, D. P.** (2006). A two-hit trigger for siRNA
471 biogenesis in plants. *Cell*, *127*(3), 565–577. https://doi.org/10.1016/j.cell.2006.09.032
472 **Axtell, M. J., & Meyers, B. C.** (2018). Revisiting criteria for plant miRNA annotation in the era
473 of big data. *The Plant Cell*, tpc.00851.2017. https://doi.org/10.1105/tpc.17.00851
474 **Barakat, A., Matassi, G., & Bernardi, G.** (1998). Distribution of genes in the genome of
475 Arabidopsis thaliana and its implications for the genome organization of plants.
476 *Proceedings of the National Academy of Sciences of the United States of America*,
477 *95*(17), 10044–10049.
478 **Baulcombe, D.** (2004). RNA silencing in plants. *Nature*, *431*(7006), 356–363.
479 https://doi.org/10.1038/nature02874
480 **Baumberger, N., & Baulcombe, D. C.** (2005). Arabidopsis ARGONAUTE1 is an RNA Slicer
481 that selectively recruits microRNAs and short interfering RNAs. *Proceedings of the*

482 *National Academy of Sciences*, *102*(33), 11928–11933.
483 https://doi.org/10.1073/pnas.0505461102
484 **Boualem, A., Laporte, P., Jovanovic, M., Laffont, C., Plet, J., Combier, J.-P., & Frugier, F.**
485 (2008). MicroRNA166 controls root and nodule development in Medicago truncatula.
486 *The Plant Journal: For Cell and Molecular Biology*, *54*(5), 876–887.
487 https://doi.org/10.1111/j.1365-313X.2008.03448.x
488 **Chapman, E. J., & Carrington, J. C.** (2007). Specialization and evolution of endogenous small
489 RNA pathways. *Nature Reviews Genetics*, *8*(11), 884–896.
490 https://doi.org/10.1038/nrg2179
491 **Chen, H.-M., Chen, L.-T., Patel, K., Li, Y.-H., Baulcombe, D. C., & Wu, S.-H.** (2010). 22-
492 nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proceedings of the*
493 *National Academy of Sciences of the United States of America*, *107*(34), 15269–15274.
494 https://doi.org/10.1073/pnas.1001738107
495 **Chen, H.-M., Li, Y.-H., & Wu, S.-H.** (2007). Bioinformatic prediction and experimental
496 validation of a microRNA-directed tandem trans-acting siRNA cascade in Arabidopsis.
497 *Proceedings of the National Academy of Sciences*, *104*(9), 3318–3323.
498 https://doi.org/10.1073/pnas.0611119104
499 **Creasey, K. M., Zhai, J., Borges, F., Van Ex, F., Regulski, M., Meyers, B. C., &**
500 **Martienssen, R. A.** (2014). miRNAs trigger widespread epigenetically activated siRNAs
501 from transposons in Arabidopsis. *Nature*, *508*(7496), 411–415.
502 https://doi.org/10.1038/nature13069
503 **Cuperus, J. T., Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Burke, R. T., Takeda, A., &**
504 **Carrington, J. C.** (2010). Unique functionality of 22-nt miRNAs in triggering RDR6-

505 dependent siRNA biogenesis from target transcripts in *Arabidopsis*. *Nature Structural &*
506 *Molecular Biology*, *17*(8), 997–1003. https://doi.org/10.1038/nsmb.1866
507 **Dotto, M. C., Petsch, K. A., Aukerman, M. J., Beatty, M., Hammell, M., & Timmermans,**
508 **M. C. P.** (2014). Genome-Wide Analysis of leafbladeless1-Regulated and Phased Small
509 RNAs Underscores the Importance of the TAS3 ta-siRNA Pathway to Maize
510 Development. *PLOS Genetics*, *10*(12), e1004826.
511 https://doi.org/10.1371/journal.pgen.1004826
512 **Elvira-Matelot, E., Hachet, M., Shamandi, N., Comella, P., Sáez-Vásquez, J., Zytnicki, M.,**
513 **& Vaucheret, H.** (2016). Arabidopsis RNASE THREE LIKE2 Modulates the Expression
514 of Protein-Coding Genes via 24-Nucleotide Small Interfering RNA-Directed DNA
515 Methylation. *The Plant Cell Online*, *28*(2), 406–425.
516 https://doi.org/10.1105/tpc.15.00540
517 **Ender, C., & Meister, G.** (2010). Argonaute proteins at a glance. *J Cell Sci*, *123*(11), 1819–
518 1823. https://doi.org/10.1242/jcs.055210
519 **Fei, Q., Xia, R., & Meyers, B. C.** (2013). Phased, Secondary, Small Interfering RNAs in
520 Posttranscriptional Regulatory Networks. *The Plant Cell Online*, *25*(7), 2400–2415.
521 https://doi.org/10.1105/tpc.113.114652
522 **Ghildiyal, M., & Zamore, P. D.** (2009). Small silencing RNAs: an expanding universe. *Nature*
523 *Reviews. Genetics*, *10*(2), 94–108. https://doi.org/10.1038/nrg2504
524 **Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., & Rokhsar, D.**
525 **S.** (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids*
526 *Research*, *40*(Database issue), D1178–D1186. https://doi.org/10.1093/nar/gkr944

527 **Grant-Downton, R., Hafidh, S., Twell, D., & Dickinson, H. G.** (2009). Small RNA Pathways
528 Are Present and Functional in the Angiosperm Male Gametophyte. *Molecular Plant*,
529 *2*(3), 500–512. https://doi.org/10.1093/mp/ssp003
530 **Griffiths-Jones, S., Saini, H. K., van Dongen, S., & Enright, A. J.** (2008). miRBase: tools for
531 microRNA genomics. *Nucleic Acids Research*, *36*(suppl_1), D154–D158.
532 https://doi.org/10.1093/nar/gkm952
533 **Groth, M., Stroud, H., Feng, S., Greenberg, M. V. C., Vashisht, A. A., Wohlschlegel, J. A.,**
534 **& Ausin, I.** (2014). SNF2 chromatin remodeler-family proteins FRG1 and -2 are
535 required for RNA-directed DNA methylation. *Proceedings of the National Academy of*
536 *Sciences of the United States of America*, *111*(49), 17666–17671.
537 https://doi.org/10.1073/pnas.1420515111
538 **Guo, Q., Qu, X., & Jin, W.** (2015). PhaseTank: genome-wide computational identification of
539 phasiRNAs and their regulatory cascades. *Bioinformatics*, *31*(2), 284–286.
540 https://doi.org/10.1093/bioinformatics/btu628
541 **Hardigan, M. A., Crisovan, E., Hamiltion, J. P., Kim, J., Laimbeer, P., Leisner, C. P., &**
542 **Buell, C. R.** (2016). Genome reduction uncovers a large dispensable genome and
543 adaptive role for copy number variation in asexually propagated Solanum tuberosum. *The*
544 *Plant Cell*, TPC2015–00538–RA. https://doi.org/10.1105/tpc.15.00538
545 **He, H., Yang, T., Wu, W., & Zheng, B.** (2015). Small RNAs in pollen. *Science China Life*
546 *Sciences*, *58*(3), 246–252. https://doi.org/10.1007/s11427-015-4800-0
547 **Howell, M. D., Fahlgren, N., Chapman, E. J., Cumbie, J. S., Sullivan, C. M., Givan, S. A.,**
548 **& Carrington, J. C.** (2007). Genome-Wide Analysis of the RNA-DEPENDENT RNA
549 POLYMERASE6/DICER-LIKE4 Pathway in Arabidopsis Reveals Dependency on

550 miRNA- and tasiRNA-Directed Targeting. *The Plant Cell*, *19*(3), 926–942.
551 https://doi.org/10.1105/tpc.107.050062
552 **Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., & Li, S.** (2009). The genome of the
553 cucumber, *Cucumis sativus* L. *Nature Genetics*, *41*(12), 1275–1281.
554 https://doi.org/10.1038/ng.475
555 **Jeong, D.-H., Thatcher, S. R., Brown, R. S. H., Zhai, J., Park, S., Rymarquis, L. A., &**
556 **Green, P. J.** (2013). Comprehensive Investigation of MicroRNAs Enhanced by Analysis
557 of Sequence Variants, Expression Patterns, ARGONAUTE Loading, and Target
558 Cleavage1[W][OA]. *Plant Physiology*, *162*(3), 1225–1245.
559 https://doi.org/10.1104/pp.113.219873
560 **Johnson, N. R., Yeoh, J. M., Coruh, C., & Axtell, M. J.** (2016). Improved Placement of Multi-
561 mapping Small RNAs. *G3: Genes, Genomes, Genetics*, *6*(7), 2103–2111.
562 https://doi.org/10.1534/g3.116.030452
563 **Kakrana, A., Li, P., Patel, P., Hammond, R., Anand, D., Mathioni, S., & Meyers, B.** (2017).
564 PHASIS: A computational suite for de novo discovery and characterization of phased,
565 siRNA-generating loci and their miRNA triggers. *bioRxiv*, 158832.
566 https://doi.org/10.1101/158832
567 **Kakrana, A., Mathioni, S. M., Huang, K., Hammond, R., Vandivier, L., Patel, P., &**
568 **Meyers, B. C.** (2018). Plant 24-nt reproductive phasiRNAs from intramolecular duplex
569 mRNAs in diverse monocots. *Genome Research*, gr.228163.117.
570 https://doi.org/10.1101/gr.228163.117
571 **Komiya, R., Ohyanagi, H., Niihama, M., Watanabe, T., Nakano, M., Kurata, N., &**
572 **Nonomura, K.-I.** (2014). Rice germline-specific Argonaute MEL1 protein binds to

phasiRNAs generated from more than 700 lincRNAs. *The Plant Journal*, *78*(3), 385–397. https://doi.org/10.1111/tpj.12483

Kutter, C., Schöb, H., Stadler, M., Meins, F., & Si-Ammour, A. (2007). MicroRNA-Mediated Regulation of Stomatal Development in Arabidopsis. *The Plant Cell*, *19*(8), 2417–2429. https://doi.org/10.1105/tpc.107.050377

Laufs, P., Peaucelle, A., Morin, H., & Traas, J. (2004). MicroRNA regulation of the CUC genes is required for boundary size control in Arabidopsis meristems. *Development*, *131*(17), 4311–4322. https://doi.org/10.1242/dev.01320

Lee, T., Gurazada, S. G. R., Zhai, J., Li, S., Simon, S. A., Matzke, M. A., & Meyers, B. C. (2012). RNA polymerase V-dependent small RNAs in Arabidopsis originate from small, intergenic loci including most SINE repeats. *Epigenetics*, *7*(7), 781–795. https://doi.org/10.4161/epi.20290

Li**, S., Vandivier, L. E., Tu, B., Gao, L., Won, S. Y., Li, S., & Chen, X.** (2014). Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in Arabidopsis reveals features and regulation of siRNA biogenesis. *Genome Research*, gr.182238.114. https://doi.org/10.1101/gr.182238.114

Matzke, M., Kanno, T., Daxinger, L., Huettel, B., & Matzke, A. J. (2009). RNA-mediated chromatin-based silencing in plants. *Current Opinion in Cell Biology*, *21*(3), 367–376. https://doi.org/10.1016/j.ceb.2009.01.025

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., & Qi, Y. (2008). Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5′ Terminal Nucleotide. *Cell*, *133*(1), 116–127. https://doi.org/10.1016/j.cell.2008.02.034

595 **Millar, A. A., & Waterhouse, P. M.** (2005). Plant and animal microRNAs: similarities and
596 differences. *Functional & Integrative Genomics*, *5*(3), 129–135.
597 https://doi.org/10.1007/s10142-005-0145-2
598 **Montgomery, T. A., Howell, M. D., Cuperus, J. T., Li, D., Hansen, J. E., Alexander, A. L.,**
599 **& Carrington, J. C.** (2008). Specificity of ARGONAUTE7-miR390 Interaction and
600 Dual Functionality in TAS3 Trans-Acting siRNA Formation. *Cell*, *133*(1), 128–141.
601 https://doi.org/10.1016/j.cell.2008.02.033
602 **Olsen, J. L., Rouzé, P., Verhelst, B., Lin, Y.-C., Bayer, T., Collen, J., & Peer, Y. V. de.**
603 (2016). The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the
604 sea. *Nature*, *530*(7590), 331–335. https://doi.org/10.1038/nature16548
605 **Ono, S., Liu, H., Tsuda, K., Fukai, E., Tanaka, K., Sasaki, T., & Nonomura, K.-I.** (2018).
606 EAT1 transcription factor, a non-cell-autonomous regulator of pollen production,
607 activates meiotic small RNA biogenesis in rice anther tapetum. *PLOS Genetics*, *14*(2),
608 e1007238. https://doi.org/10.1371/journal.pgen.1007238
609 **Onodera, Y., Haag, J. R., Ream, T., Nunes, P. C., Pontes, O., & Pikaard, C. S.** (2005). Plant
610 Nuclear RNA Polymerase IV Mediates siRNA and DNA Methylation-Dependent
611 Heterochromatin Formation. *Cell*, *120*(5), 613–622.
612 https://doi.org/10.1016/j.cell.2005.02.007
613 **Panda, K., Ji, L., Neumann, D. A., Daron, J., Schmitz, R. J., & Slotkin, R. K.** (2016). Full-
614 length autonomous transposable elements are preferentially targeted by expression-
615 dependent forms of RNA-directed DNA methylation. *Genome Biology*, *17*, 170.
616 https://doi.org/10.1186/s13059-016-1032-y

617 **Polydore, S., & Axtell, M. J.** (2018). Analysis of RDR1/RDR2/RDR6-independent small RNAs
618 in Arabidopsis thaliana improves MIRNA annotations and reveals unexplained types of
619 short interfering RNA loci. *The Plant Journal: For Cell and Molecular Biology*, *94*(6),
620 1051–1063. https://doi.org/10.1111/tpj.13919
621 **Qi, Y., Denli, A. M., & Hannon, G. J.** (2005). Biochemical Specialization within Arabidopsis
622 RNA Silencing Pathways. *Molecular Cell*, *19*(3), 421–428.
623 https://doi.org/10.1016/j.molcel.2005.06.014
624 **Rajagopalan, R., Vaucheret, H., Trejo, J., & Bartel, D. P.** (2006). A diverse and
625 evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes & Development*,
626 *20*(24), 3407–3425. https://doi.org/10.1101/gad.1476406
627 **Ronemus, M., Vaughn, M. W., & Martienssen, R. A.** (2006). MicroRNA-Targeted and Small
628 Interfering RNA–Mediated mRNA Degradation Is Regulated by Argonaute, Dicer, and
629 RNA-Dependent RNA Polymerase in Arabidopsis. *The Plant Cell*, *18*(7), 1559–1574.
630 https://doi.org/10.1105/tpc.106.042127
631 **Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., &**
632 **Jackson, S. A.** (2014). A reference genome for common bean and genome-wide analysis
633 of dual domestications. *Nature Genetics*, *46*(7), 707–713. https://doi.org/10.1038/ng.3008
634 **Sharma, B., & Kramer, E. M.** (2014). THE MADS-BOX GENE FAMILY OF THE BASAL
635 EUDICOT AND HYBRID AQUILEGIA COERULEA "ORIGAMI"
636 (RANUNCULACEAE). *Annals of the Missouri Botanical Garden*, *99*(3), 313–322.
637 **Si-Ammour, A., Windels, D., Arn-Bouldoires, E., Kutter, C., Ailhas, J., Meins, F., &**
638 **Vazquez, F.** (2011). miR393 and Secondary siRNAs Regulate Expression of the

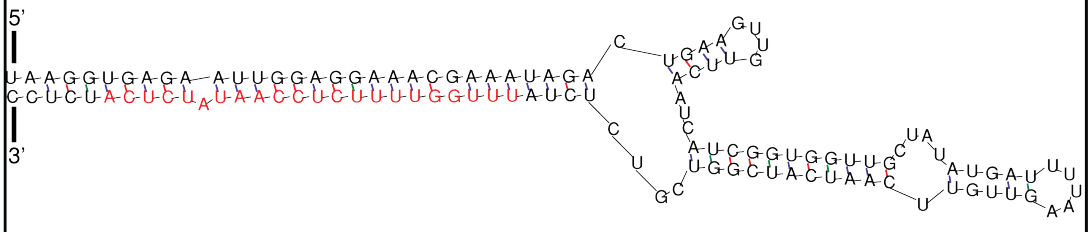
639 TIR1/AFB2 Auxin Receptor Clade and Auxin-Related Development of Arabidopsis
640 Leaves. *Plant Physiology*, *157*(2), 683–691. https://doi.org/10.1104/pp.111.180083
641 **Song, X., Li, P., Zhai, J., Zhou, M., Ma, L., Liu, B., & Cao, X.** (2008). Roles of DCL4 and
642 DCL3b in rice phased small RNA biogenesis. *The Plant Journal*, *69*(3), 462–474.
643 https://doi.org/10.1111/j.1365-313X.2011.04805.x
644 **Springer, N. M., Ying, K., Fu, Y., Ji, T., Yeh, C.-T., Jia, Y., & Schnable, P. S.** (2009). Maize
645 Inbreds Exhibit High Levels of Copy Number Variation (CNV) and Presence/Absence
646 Variation (PAV) in Genome Content. *PLOS Genetics*, *5*(11), e1000734.
647 https://doi.org/10.1371/journal.pgen.1000734
648 **Sugiyama, T., Cam, H., Verdel, A., Moazed, D., & Grewal, S. I. S.** (2005). RNA-dependent
649 RNA polymerase is an essential component of a self-enforcing loop coupling
650 heterochromatin assembly to siRNA production. *Proceedings of the National Academy of*
651 *Sciences of the United States of America*, *102*(1), 152.
652 https://doi.org/10.1073/pnas.0407641102
653 **Tamim, S., Cai, Z., Mathioni, S., Zhai, J., Teng, C., Zhang, Q., & Meyers, B. C.** (2018). Cis-
654 directed cleavage and nonstoichiometric abundances of 21-nt reproductive phasiRNAs in
655 grasses. *bioRxiv*, 243907. https://doi.org/10.1101/243907
656 **Tolia, N. H., & Joshua-Tor, L.** (2007). Slicer and the Argonautes. *Nature Chemical Biology*,
657 *3*(1), 36–43. https://doi.org/10.1038/nchembio848
658 **Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A. C., &**
659 **Crété, P.** (2004). Endogenous trans-Acting siRNAs Regulate the Accumulation of
660 Arabidopsis mRNAs. *Molecular Cell*, *16*(1), 69–79.
661 https://doi.org/10.1016/j.molcel.2004.09.028

