

1 **Full title:**

2
3 **Exploiting somatic piRNAs in the whitefly, *Bemisia tabaci***
4 **enables a new mode of gene silencing through RNA feeding**

5
6 **Running title:**

7
8 **Facile piRNA-mediated gene silencing in the whitefly, *Bemisia***
9 ***tabaci***

10
11 Mosharrof Mondal¹, Judith K. Brown¹, Alex Flynt^{2,3}

12
13 ¹School of Plant Sciences, University of Arizona, Tucson, Arizona, 85721, United
14 States of America

15
16 ²Cellular and Molecular Biology, University of Southern Mississippi, Hattiesburg,
17 Mississippi, 39406, United States of America

18
19
20 Corresponding author: Alex Flynt

21
22 Email: alex.flynt@usm.edu

23
24
25
26
27 **Summary Blurb**

28 RNAi usually relies on Dicer produced siRNAs to induce gene silencing. In many
29 arthropods another type of RNAi is present in the soma—the piRNA pathway. This
30 work finds exploiting this biology is a viable alternative for gene knockdown.
31

Abstract

RNA interference (RNAi) promises to reshape pest control technologies by being nontoxic, biodegradable, and species-specific. However, due to the plastic nature of RNAi there is a significant variability in animal responses. In this study we investigate small RNA pathways and processing of ingested RNAi trigger molecules in a hemipteran plant pest and virus vector, the whitefly *Bemisia tabaci* (Genn.). Unlike *Drosophila*, the animal where the paradigm for insect RNAi technology was established, whitefly has abundant somatic piwi-associated RNAs (piRNAs). For many years this class of small RNA was thought to be germline restricted but is common in the soma of many invertebrates. We sought to exploit this for a novel gene silencing approach. The main principle of piRNA biogenesis is the recruitment of target RNA fragments into the pathway. As such we designed synthetic RNAs to possess complementarity to the loci we annotated. Following feeding of these exogenous piRNA triggers knockdown as effective as conventional siRNA-only approaches was observed. These results demonstrate a new approach for RNAi technology that could be applicable to dsRNA-recalcitrant pest species, which could be fundamental to realizing insecticidal RNAi against plant pests and/or vectors of plant pathogens.

Introduction

RNA interference (RNAi) technology has been shown to be applicable as a low-toxicity biopesticide to control agricultural insect pests and vectors of plant pathogens through silencing essential, biologically-relevant genes (Zotti and

55 Smagghe 2015). RNAi shows great potential to be highly species-specific and
56 thereby spares beneficial organisms, and is non-toxic to humans and other
57 animal consumers. The RNAi approach for insect pest/vector control relies on the
58 ingestion of long double-stranded RNAs (dsRNAs) to trigger gene silencing via
59 siRNA production following Dicer processing (Head, Carroll et al. 2017, Knorr,
60 Fishilevich et al. 2018). While a number of products are available, some
61 arthropod pests exhibit moderate or only minor sensitivity to dsRNA upon
62 ingestion (Yu, Christiaens et al. 2013, Zhu and Palli 2020). This suggests that to
63 fully realize this strategy across most or all arthropods, RNAi triggers may require
64 unique engineering relevant to each target species (Shukla, Kalsi et al. 2016,
65 Parsons, Mondal et al. 2018).

66
67 Here, RNAi pathways were investigated for the whitefly *Bemisia tabaci*
68 (Genn.) (Aleyrodidae, Hemiptera) to characterize the fundamental features that
69 might be exploited to improve RNAi approach(es). As a group, this whitefly is
70 considered a cryptic or sibling species. Although most *B. tabaci* are relatively
71 benign, at least two variants/cryptic species transmit plant viruses and are among
72 the most invasive species causing damage to crops grown in subtropical,
73 tropical, and mild temperate parts of the world (J K Brown, D R Frohlich et al.
74 1995, Brown 2010, Chen, Hasegawa et al. 2016, de Moya, Brown et al. 2019,
75 Grover, Jindal et al. 2019). Chemical pesticides can be toxic to the environment
76 and consumers of these products, and regularly have become ineffective when
77 resistance develops (Chen, Hasegawa et al. 2016). *B. tabaci* is closely related to