662 **Wang, W., Haberer, G., Gundlach, H., Gläßer, C., Nussbaumer, T., Luo, M. C., & Messing,**
663 **J.** (2014). The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction
664 fast growth and aquatic lifestyle. *Nature Communications*, *5*, 3311.
665 https://doi.org/10.1038/ncomms4311
666 **Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., & Zhang, Z. (**2011). The genome of
667 the mesopolyploid crop species *Brassica rapa*. *Nature Genetics*, *43*(10), 1035–1039.
668 https://doi.org/10.1038/ng.919
669 **Williams, L., Grigg, S. P., Xie, M., Christensen, S., & Fletcher, J. C.** (2005). Regulation of
670 Arabidopsis shoot apical meristem and lateral organ formation by microRNA miR166g
671 and its AtHD-ZIP target genes. *Development*, *132*(16), 3657–3668.
672 https://doi.org/10.1242/dev.01942
673 **Willmann, M. R., Endres, M. W., Cook, R. T., & Gregory, B. D.** (2011). The Functions of
674 RNA-Dependent RNA Polymerases in Arabidopsis. *The Arabidopsis Book*, e0146.
675 https://doi.org/10.1199/tab.0146
676 **Ye, R., Chen, Z., Lian, B., Rowley, M. J., Xia, N., Chai, J., & Qi, Y.** (2016). A Dicer-
677 Independent Route for Biogenesis of siRNAs that Direct DNA Methylation in
678 Arabidopsis. *Molecular Cell*, *61*(2), 222–235.
679 https://doi.org/10.1016/j.molcel.2015.11.015
680 **Zhai, J., Bischof, S., Wang, H., Feng, S., Lee, T., Teng, C., & Jacobsen, S. E.** (2015). One
681 precursor One siRNA model for Pol IV- dependent siRNAs Biogenesis. *Cell*, *163*(2),
682 445–455. https://doi.org/10.1016/j.cell.2015.09.032
683 **Zhai, J., Jeong, D.-H., Paoli, E. D., Park, S., Rosen, B. D., Li, Y., & Meyers, B. C.** (2011).
684 MicroRNAs as master regulators of the plant NB-LRR defense gene family via the

685 production of phased, trans-acting siRNAs. *Genes & Development*, *25*(23), 2540–2553.
686 https://doi.org/10.1101/gad.177527.111
687 **Zhai, J., Zhang, H., Arikit, S., Huang, K., Nan, G.-L., Walbot, V., & Meyers, B. C.** (2015).
688 Spatiotemporally dynamic, cell-type–dependent premeiotic and meiotic phasiRNAs in
689 maize anthers. *Proceedings of the National Academy of Sciences*, *112*(10), 3146–3151.
690 https://doi.org/10.1073/pnas.1418918112
691 **Zhang, H., Xia, R., Meyers, B. C., & Walbot, V.** (2015). Evolution, functions, and mysteries of
692 plant ARGONAUTE proteins. *Current Opinion in Plant Biology*, *27*, 84–90.
693 https://doi.org/10.1016/j.pbi.2015.06.011
694 **Zheng, Y., Wang, S., & Sunkar, R.** (2014). Genome-Wide Discovery and Analysis of Phased
695 Small Interfering RNAs in Chinese Sacred Lotus. *PLoS ONE*, *9*(12).
696 https://doi.org/10.1371/journal.pone.0113790
697 **Zuker, M.** (2003). Mfold web server for nucleic acid folding and hybridization prediction.
698 *Nucleic Acids Research*, *31*(13), 3406–3415.
699
700
701
702
703
704 **Supplementary Figures**

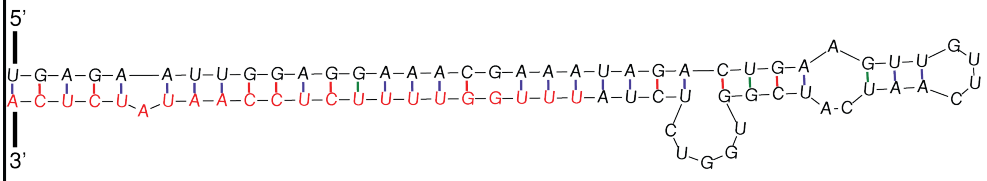
Eudicots

Citrus sinensis (Sweet Orange)



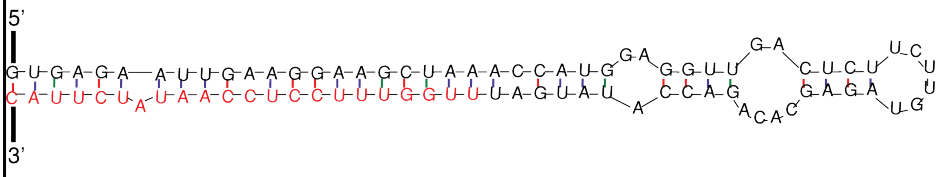
Scaffold02893:5,816-6,237

Citrus clementina (Clementine)



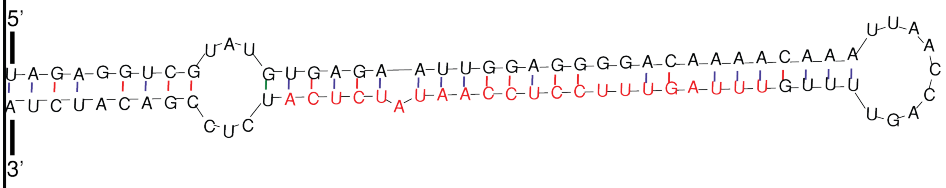
Scaffold_7:11,677,716-11,678,137

Populus trichocarpa (Black Cottonwood)



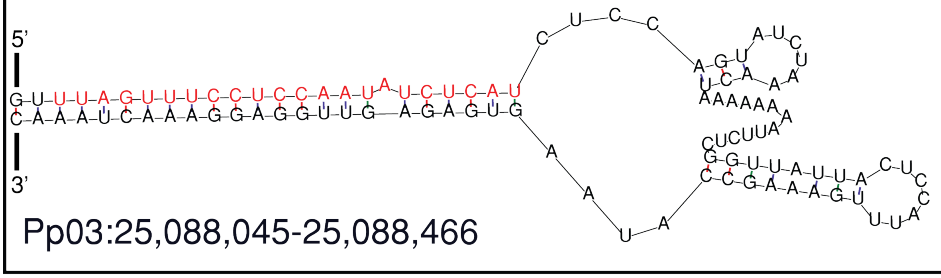
Chr18:16,602,794-16,603,213

Fragaria vesca (Stawberry)



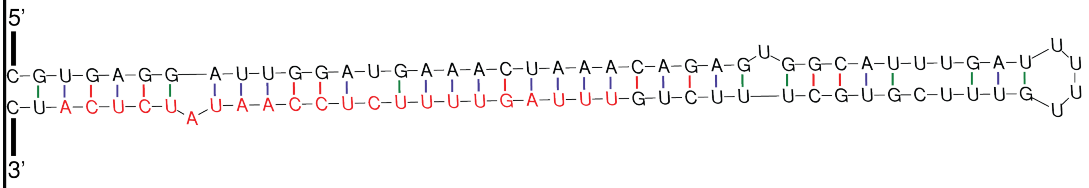
LG6:33,999,807-34,000,228

Prunus persica (Peach)



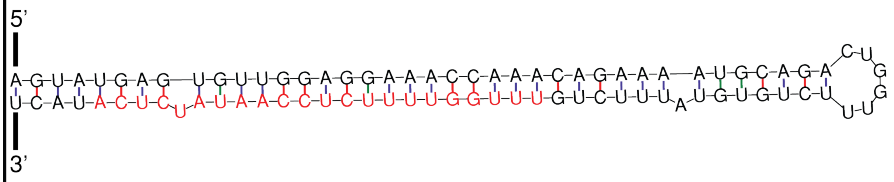
Pp03:25,088,045-25,088,466

Daucus carota (Carrot)



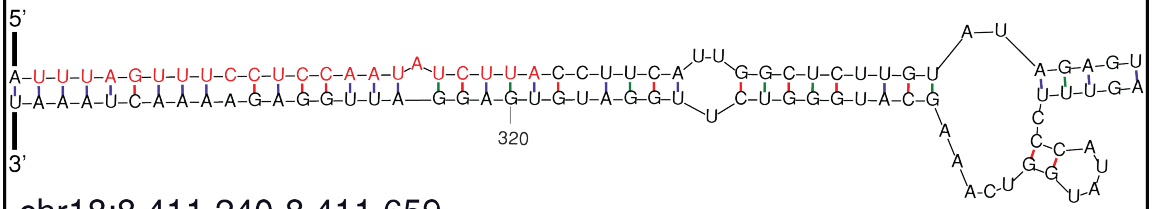
DCARv2_Chr3:50,126,000-50,126,421

Linum usitatissimum (Flax)



Scaffold917:325,461-325,882

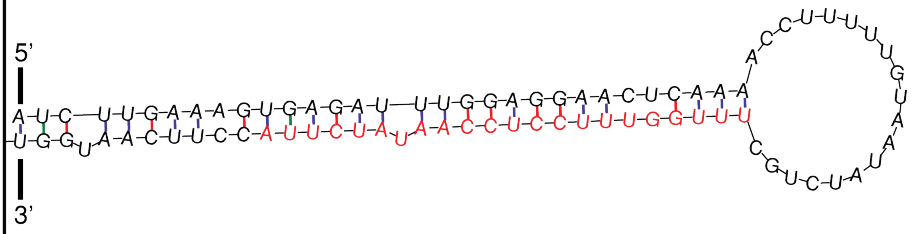
Salix purpurea



chr18:8,411,240-8,411,659

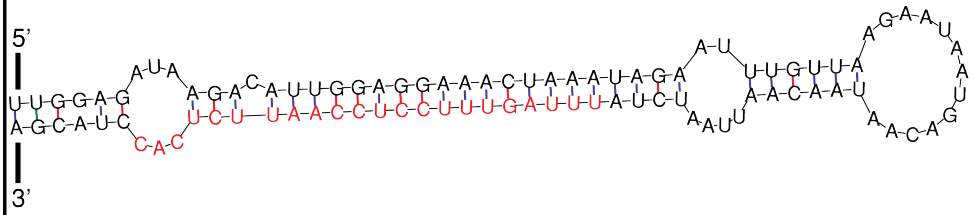
706

Manihot esculenta (Cassava)



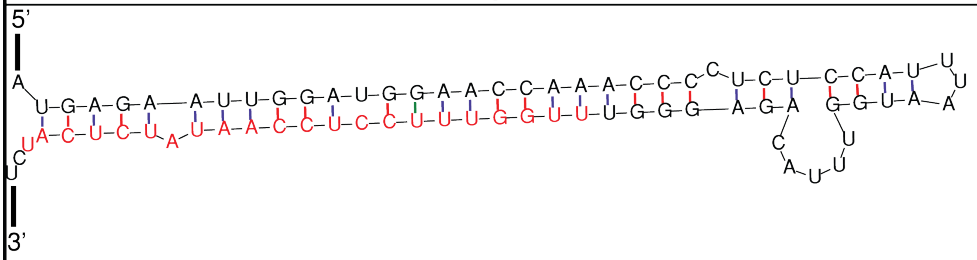
Chromosome09:25,243,068-25,243,487

Vitis vinifera (Grapevine)



chr18_random:2,094,218-2,094,599

Aquilegia coerulea

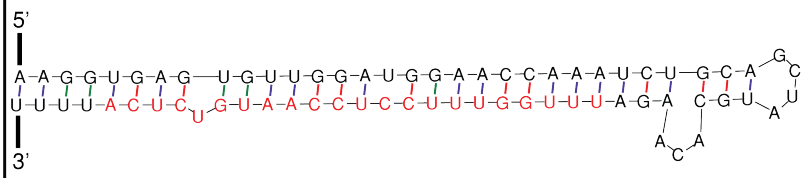


Chr5:8,028,124-8,028,545

707

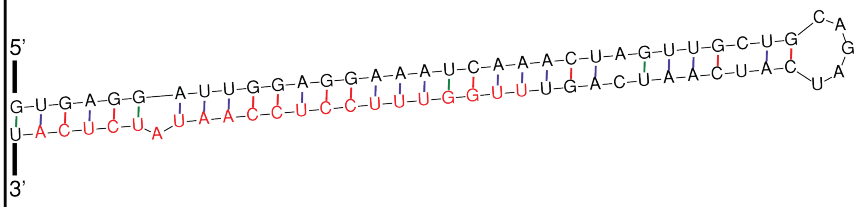
Monocots

Panicum hallii



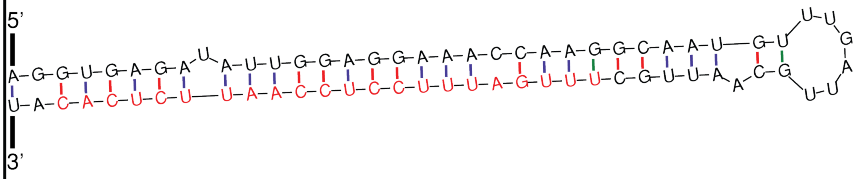
chr18_random:2,094,218-2,094,599

Ananas comosus (Pineapple)



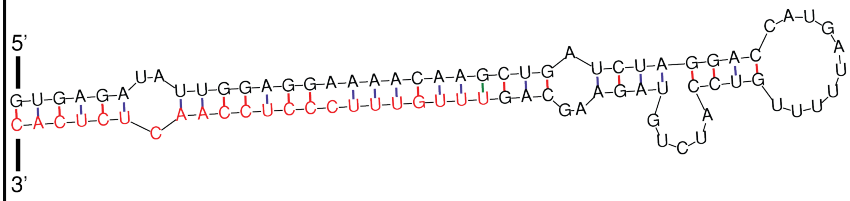
LG24:194,094-194,515

Musa acuminata (Banana)



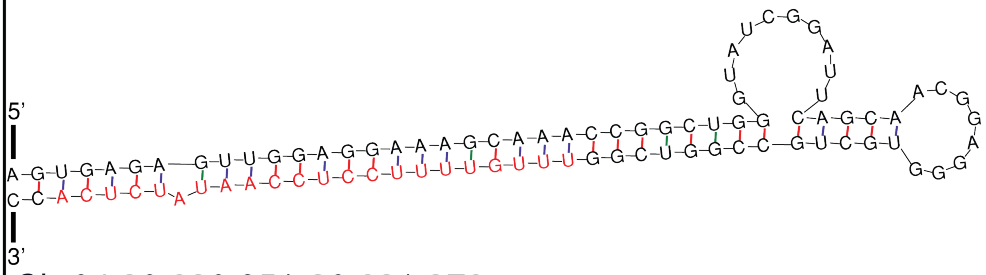
chr3:26,582,014-26,582,394

Brachypodium stacei



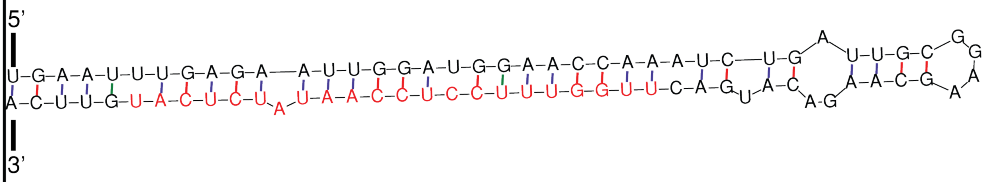
Chr03:1,620,495-1,620,875

Zea mays (Maize)



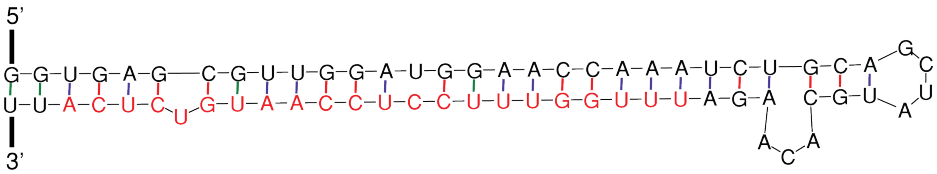
Chr04:30,230,951-30,231,372

Sorghum bicolor



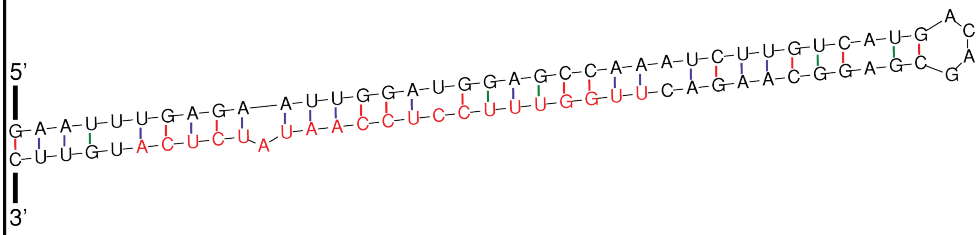
Chr07:2,205,924-2,206,344

Panicum virgatum (Switchgrass)



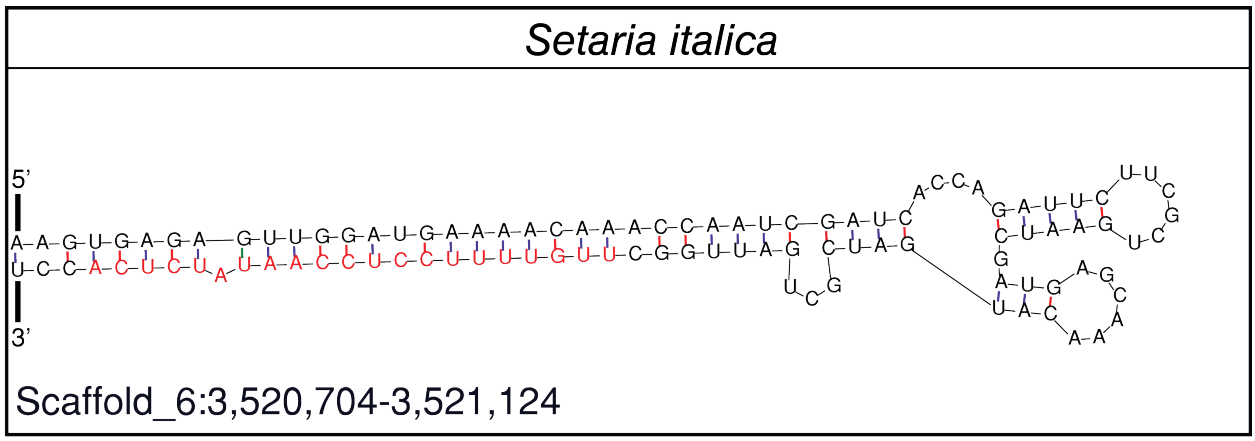
Chr06N:2,434,933-2,435,354

Oropetium thomaeum



Oropetium_genomic_20141112_019:926,647-927,067

709



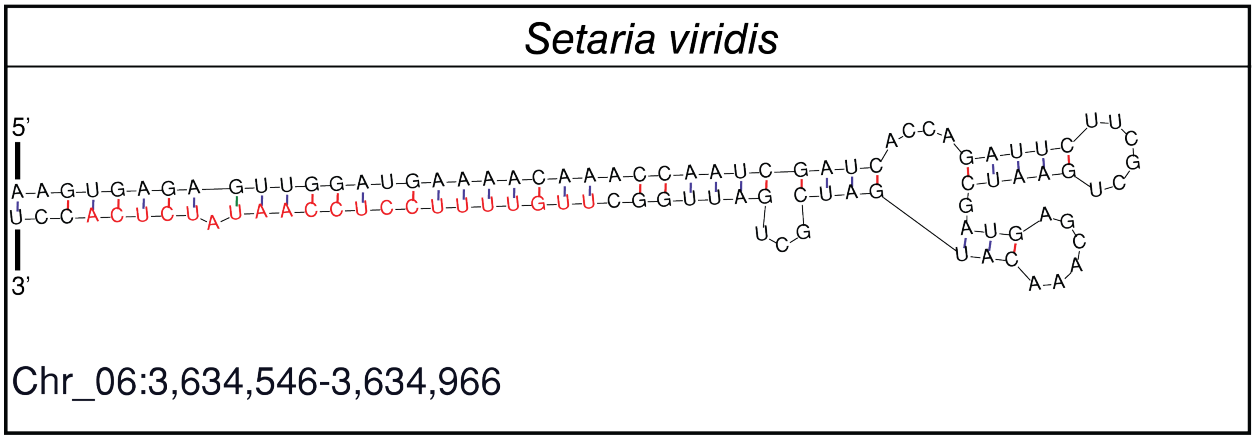
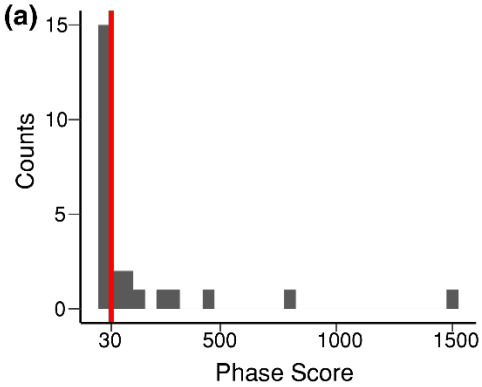


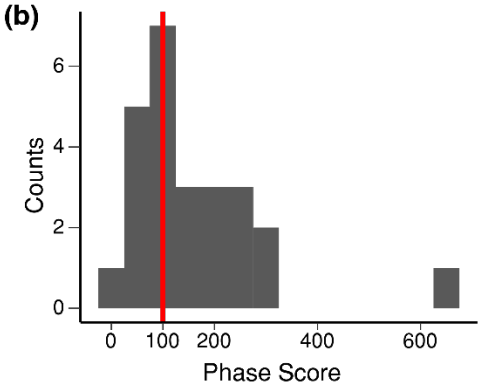
Figure S1. The RNA secondary structure of BLAST-validated *MIR2275* homologs.

Secondary structures of loci were predicted via mFOLD. If the potential *MIR2275* homolog formed a stem-and-loop structure, it is shown above. The red nucleotides denote the mature *MIR2275* homolog in the secondary structure.



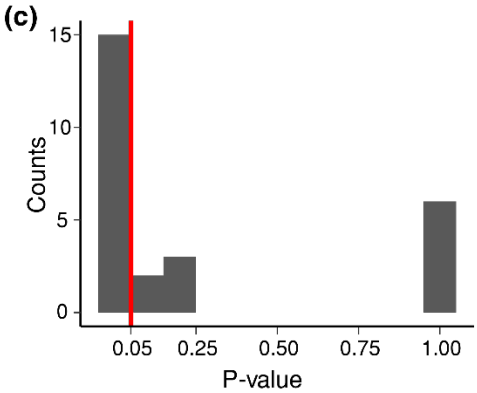
$$S = PR * PN * \ln(1+PArpm)$$

S is the Phase Score.
PR is the Phase Ratio.
PN is the number of distinct, "in-phase" reads.
PArpm is the accumulation of "in-phase" reads (in Reads Per Million).



$$S = \ln\left[\left(1 + C * \frac{\sum_{i=1}^{C} P_i}{\sum U}\right)^{n-2}\right], n>4$$

S is the Phase Score.
C is the number of 21 nt or 24 nt Phase Cycles that fits into the *PHAS* locus.
P is the number of "in-phase" reads.
U is the number of "out-of-phase" reads.
n is the number of phase cycles with at least one read.



$$\text{P-val}\ (X=k) = \sum_{X=k}^{m} \frac{\binom{(H-1)m}{n-k}\binom{m}{k}}{\binom{Hm}{n}}$$

P-val= p-value with null hypothesis that a locus is a true *PHAS*.
k is the number of unique 21 nt or 24 nt "in phase" reads.
m is the number of phase cycles.
H is 21 or 24.
n is the number of unique 21 nt or 24 nt reads in a particular locus.

Figure S2. Determination of phase score cutoffs using three different algorithms

(a) Small RNA accumulation at 25 known 21 nt *PHAS* loci (Table S1) was analyzed using the *PHAS*-Test algorithm from Guo et al. (2015), which is used by ShortStack. Variables and formula are shown. Red line represents the cut-off used in this study to determine if a locus was phased or not; scores above the red line were considered 'phased'.

(b) Same as in panel a except for the *PHAS*-Test algorithm from Zheng et al. (2014).

(c) Same as in panel a except for the *PHAS*-Test algorithm from Dotto et al. (2014); scores below the red-line were considered 'phased'.

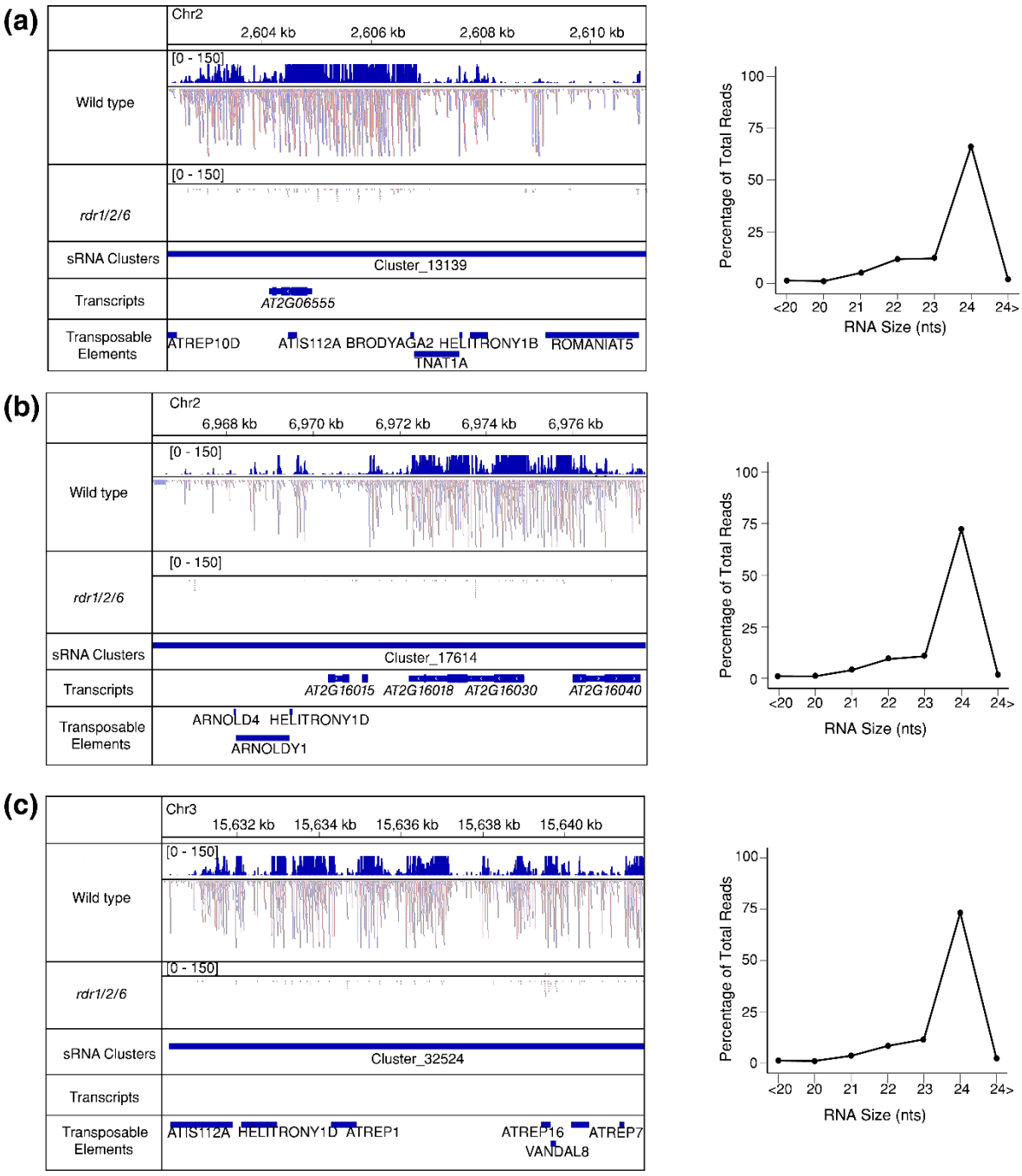
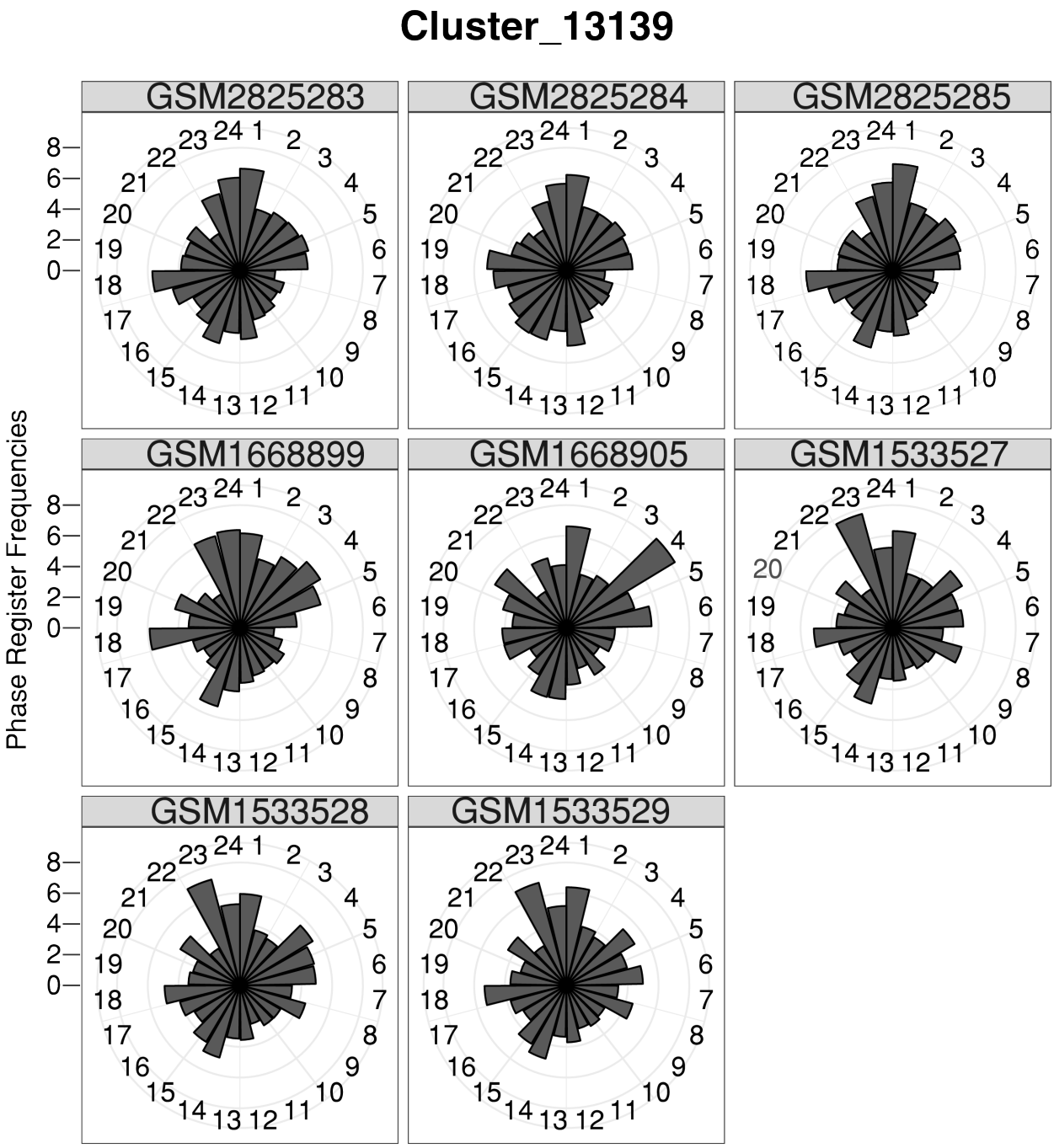


Figure S3. Genome browser snapshots of the three *Arabidopsis thaliana PHAS*-Test passing clusters.

(a) Alignments and read size distribution of Cluster_13139. Red reads represent positive-strand mapped genes, while blue reads are those that mapped to the negative strand. Numbers in the brackets are the range of coverage shown in Reads per Million.

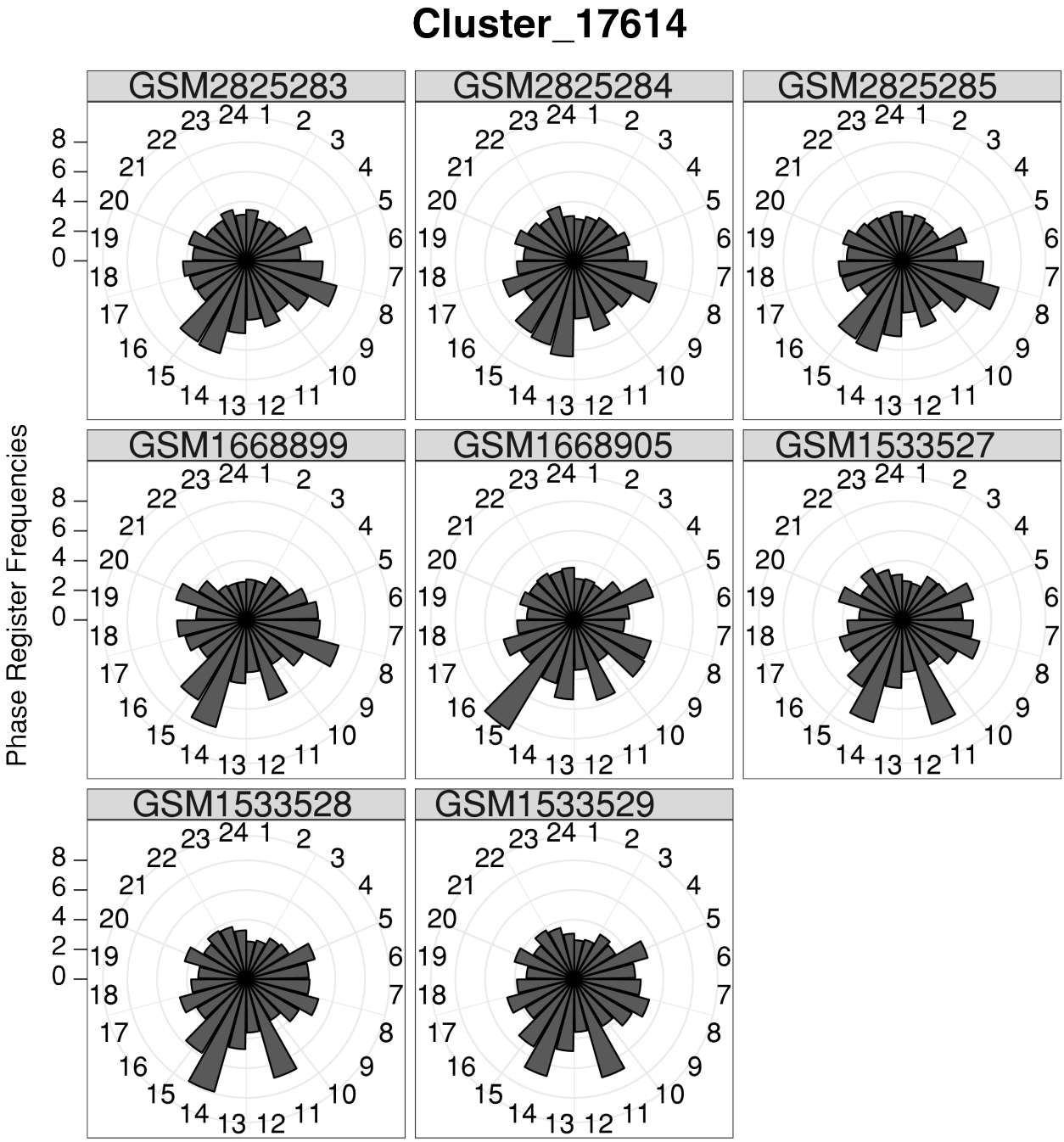
(b) Same as a, except for Cluster_17614.