greenhouse and spiraling whiteflies, and several other related phloem-feeding pests/pathogen vectors, including aphids, mealybugs, and psyllids. Previous RNAi studies with *B. tabaci* have demonstrated gene silencing in response to long dsRNA feeding, however, processing modes of these molecules and those in other non-holometabolous insects have not been characterized at the level of small RNA effector populations (Jaubert-Possamai, Le Trionnaire et al. 2007, Zha, Peng et al. 2011, El-Shesheny, Hajeri et al. 2013, Thakur, Upadhyay et al. 2014, Li, Zhang et al. 2016, Wang, Wang et al. 2016, Vyas, Raza et al. 2017, Grover, Jindal et al. 2019, Kanakala, Kontsedalov et al. 2019). In this study the behavior of these molecules is investigated in the context of an extant RNAi mechanism in *B. tabaci*.

Multiple biogenesis modes are reported for animal small RNAs, generally though three main classes are recognized: microRNAs (miRNAs), small-interfering RNAs (siRNAs), and Piwi-associated RNAs (piRNAs) (Carthew and Sontheimer 2009). miRNAs are deeply conserved, short-hairpin derived RNAs that are present in the cells of nearly all metazoans (Bartel 2018). In contrast, the biology of the other two small RNA classes is highly variable, likely due their role in defense against invasive nucleic acids like transposable elements and viruses (Okamura 2012). Indeed, their overall roles in animals has been observed to diverge even among family members (Ozata, Gainetdinov et al. 2019). Each small RNA variety is sorted to distinct Argonaute (Ago)/Piwi proteins with Ago's binding siRNAs and miRNAs, and Piwi's binding piRNAs. siRNAs are 20-23nt

101 products of Dicer cleavage, usually from a long > 100nt dsRNA molecule. While
102 important to antiviral response many endogenous siRNA (endo-siRNAs) species
103 derived from hairpin RNAs (hpRNAs) or cis-NATs can be found in arthropod
104 genomes (Fagegaltier, Bougé et al. 2009, Lau, Robine et al. 2009, Claycomb
105 2014).

106
107 In comparison, piRNAs, are Dicer independent and produced by several
108 methods that include endonucleolytic “slicer” activity present in Piwi proteins
109 (Yamaguchi, Oe et al. 2020). They and are typically 26-30nt long and have a “U”
110 residue at their 5’ end. In *Drosophila* two biogenesis modes of piRNAs have been
111 observed that are mediated by three piwi proteins: PIWI, Aubergine (Aub), and
112 Argonaute 3 (Ago3) (Ozata, Gainetdinov et al. 2019). The “ping-pong” mode
113 involves alternating target RNA “slicing” by Piwi proteins. This is an amplifying
114 mechanism where processed RNAs are recruited as new piRNAs. Aub cleaves
115 RNAs that load into Ago3 as secondary piRNAs. Reciprocally, Ago3 substrates
116 load into Aub. The other mechanism relies on piwi proteins cleaving designated
117 transcripts that then become substrates for the RNase Zucchini (Zuc), which the
118 subsequently load into PIWI/Aub in *Drosophila* (Gainetdinov, Colpan et al. 2018).
119 In *Drosophila*, piRNAs are associated with germline, however, in many other
120 arthropods piRNAs are found in soma, including hemipterans (Lewis, Quarles et
121 al. 2018).