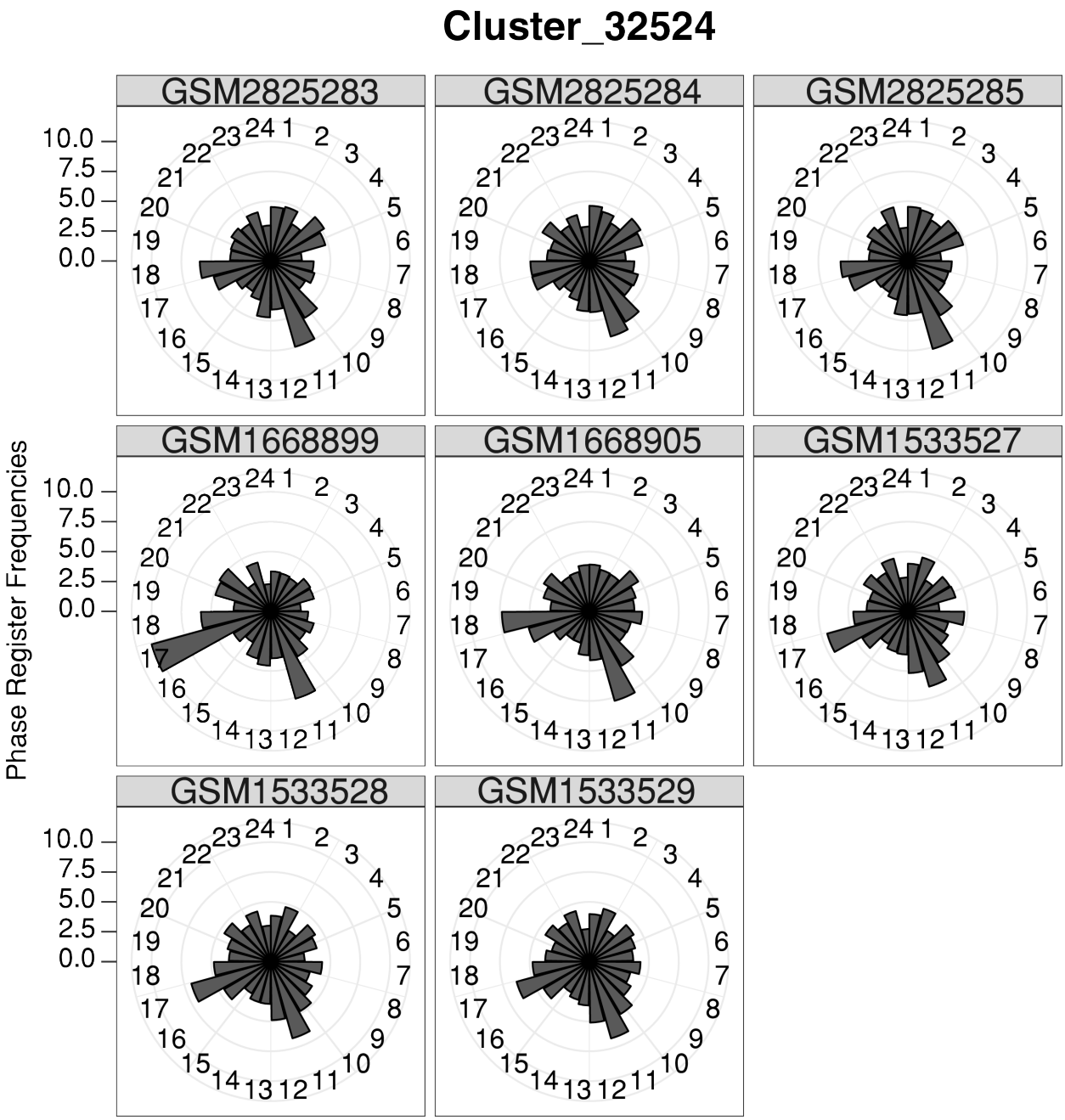
(c) Same as a, except for Cluster_32524.


Cluster_13139
GSM2825283
GSM2825284
GSM2825285
GSM1668899
GSM1668905
GSM1533527
GSM1533528
GSM1533529
Phase Register Frequencies
0
2
4
6
8
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

732

733


Cluster_17614
GSM2825283
GSM2825284
GSM2825285
GSM1668899
GSM1668905
GSM1533527
GSM1533528
GSM1533529
Phase Register Frequencies


Cluster_32524
GSM2825283
GSM2825284
GSM2825285
GSM1668899
GSM1668905
GSM1533527
GSM1533528
GSM1533529
Phase Register Frequencies
10.0
7.5
5.0
2.5
0.0

736

737

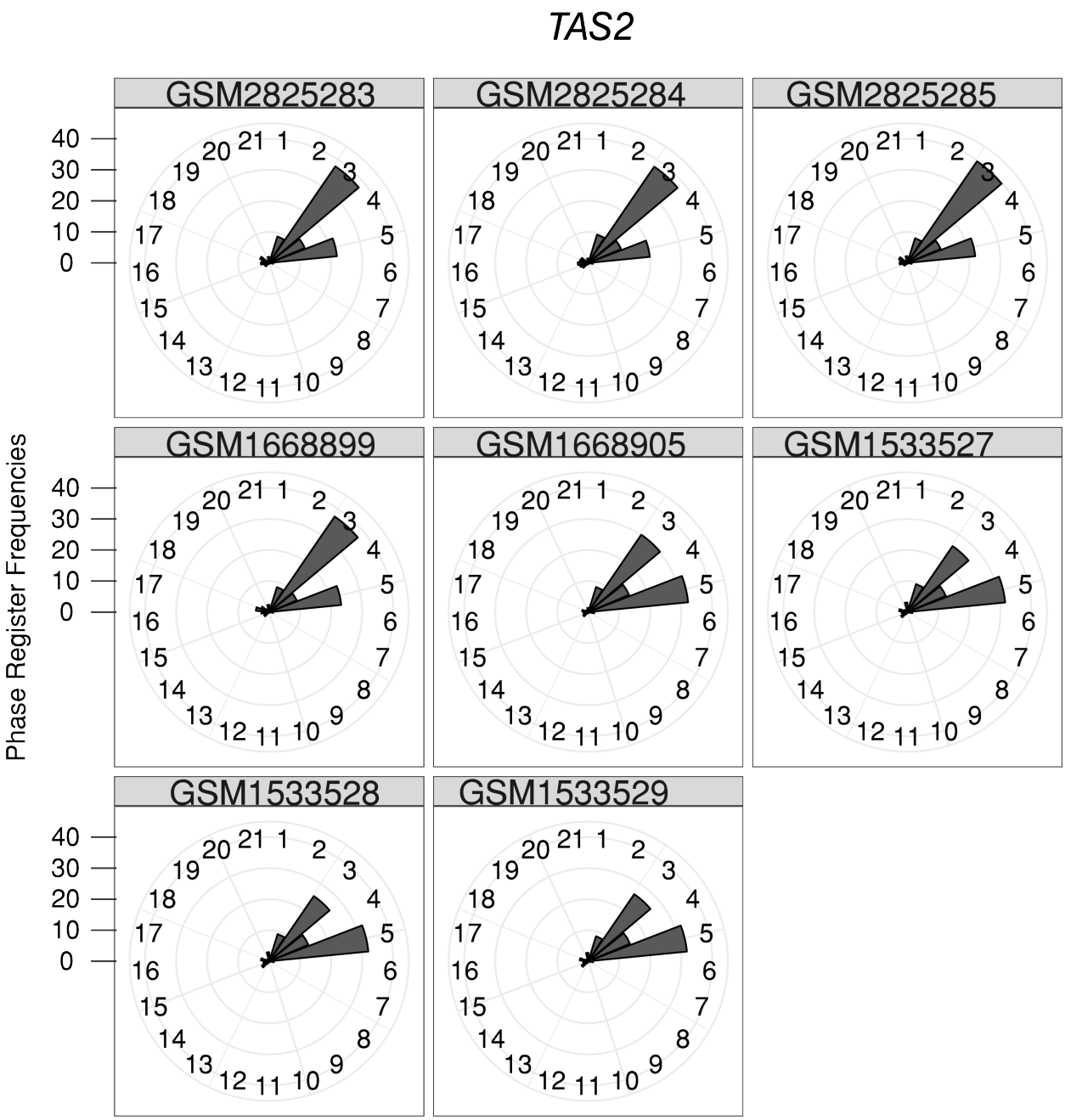


Figure S4. The three *PHAS*-Test passing loci have inconsistent phase register frequencies.

Frequencies of phase registers where calculated in eight publicly available wild-type inflorescence libraries from *A. thaliana* (the accession numbers are listed in the grey boxes). *TAS2* is shown for comparison. Phase register frequencies are calculated via the following formula: (number of reads "in phase"/ total number of reads at locus)*100.

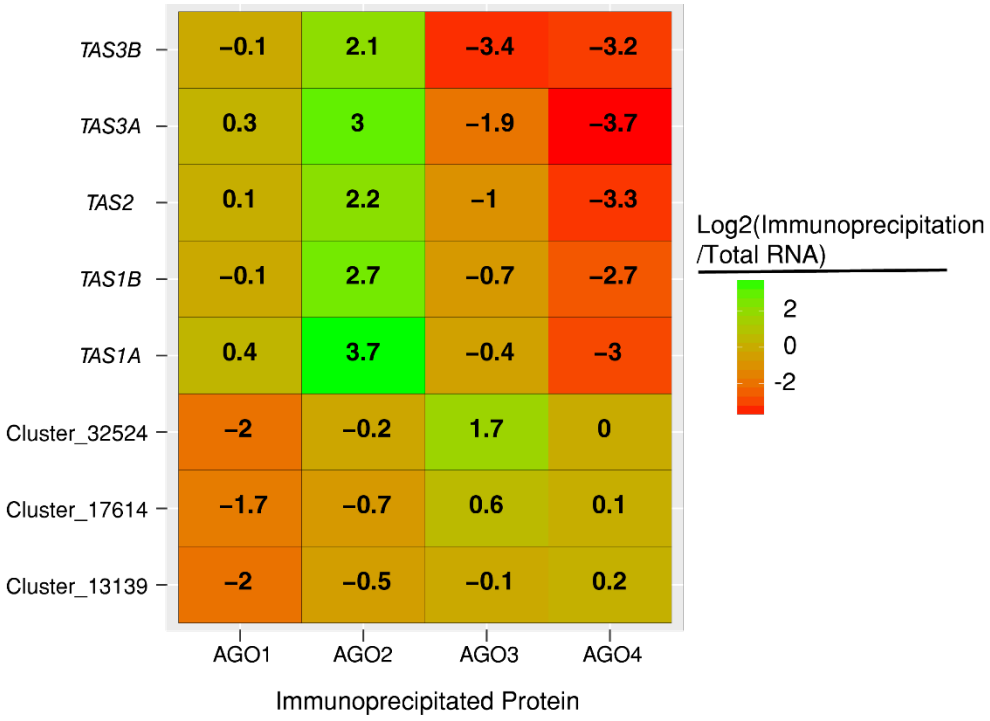
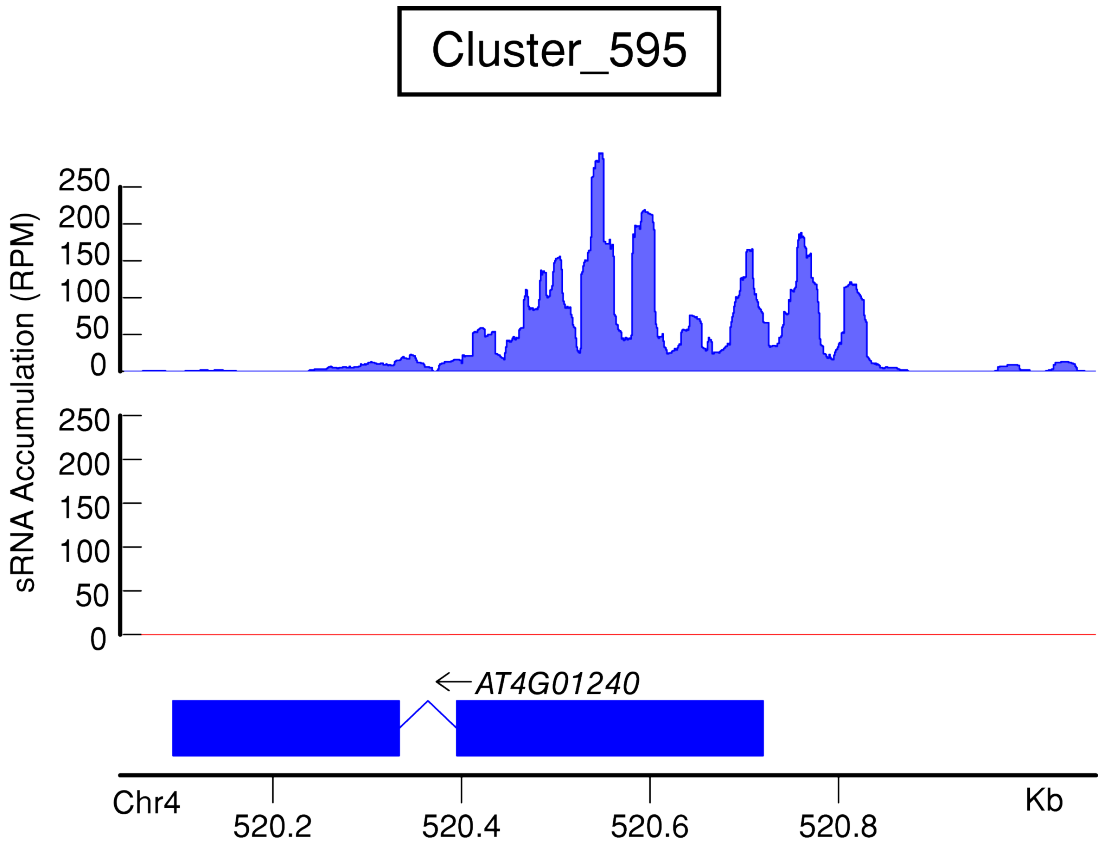
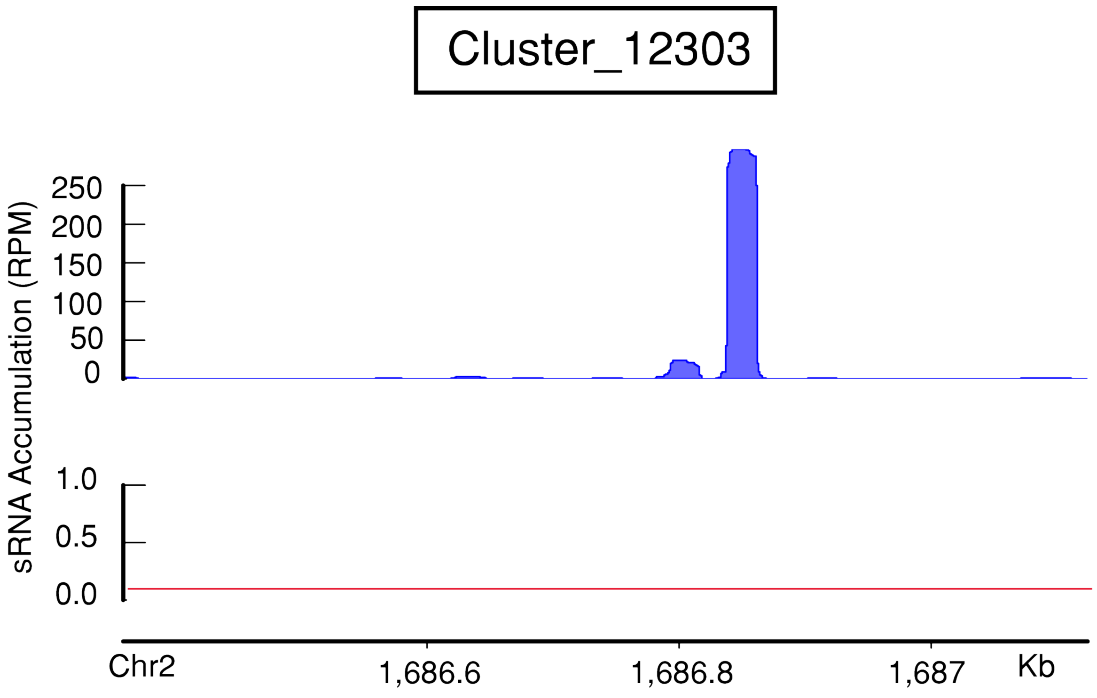


Figure S5. The three putative 24 nt *PHAS* loci show AGO-loading profiles that are distinct from *TAS* loci.

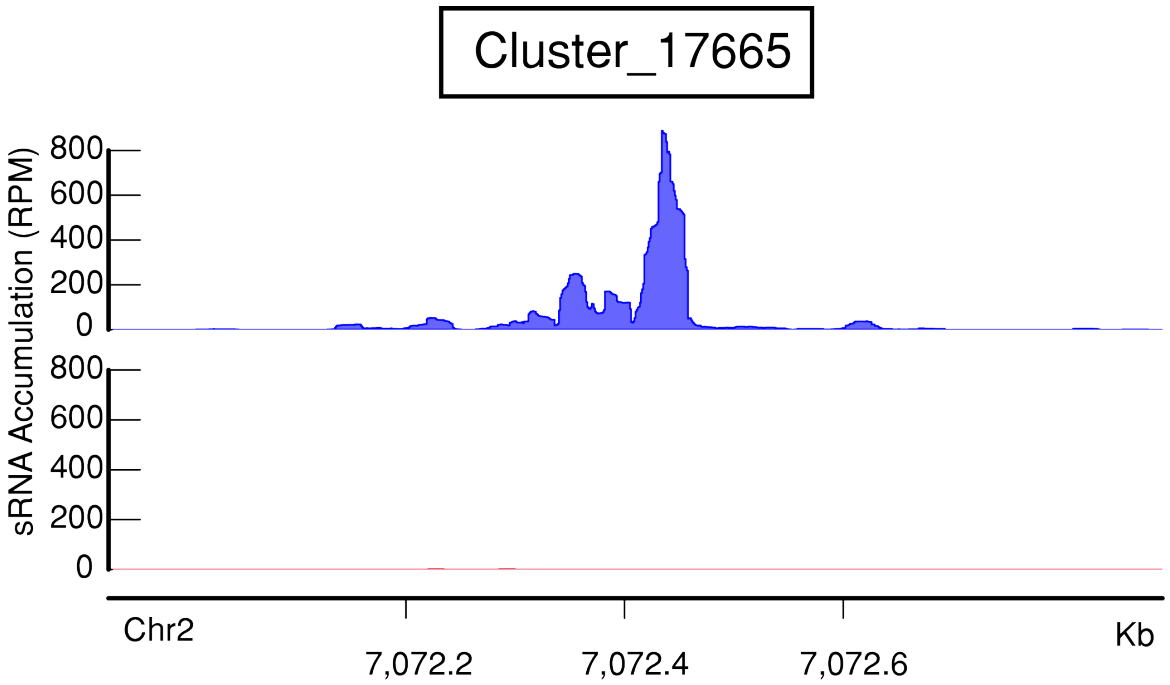
Small RNAs from immunoprecipitation protein were aligned to the *A. thaliana* (TAIR10) genome. Numbers indicate the ratio of sRNA accumulation between immunoprecipitated and total libraries (in RPMs).


Cluster_595
sRNA Accumulation (RPM)
250
200
150
100
50
0
250
200
150
100
50
0
AT4G01240
Chr4
520.2
520.4
520.6
520.8
Kb

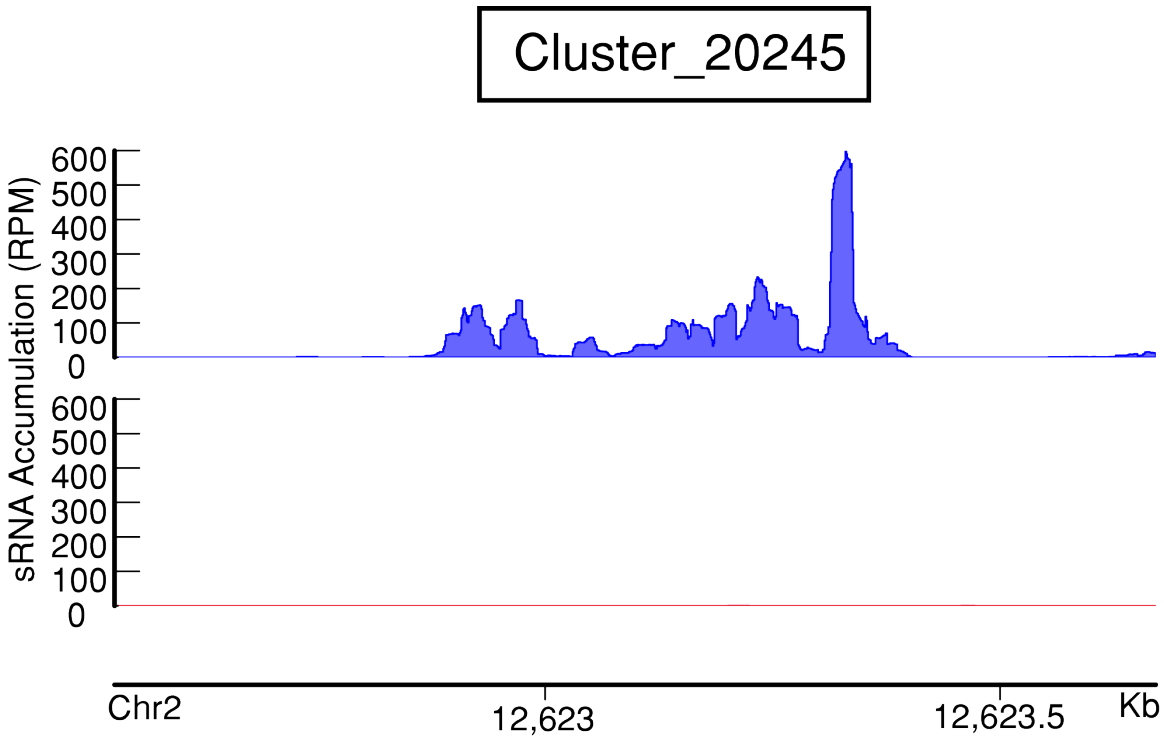
750


Cluster_12303
sRNA Accumulation (RPM)
250
200
150
100
50
0
1.0
0.5
0.0
Chr2
1,686.6
1,686.8
1,687
Kb

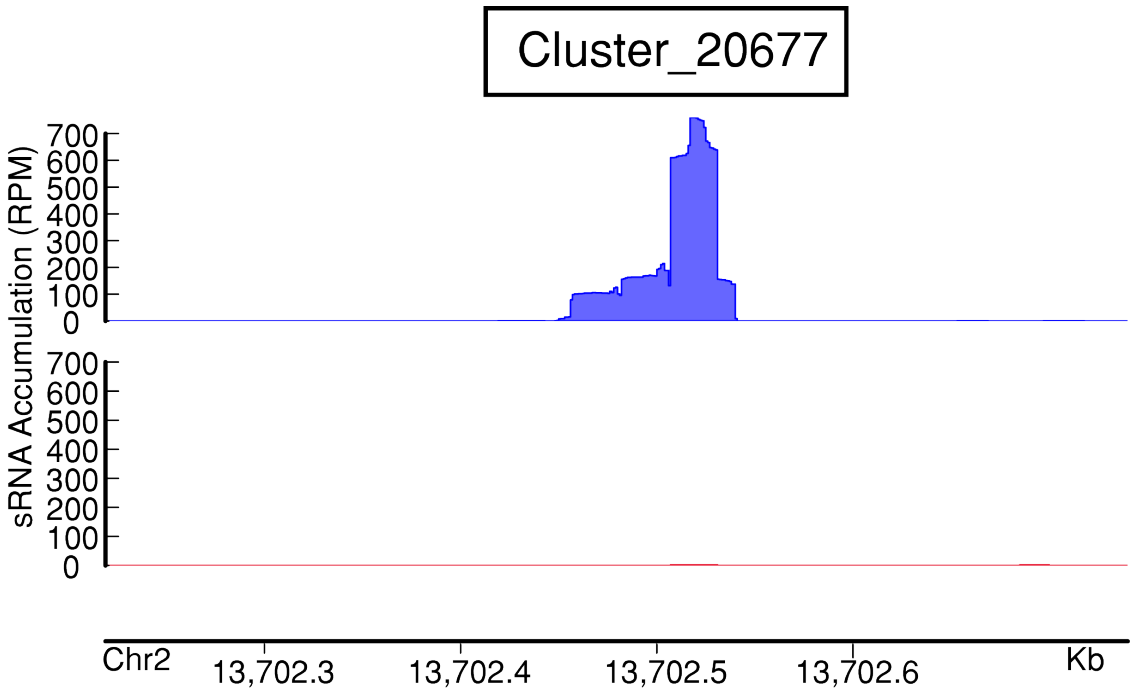
751


Cluster_17665
sRNA Accumulation (RPM)
800
600
400
200
0
800
600
400
200
0
Chr2
7,072.2
7,072.4
7,072.6
Kb

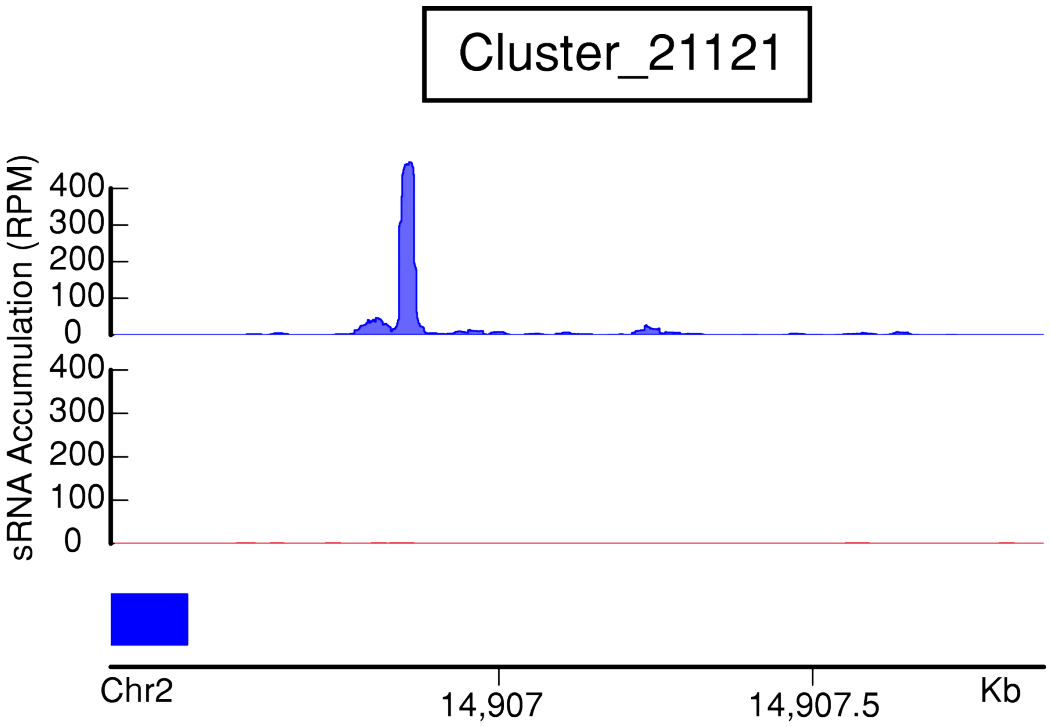
752


Cluster_20245
sRNA Accumulation (RPM)
600
500
400
300
200
100
0
600
500
400
300
200
100
0
Chr2
12,623
12,623.5
Kb

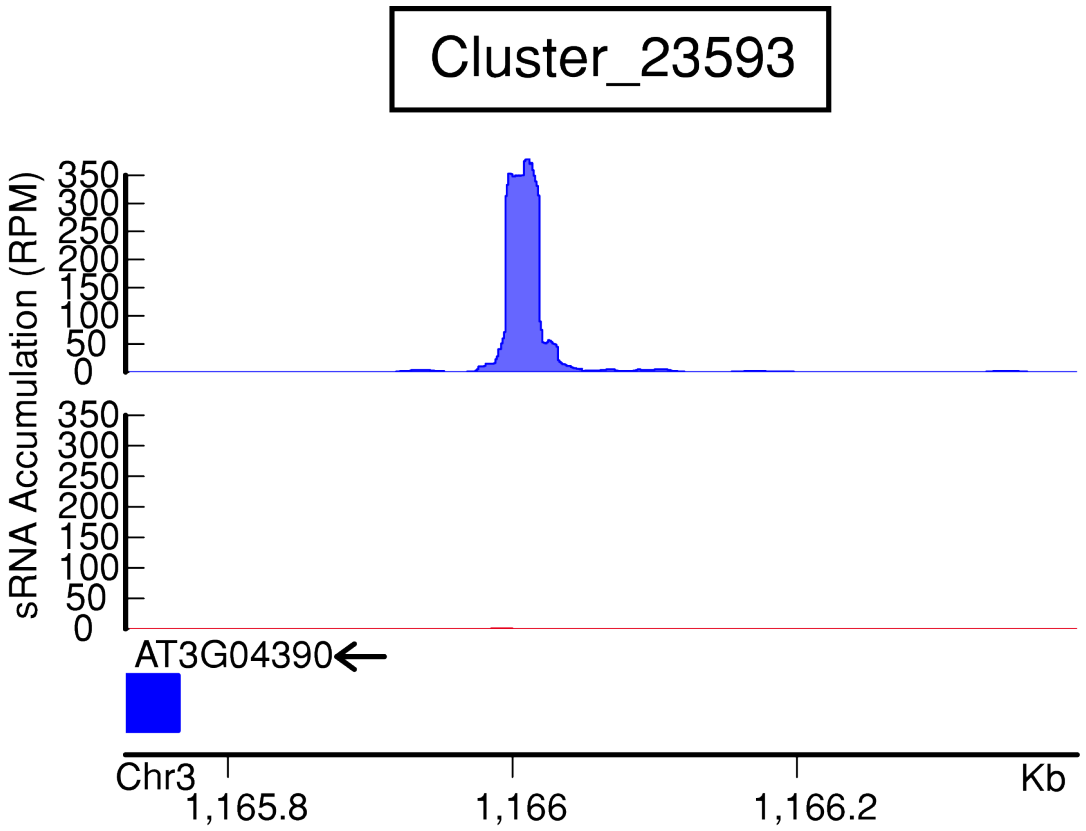
753


Cluster_20677
sRNA Accumulation (RPM)
700
600
500
400
300
200
100
0
Chr2
13,702.3
13,702.4
13,702.5
13,702.6
Kb

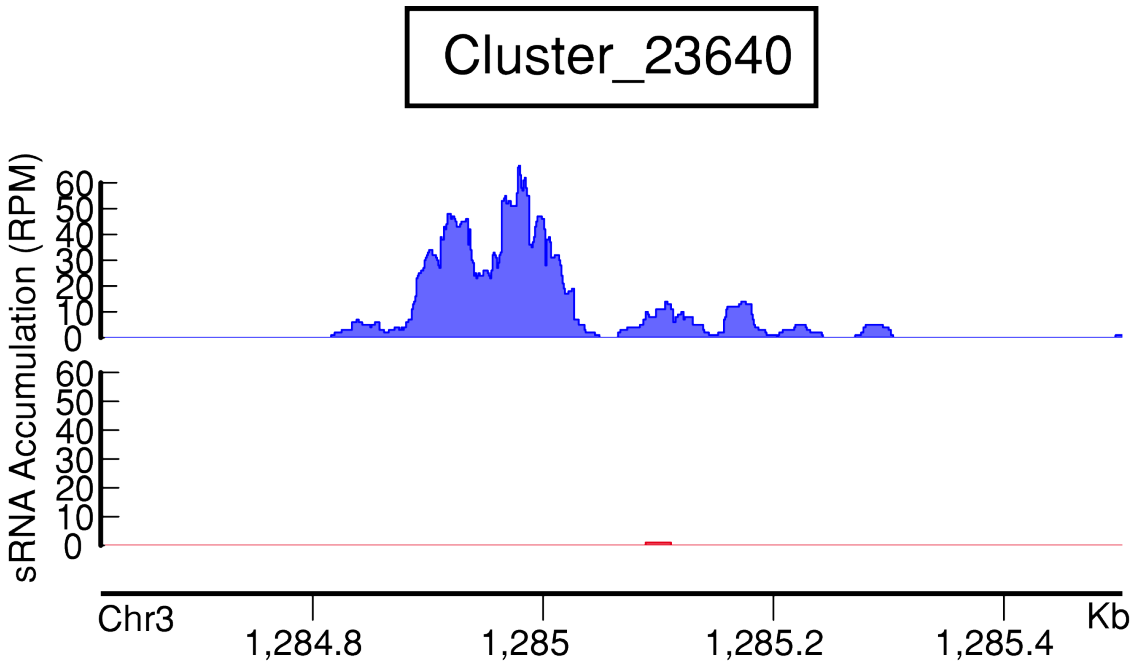
754


Cluster_21121
sRNA Accumulation (RPM)
400
300
200
100
0
400
300
200
100
0
Chr2
14,907
14,907.5
Kb

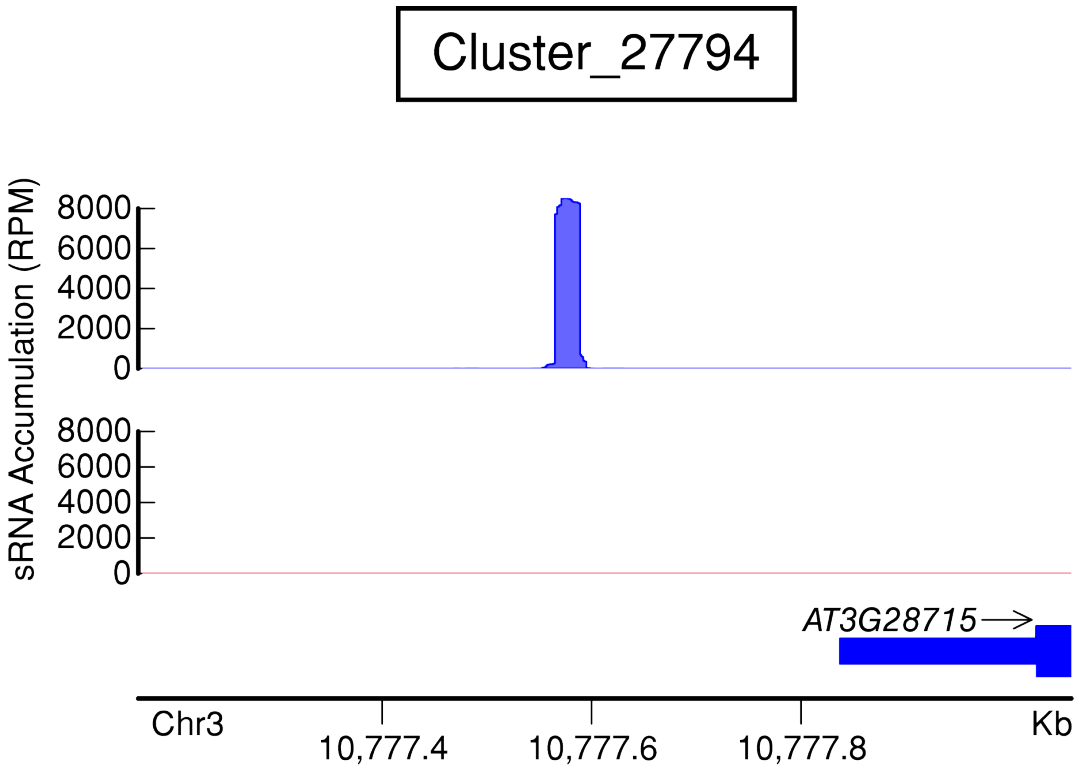
755


Cluster_23593
sRNA Accumulation (RPM)
350
300
250
200
150
100
50
0
AT3G04390
Chr3
1,165.8
1,166
1,166.2
Kb

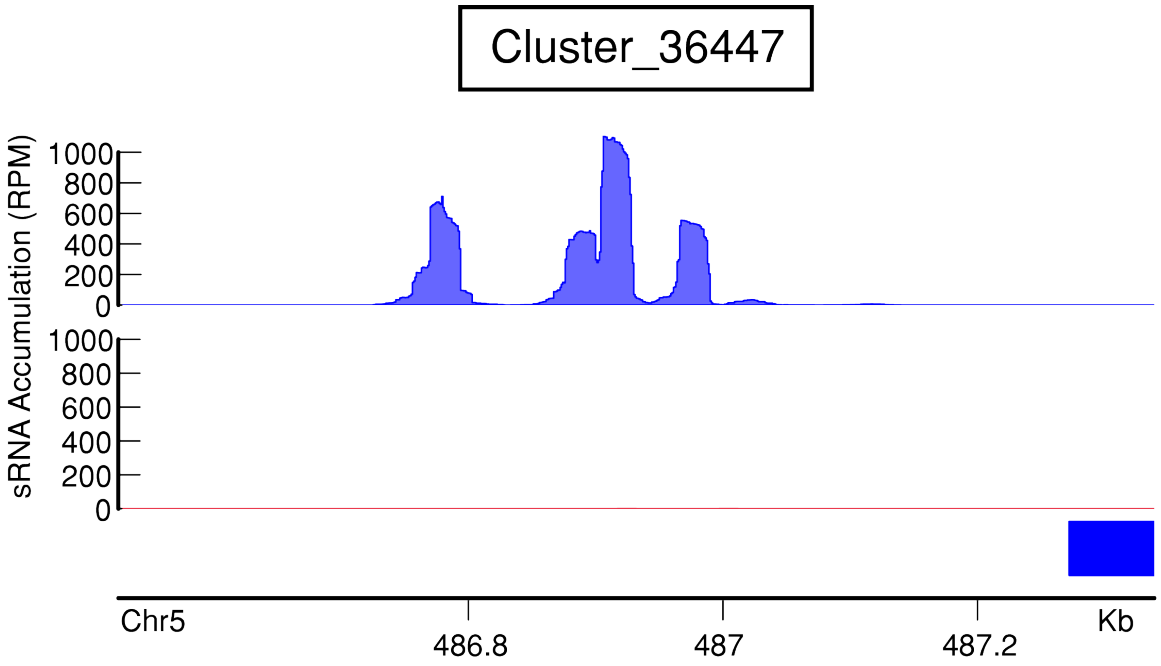
756


Cluster_23640
sRNA Accumulation (RPM)
60
50
40
30
20
10
0
60
50
40
30
20
10
0
Chr3
1,284.8
1,285
1,285.2
1,285.4
Kb

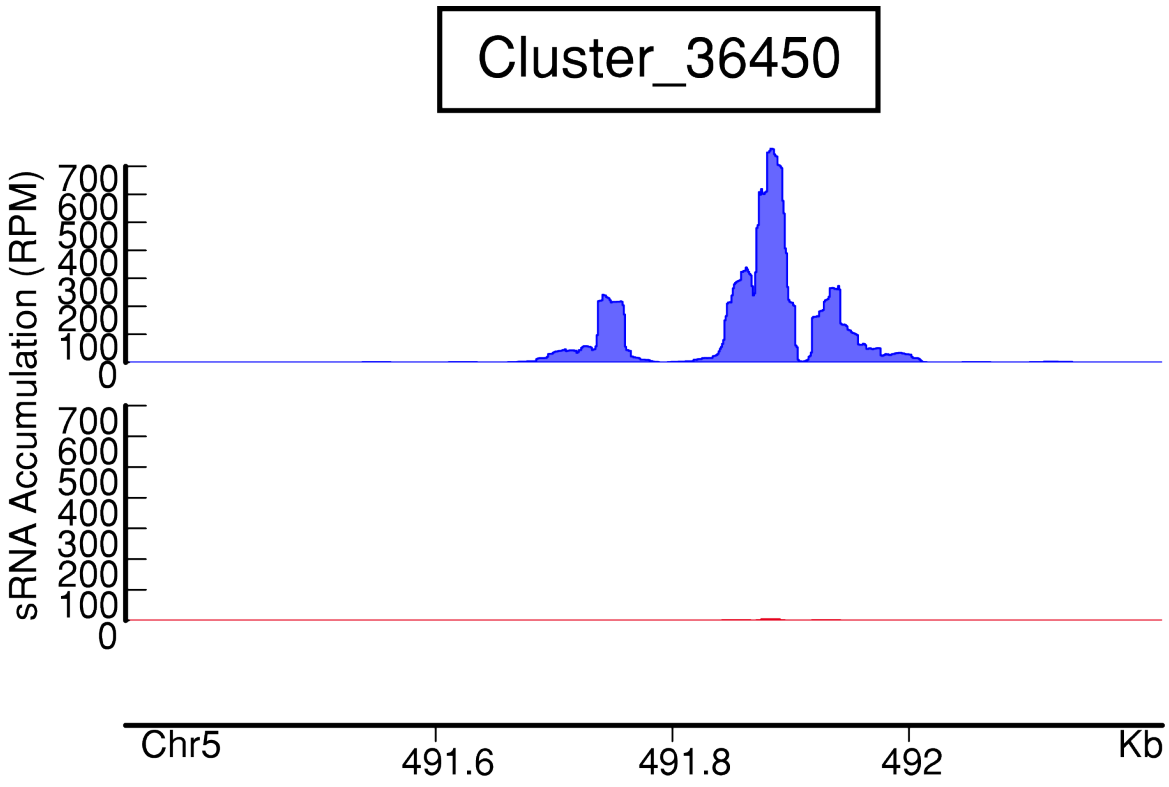
757


Cluster_27794
sRNA Accumulation (RPM)
8000
6000
4000
2000
0
8000
6000
4000
2000
0
AT3G28715
Chr3
10,777.4
10,777.6
10,777.8
Kb

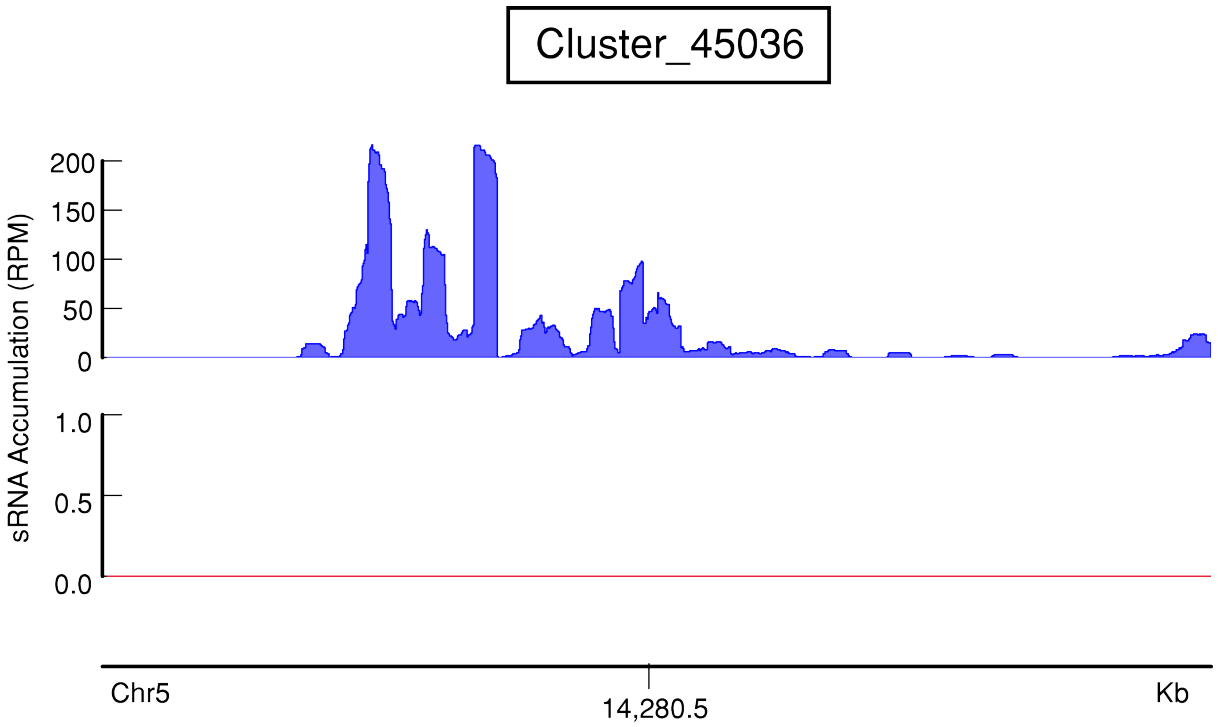
758


Cluster_36447
sRNA Accumulation (RPM)
1000
800
600
400
200
0
1000
800
600
400
200
0
Chr5
486.8
487
487.2
Kb

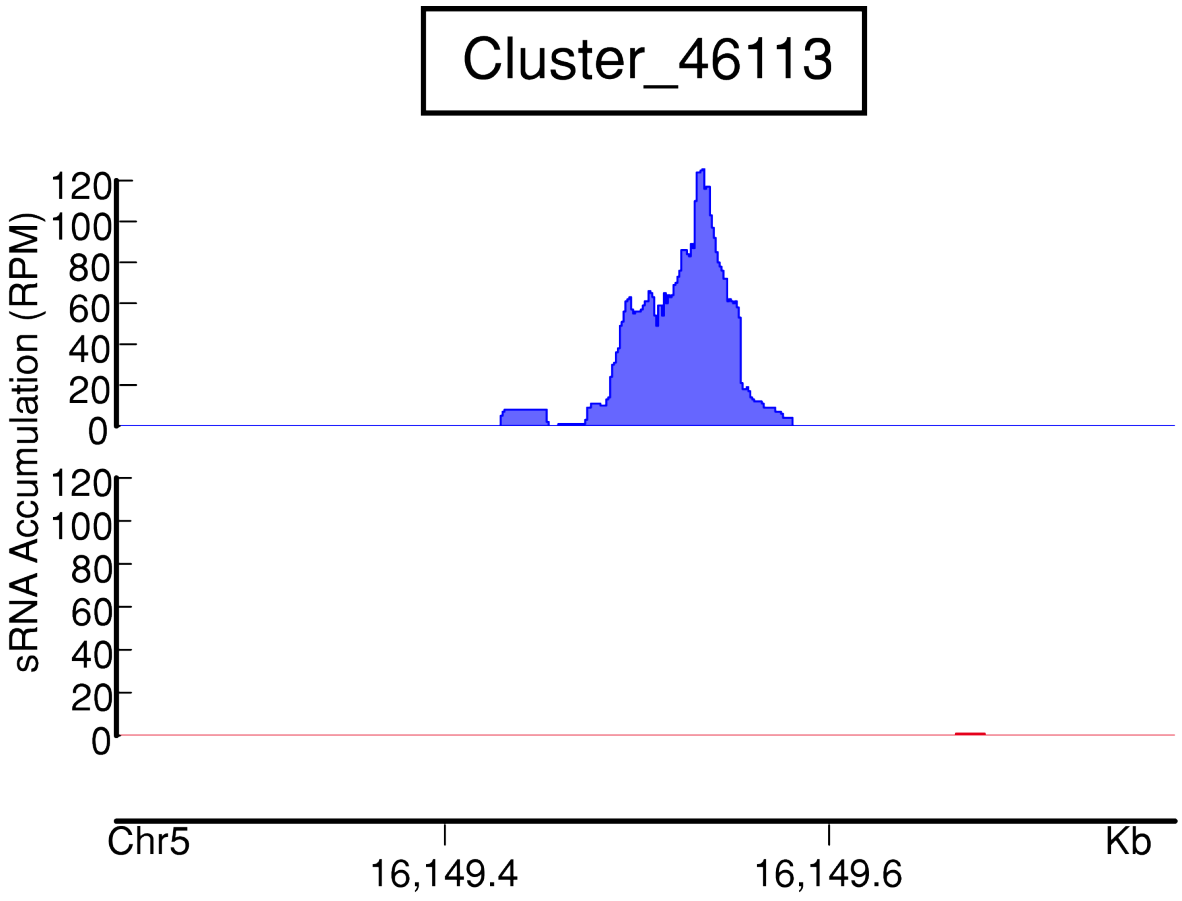
759


Cluster_36450
sRNA Accumulation (RPM)
700
600
500
400
300
200
100
0
700
600
500
400
300
200
100
0
Chr5
491.6
491.8
492
Kb

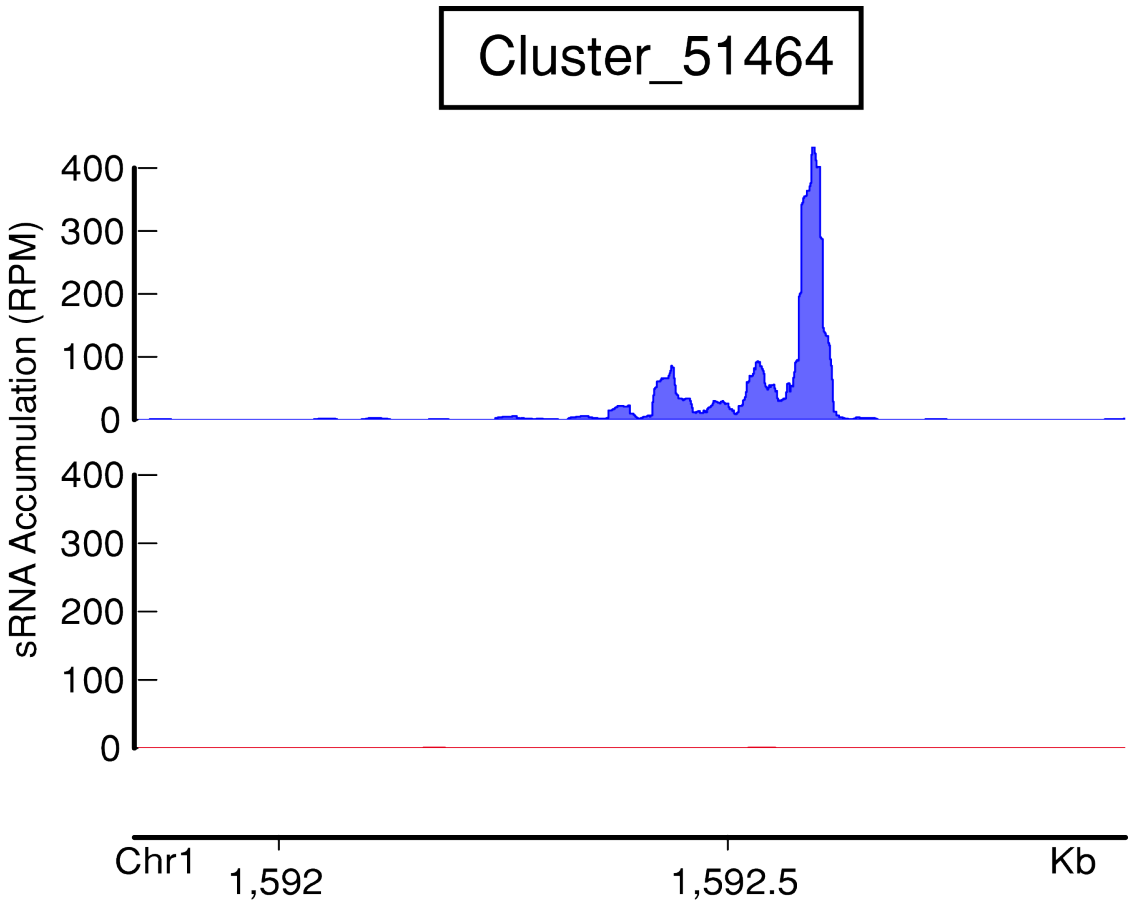
760


Cluster_45036
sRNA Accumulation (RPM)
200
150
100
50
0
1.0
0.5
0.0
Chr5
14,280.5
Kb

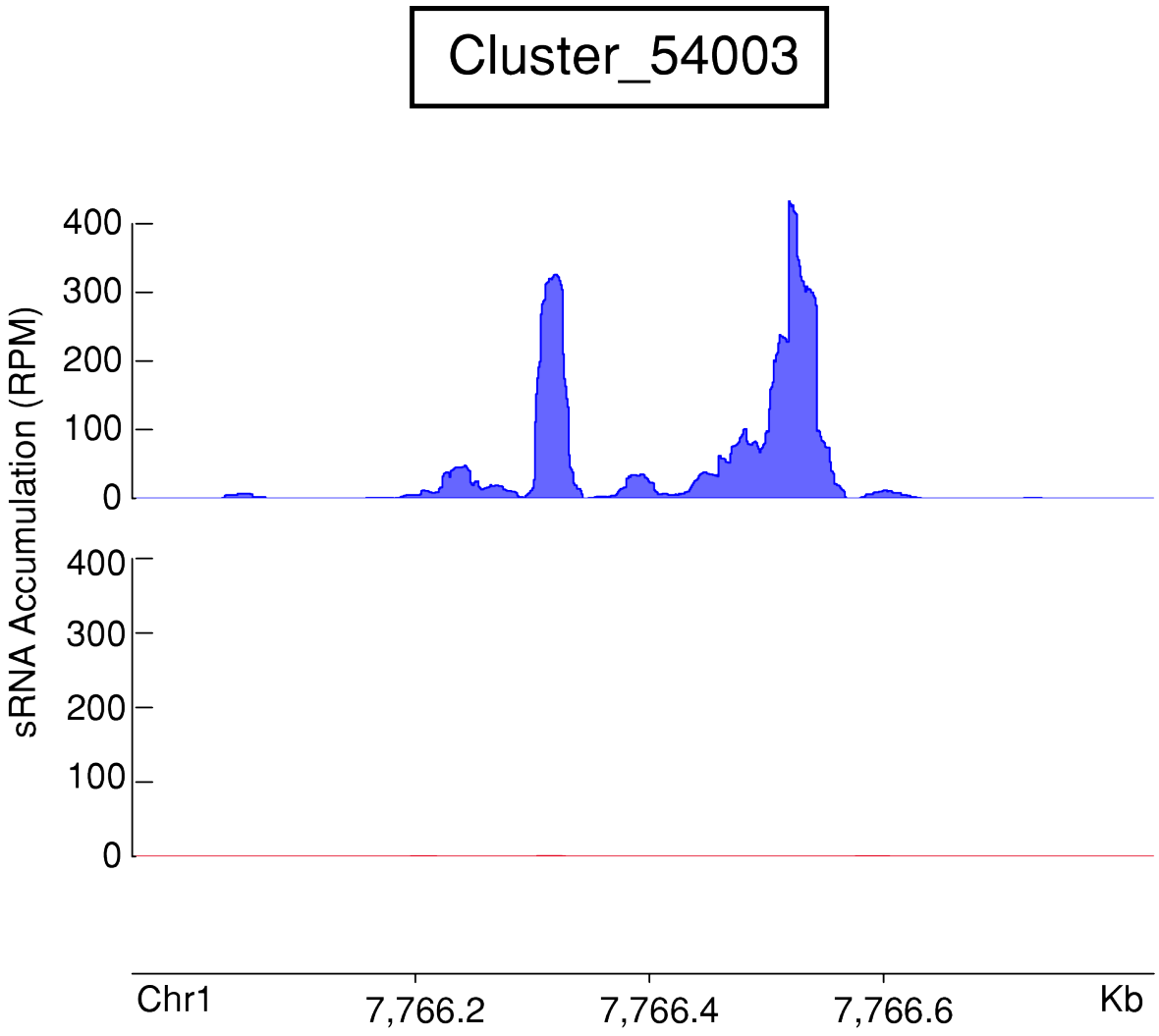
761


Cluster_46113
sRNA Accumulation (RPM)
120
100
80
60
40
20
0
120
100
80
60
40
20
0
Chr5
16,149.4
16,149.6
Kb

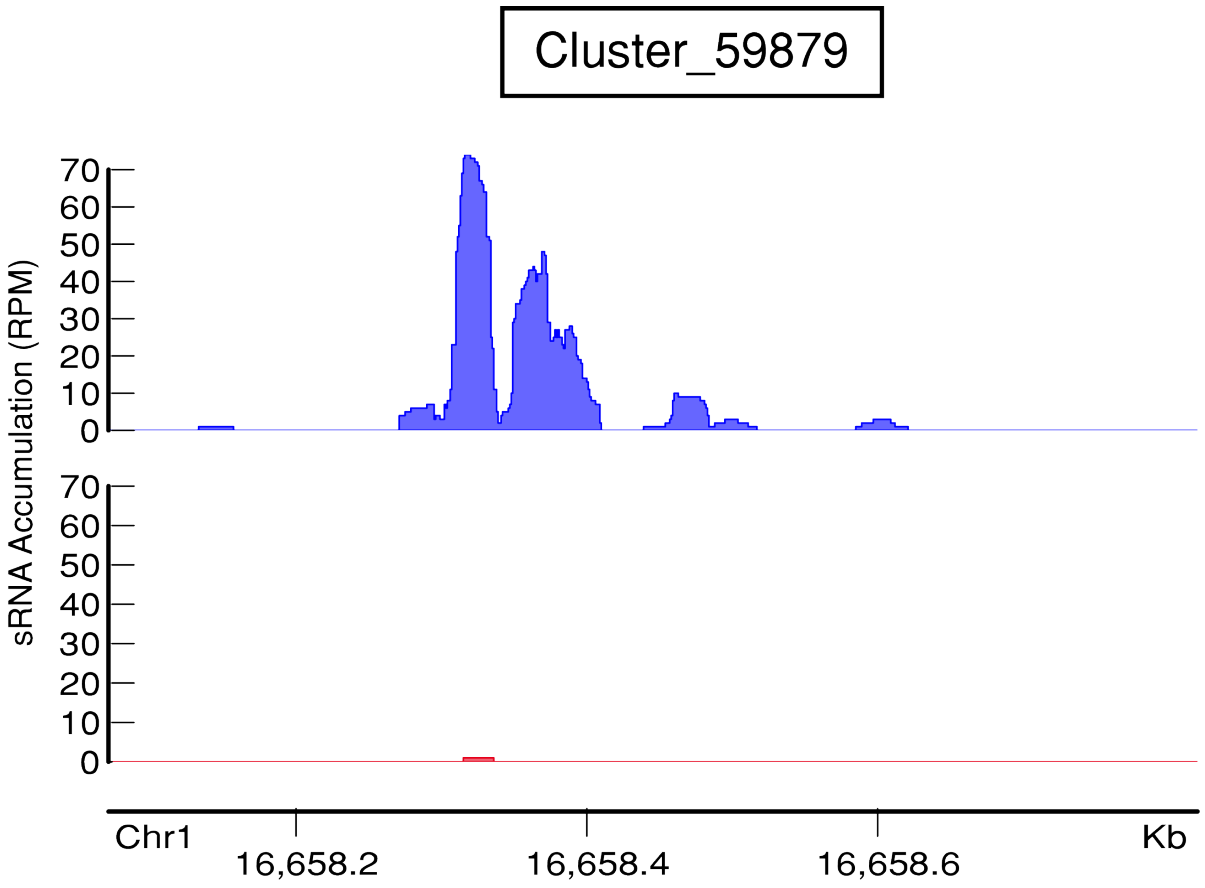
762


Cluster_51464
sRNA Accumulation (RPM)
400
300
200
100
0
400
300
200
100
0
Chr1
1,592
1,592.5
Kb

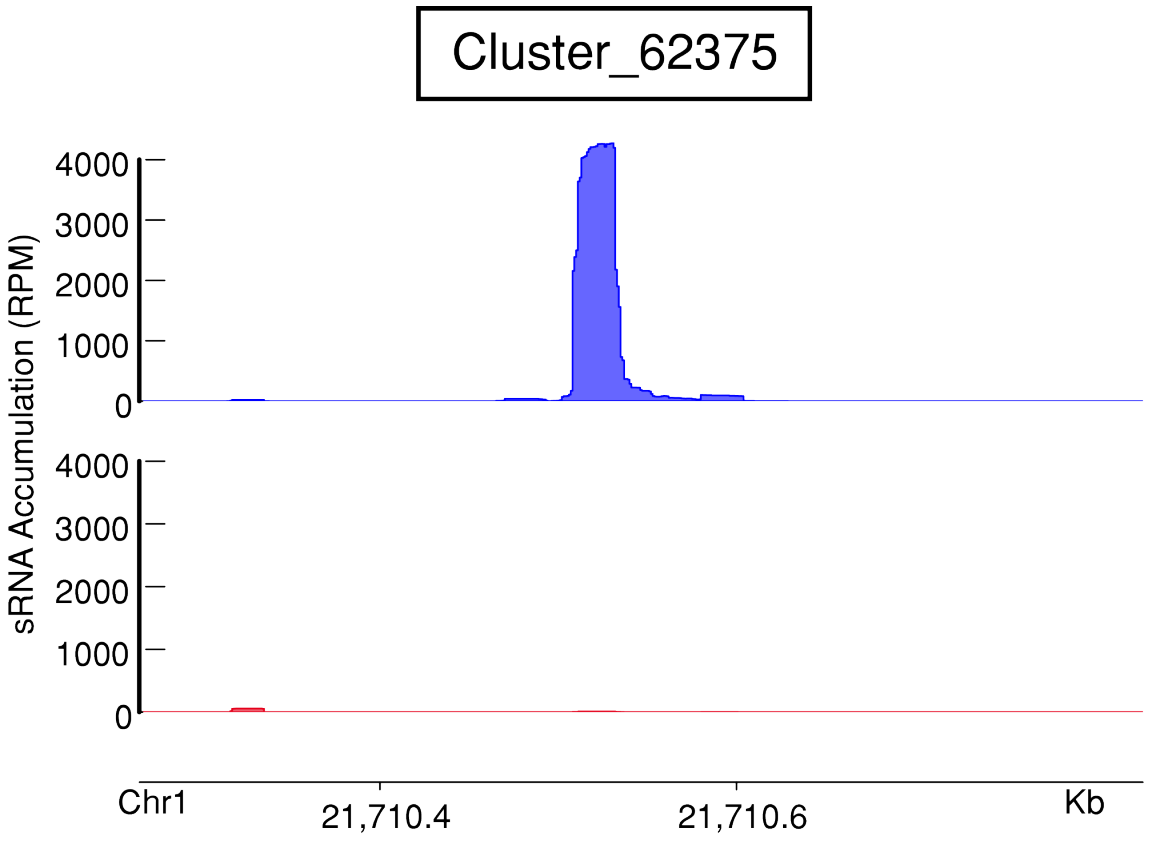
763


Cluster_54003
sRNA Accumulation (RPM)
400
300
200
100
0
400
300
200
100
0
Chr1
7,766.2
7,766.4
7,766.6
Kb

764


Cluster_59879
sRNA Accumulation (RPM)
70
60
50
40
30
20
10
0
70
60
50
40
30
20
10
0
Chr1
16,658.2
16,658.4
16,658.6
Kb

765


Cluster_62375
sRNA Accumulation (RPM)
4000
3000
2000
1000
0
4000
3000
2000
1000
0
Chr1
21,710.4
21,710.6
Kb

766

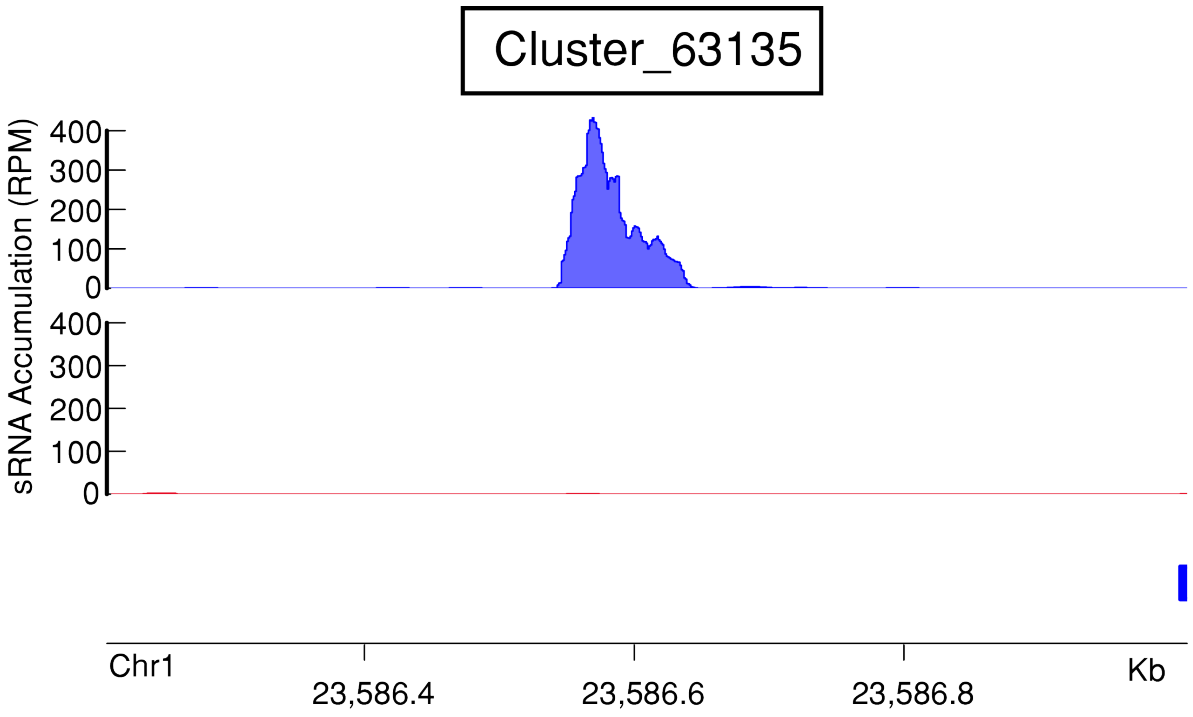


Figure S6. Small RNA coverage at *rdr6*-dependent, 24 nt-dominated small RNA clusters.

Charts represent small RNA accumulation in RPMs at a locus (regardless of the strandedness of a read). Top chart (blue) represents accumulation in wild-type libraries, and the bottom chart (red) represents the same in *rdr1/2/6* libraries. Arrows on the gene names represent the strandedness of that gene. None had significant phasing.

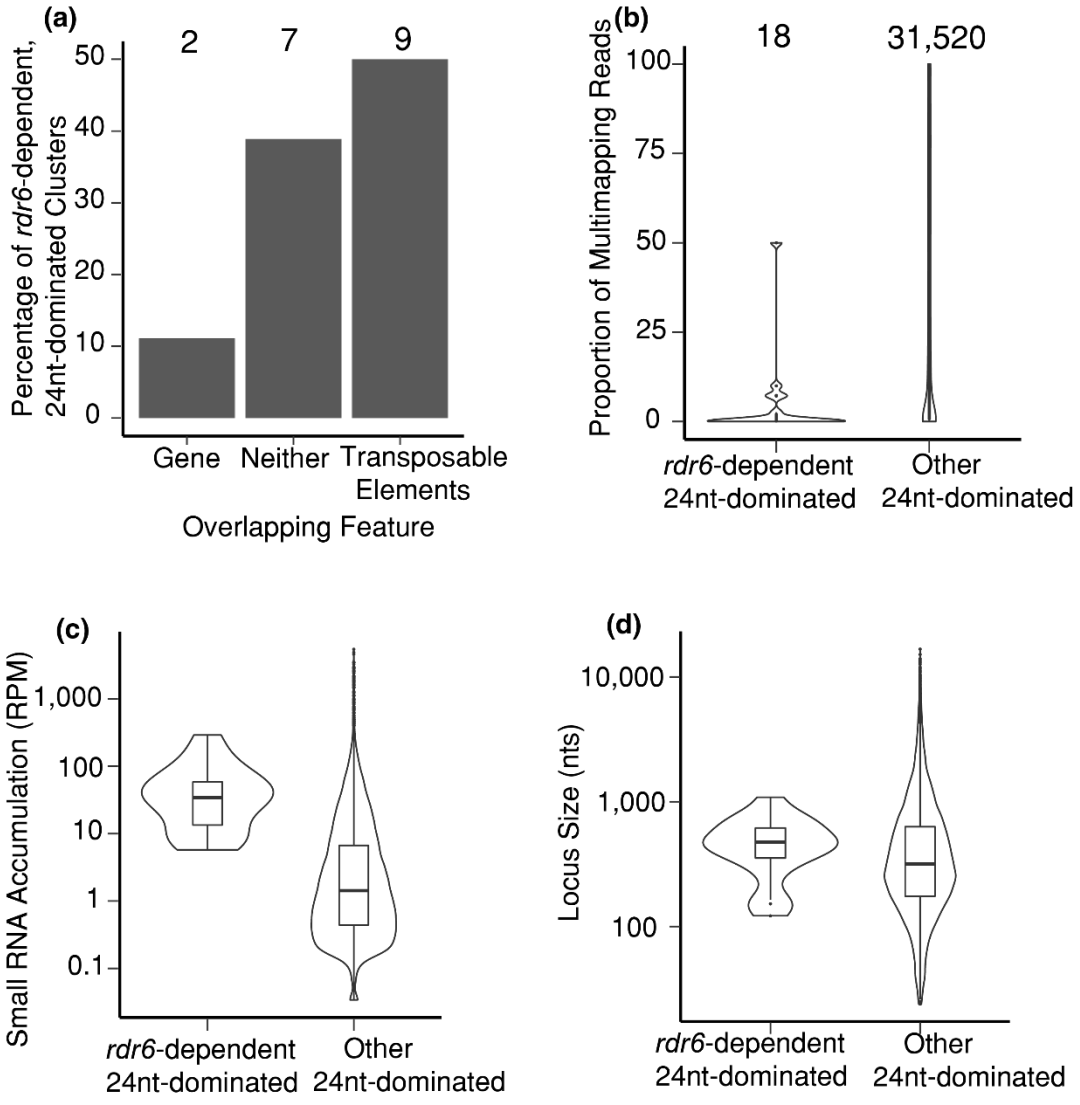


Figure S7. *rdr6*-dependent, 24 nt-dominated clusters' characteristics compared to other 24nt-dominated clusters.

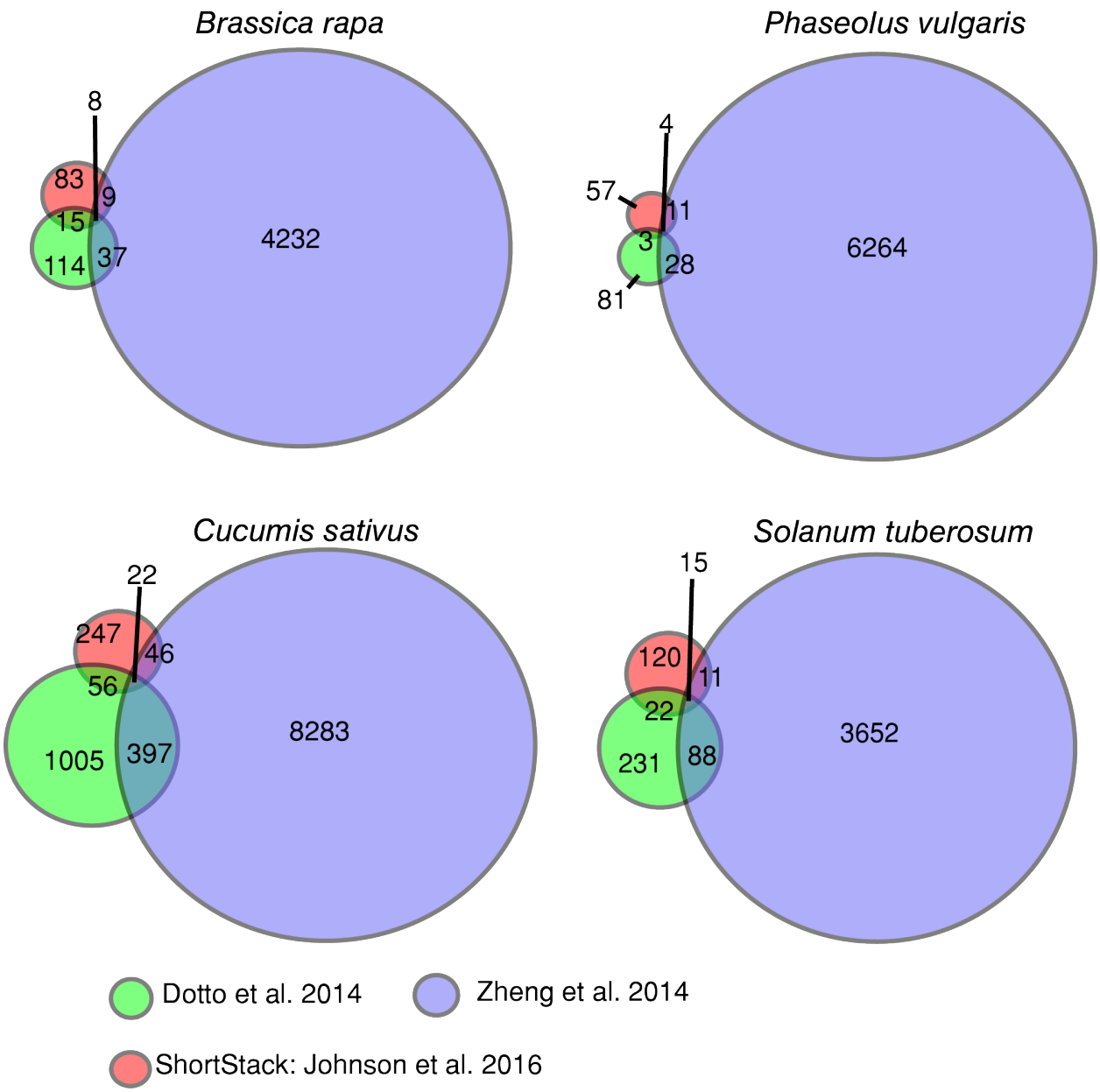
(a) The overlap between *rdr6*-dependent, 24 nt-dominated clusters with genes and transposons was calculated as in Figure 3a. Numbers at the top indicate the count in each category.

(b) The proportion of multi-mapping reads produced at *rdr6*-dependent, 24 nt-dominated clusters compared to other 24 nt-dominated clusters. Numbers at the top indicate the count in each category.

(c) Same as panel b except showing small RNA accumulation (in RPM). Amount in each category is the same as panel b.

(d) Same as panel b except showing length (in nts). Amount in each category is the same as panel b.

786



787

788 **Figure S8. Several 24 nt-dominated small RNA loci pass *PHAS*-detection algorithms in**
789 **four other eudicots.**

790 Venn diagram shows numbers of 24 nt-dominated loci that were called 'phased' by the indicated
791 algorithms. Species examined is shown above the graphs.

792 **Table S1. List of 25 known 21 nt *PHAS* loci in *A. thaliana.***

Chr.	Start	Stop	Locus Name	Source
Chr2	11721539	11722468	*TAS1a*	(Vazquez et al., 2004)
Chr1	18549204	18550042	*TAS1b*	(Allen et al., 2005)
Chr2	16537288	16538277	*TAS1c*	(Allen et al., 2005)
Chr2	16539384	16540417	*TAS2*	(Allen et al., 2005)
Chr3	5861491	5862437	*TAS3a*	(Montgomery et al., 2008)
Chr3	9415004	9422587	*TAS4*	(Rajagopalan et al., 2006)
Chr1	23299057	23300958	*PPR-At1g62910*	(Ronemus et al., 2006)
Chr1	23412730	23415149	*PPR-At1g63130*	(Ronemus et al., 2006)
Chr1	23306534	23308683	*PPR-At1g62930*	(Ronemus et al., 2006)
Chr1	23387631	23390816	*PPR-At1g63080*	(Ronemus et al., 2006)
Chr1	23507320	23509053	*PPR-At1g63400*	(Ronemus et al., 2006)
Chr1	23419396	23421579	*PPR-At1g63150*	(Ronemus et al., 2006)
Chr1	23385324	23387167	*PPR-At1g63070*	(Ronemus et al., 2006)
Chr1	23489840	23491519	*PPR-At1g63330*	(Ronemus et al., 2006)
Chr1	23176930	23179248	*PPR-At1g62590*	(Ronemus et al., 2006)
Chr5	15555156	15558732	*TIR-NBS-LRR-At5g38850*	(Howell et al., 2007)
Chr1	4368760	4371293	*AFB3*	(Si-Ammour et al., 2011)
Chr5	16638370	16641728	*ATCHX18*	(Howell et al., 2007)
Chr1	17886098	17892586	*AGO1*	(Axtell et al., 2006)
Chr3	23273116	23276375	*TIR1*	(Si-Ammour et al., 2011)
Chr4	1404887	1407139	*AFB1*	(Si-Ammour et al., 2011)
Chr3	9867845	9870640	*AFB2*	(Si-Ammour et al., 2011)
Chr5	15757717	15758109	*SLG*	(Chen et al., 2007)
Chr4	8380848	8383496	*CC-NBS-LRR-At4g14610*	(Zhai et al., 2011)
Chr4	8146345	8152131	*MET2*	(Chen et al., 2010)

793

794 **Reference Cited (Table S1)**

795

796 **Allen, E., Xie, Z., Gustafson, A. M., & Carrington, J. C.** (2005). microRNA-directed phasing

797 during trans-acting siRNA biogenesis in plants. *Cell*, *121*(2), 207–221.

798 https://doi.org/10.1016/j.cell.2005.04.004

Axtell, M. J., Jan, C., Rajagopalan, R., & Bartel, D. P. (2006). A two-hit trigger for siRNA biogenesis in plants. *Cell*, *127*(3), 565–577. https://doi.org/10.1016/j.cell.2006.09.032

Chen, H.-M., Chen, L.-T., Patel, K., Li, Y.-H., Baulcombe, D. C., & Wu, S.-H. (2010). 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(34), 15269–15274. https://doi.org/10.1073/pnas.1001738107

Chen, H.-M., Li, Y.-H., & Wu, S.-H. (2007). Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in Arabidopsis. *Proceedings of the National Academy of Sciences*, *104*(9), 3318–3323. https://doi.org/10.1073/pnas.0611119104

Howell, M. D., Fahlgren, N., Chapman, E. J., Cumbie, J. S., Sullivan, C. M., Givan, S. A., & Carrington, J. C. (2007). Genome-Wide Analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 Pathway in Arabidopsis Reveals Dependency on miRNA- and tasiRNA-Directed Targeting. *The Plant Cell*, *19*(3), 926–942. https://doi.org/10.1105/tpc.107.050062

Montgomery, T. A., Howell, M. D., Cuperus, J. T., Li, D., Hansen, J. E., Alexander, A. L., & Carrington, J. C. (2008). Specificity of ARGONAUTE7-miR390 Interaction and Dual Functionality in TAS3 Trans-Acting siRNA Formation. *Cell*, *133*(1), 128–141. https://doi.org/10.1016/j.cell.2008.02.033

Rajagopalan, R., Vaucheret, H., Trejo, J., & Bartel, D. P. (2006). A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes & Development*, *20*(24), 3407–3425. https://doi.org/10.1101/gad.1476406

Ronemus, M., Vaughn, M. W., & Martienssen, R. A. (2006). MicroRNA-Targeted and Small Interfering RNA–Mediated mRNA Degradation Is Regulated by Argonaute, Dicer, and RNA-Dependent RNA Polymerase in Arabidopsis. *The Plant Cell*, *18*(7), 1559–1574. https://doi.org/10.1105/tpc.106.042127

Si-Ammour, A., Windels, D., Arn-Bouldoires, E., Kutter, C., Ailhas, J., Meins, F., & Vazquez, F. (2011). miR393 and Secondary siRNAs Regulate Expression of the TIR1/AFB2 Auxin Receptor Clade and Auxin-Related Development of Arabidopsis Leaves. *Plant Physiology*, *157*(2), 683–691. https://doi.org/10.1104/pp.111.180083

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A. C., & Crété, P. (2004). Endogenous trans-Acting siRNAs Regulate the Accumulation of Arabidopsis mRNAs. *Molecular Cell*, *16*(1), 69–79. https://doi.org/10.1016/j.molcel.2004.09.028

Zhai, J., Jeong, D.-H., Paoli, E. D., Park, S., Rosen, B. D., Li, Y., & Meyers, B. C. (2011). MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes & Development*, *25*(23), 2540–2553. https://doi.org/10.1101/gad.177527.111

846 **Table S2. *A. thaliana* small RNA libraries used in this study**

Accession Number	Genotype	3' Adapter (First 8 nts)	Source
GSM2825283	Wild-type Replicate 1	TGGAATTC	(Polydore & Axtell, 2018)
GSM2825284	Wild-type Replicate 2	TGGAATTC	(Polydore & Axtell, 2018)
GSM2825285	Wild-type Replicate 3	TGGAATTC	(Polydore & Axtell, 2018)
GSM2825286	*rdr1-1/2-1/6-15* Replicate 1	TGGAATTC	(Polydore & Axtell, 2018)
GSM2825287	*rdr1-1/2-1/6-15* Replicate 2	TGGAATTC	(Polydore & Axtell, 2018)
GSM2825288	*rdr1-1/2-1/6-15* Replicate 3	TGGAATTC	(Polydore & Axtell, 2018)
GSM1533527	Wild-type Replicate 1	TGGAATTC	(Groth et al., 2014)
GSM1533528	Wild-type Replicate 2	TGGAATTC	(Groth et al., 2014)
GSM1533529	Wild-type Replicate 3	TGGAATTC	(Groth et al., 2014)
GSM1533542	*dcl3* Replicate 1	TGGAATTC	(Groth et al., 2014)
GSM1533543	*dcl3* Replicate 2	TGGAATTC	(Groth et al., 2014)
GSM1533544	*dcl3* Replicate 3	TGGAATTC	(Groth et al., 2014)
GSM1845210	Wild-type Replicate 1	AGATCGGA	(Elvira-Matelot et al., 2016)
GSM1845211	Wild-type Replicate 2	AGATCGGA	(Elvira-Matelot et al., 2016)
GSM1845212	Wild-type Replicate 3	AGATCGGA	(Elvira-Matelot et al., 2016)
GSM1845222	*dcl2-1/3-1/4-2t* Replicate 1	AGATCGGA	(Elvira-Matelot et al., 2016)
GSM1845223	*dcl2-1/3-1/4-2t* Replicate 2	AGATCGGA	(Elvira-Matelot et al., 2016)
GSM1845224	*dcl2-1/3-1/4-2t* Replicate 3	AGATCGGA	(Elvira-Matelot et al., 2016)
GSM1087973	Wild-type Replicate 1	TCGTATGC	(Jeong et al., 2013)
GSM1087974	Wild-type Replicate 2	TCGTATGC	(Jeong et al., 2013)
GSM1087975	*dcl1-7* Replicate 1	TCGTATGC	(Jeong et al., 2013)
GSM1087976	*dcl1-7* Replicate 2	TCGTATGC	(Jeong et al., 2013)
GSM1377370	Wild-type Replicate 1	TGGAATTC	(Li et al., 2014)
GSM1377371	Wild-type Replicate 2	TGGAATTC	(Li et al., 2014)
GSM1377372	*nrpd1-3* Replicate 1	TGGAATTC	(Li et al., 2014)
GSM1377373	*nrpd1-3* Replicate 2	TGGAATTC	(Li et al., 2014)
GSM1377376	*rdr2-1* Replicate 1	TGGAATTC	(Li et al., 2014)

GSM1377377	*rdr2-1* Replicate 2	TGGAATTC	(Li et al., 2014)
GSM2102962	Wild-type Replicate 1	TGGAATTC	(Panda et al., 2016)
GSM2102963	Wild-type Replicate 2	TGGAATTC	(Panda et al., 2016)
GSM2102965	*rdr6-15* Replicate 1	TGGAATTC	(Panda et al., 2016)
GSM2102462	*rdr6-15* Replicate 2	TGGAATTC	(Panda et al., 2016)
GSM893112	Wild-type Replicate 1	CACTCGGG	(Lee et al., 2012)
GSM893113	Wild-type Replicate 2	CACTCGGG	(Lee et al., 2012)
GSM893114	Wild-type Replicate 3	CACTCGGG	(Lee et al., 2012)
GSM893115	*nrpb1-1 (nrpe)* Replicate 1	CACTCGGG	(Lee et al., 2012)
GSM893116	*nrpb1-1 (nrpe)* Replicate 2	CACTCGGG	(Lee et al., 2012)
GSM893117	*nrpb1-1 (nrpe)* Replicate 3	CACTCGGG	(Lee et al., 2012)
GSM1668899	Wild-type Replicate 1	TGGAATTC	(Zhai et al., 2015)
GSM1668905	Wild-type Replicate 2	TGGAATTC	(Zhai et al., 2015)

Reference List (Table S2)

Elvira-Matelot, E., Hachet, M., Shamandi, N., Comella, P., Sáez-Vásquez, J., Zytnicki, M., & Vaucheret, H. (2016). Arabidopsis RNASE THREE LIKE2 Modulates the Expression of Protein-Coding Genes via 24-Nucleotide Small Interfering RNA-Directed DNA Methylation. *The Plant Cell Online*, *28*(2), 406–425. https://doi.org/10.1105/tpc.15.00540

Groth, M., Stroud, H., Feng, S., Greenberg, M. V. C., Vashisht, A. A., Wohlschlegel, J. A., & Ausin, I. (2014). SNF2 chromatin remodeler-family proteins FRG1 and -2 are required for RNA-directed DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(49), 17666–17671. https://doi.org/10.1073/pnas.1420515111

Jeong, D.-H., Thatcher, S. R., Brown, R. S. H., Zhai, J., Park, S., Rymarquis, L. A., & Green, P. J. (2013). Comprehensive Investigation of MicroRNAs Enhanced by Analysis of Sequence Variants, Expression Patterns, ARGONAUTE Loading, and Target Cleavage1[W][OA]. *Plant Physiology*, *162*(3), 1225–1245. https://doi.org/10.1104/pp.113.219873

863 **Lee, T., Gurazada, S. G. R., Zhai, J., Li, S., Simon, S. A., Matzke, M. A., & Meyers, B. C.**
864 (2012). RNA polymerase V-dependent small RNAs in Arabidopsis originate from small,
865 intergenic loci including most SINE repeats. *Epigenetics*, *7*(7), 781–795.
866 https://doi.org/10.4161/epi.20290
867 **Li, S., Vandivier, L. E., Tu, B., Gao, L., Won, S. Y., Li, S., & Chen, X.** (2014). Detection of
868 Pol IV/RDR2-dependent transcripts at the genomic scale in Arabidopsis reveals features
869 and regulation of siRNA biogenesis. *Genome Research*, gr.182238.114.
870 https://doi.org/10.1101/gr.182238.114
871 **Panda, K., Ji, L., Neumann, D. A., Daron, J., Schmitz, R. J., & Slotkin, R. K.** (2016). Full-
872 length autonomous transposable elements are preferentially targeted by expression-
873 dependent forms of RNA-directed DNA methylation. *Genome Biology*, *17*, 170.
874 https://doi.org/10.1186/s13059-016-1032-y
875 **Polydore, S., & Axtell, M. J**. (2018). Analysis of RDR1/RDR2/RDR6-independent small RNAs
876 in Arabidopsis thaliana improves MIRNA annotations and reveals unexplained types of
877 short interfering RNA loci. *The Plant Journal: For Cell and Molecular Biology*, *94*(6),
878 1051–1063. https://doi.org/10.1111/tpj.13919
879 **Zhai, J., Bischof, S., Wang, H., Feng, S., Lee, T., Teng, C., & Jacobsen, S. E.** (2015). One
880 precursor One siRNA model for Pol IV- dependent siRNAs Biogenesis. *Cell*, *163*(2),
881 445–455. https://doi.org/10.1016/j.cell.2015.09.032