Relative to *Drosophila*, whitefly has additional Ago/Piwi proteins and exhibits somatic expression of Piwis. Using small RNA sequencing datasets miRNAs, endo-siRNAs, and piRNAs were annotated. Many of the loci resembled those described in *Drosophila*, however, we found many instances where there was production of both piRNAs and siRNAs. Using characteristics of these endogenous loci we designed RNAi triggers that exploit piRNA pathways. Gene silencing triggered by the piRNA pathway were equally as efficient as the siRNA mediated silencing of endogenous genes. Somatic piRNAs are widespread among insects, however, their application as a pest control tool is yet to be developed. Other hemipteran insect pests and vectors of pathogens to plants and humans, such as pea aphid and the kissing bug, respectively, have been found to produce somatic piRNAs (Brito, Julio et al. 2018, Lewis, Quarles et al. 2018). Insights of this study are expected to apply directly to these pests as well as many others.

Results

Whitefly RNAi pathways

Due to the divergent nature of siRNA and piRNA biology, species specific design is necessary to fully exploit these pathways for effective gene silencing. To characterize the RNAi pathways of whitefly we first sought to identify the collection of Ago/Piwi proteins encoded in the whitefly 'B biotype' (also known as MEAM1) genome (MEAM1v1.2) using existing annotations and BLAST to curate sequences (Fig 1A) (Chen, Hasegawa et al. 2016). These sequences were then

146 compared to Ago/Piwi proteins from *D. melanogaster*, *Tribolium castaneum* (red
147 flour beetle), and subset from *C. elegans*. We found seven members of this
148 family encoded in the whitefly genome. Three of the genes belong to the Ago
149 family with one clearly related to miRNA-loading factors. The two other Agos
150 group with the siRNA-associated Agos and appear to be a clade specific
151 duplication as they are not homologs of the two si-Agos seen in *T. castaneum*.
152 This is consistent with the diverging nature of siRNA biology and opens up the
153 possibility that si-Ago function in whitefly might be distinct from fruit flies and
154 beetles. The other four members belong to the piwi clade. One of the whitefly
155 Piwi proteins is a homolog of DmeAgo3 while the other three group with
156 DmePIWI/DmeAub. This indicates that ping-pong biogenesis is likely present in
157 these animals. The phasing piRNA pathway is also presumably operative with
158 apparent homologs of Zuc (Bta02312) and the RNA helicase Armitage
159 (Bta07189) (Ishizu, Kinoshita et al. 2019).

160
161 piRNAs are found in somatic tissues of many insect orders, including
162 hemipterans (Huang, Fejes Toth et al. 2017, Lewis, Quarles et al. 2018). To
163 verify if this is also pertinent for whitefly, we investigated RNAi factor expression
164 in the whitefly guts, salivary glands, and whole body (Fig 1B) (Cicero and Brown
165 2011). PolyA sequencing libraries from extirpated whitefly guts, salivary glands,
166 and whole body were mapped to the RNAi factor sequences from MEAM1v1.2,
167 per above (Chen, Hasegawa et al. 2016). The Alignments were then used to
168 calculate RPKM values for each transcript. Expression of BtaAgo1 and BtaAgo2

was found in all tissues, as well as two Piwi's (BtaPiwi3 and BtaAgo3), along with BtaZuc, and BtaArmi (Fig 1B). This contrasts with BtaPiwi2, which is enriched in whole body presumably due to the inclusion of RNAs originating from gonad tissues. This suggests that similar to other hemipterans, whiteflies have somatic piRNAs with both ping-pong and phasing piRNA biogenesis modes being present. Significantly, somatic piRNAs are likely present in whitefly gut, the tissue that would be the primary target of ingested RNAi trigger molecules.

To further investigate whitefly RNAi pathways endogenous small RNA populations from whole body mixed adults (male and female) were examined using small RNA sequencing libraries mapped against MEAM1v1.2. From this alignment we first annotated miRNAs using miRDeep2 (Friedlander, Mackowiak et al. 2012). Subtracting miRNA-derived reads from datasets would allow focus on non-miRNA small RNA loci such as endo-siRNAs and piRNAs, which unlike miRNAs might have whitefly specific biology. 202 miRNAs are identified with high confidence with 89 being conserved in *Drosophila* (Fig 1C and Table S1). We also identified 124 additional miRNAs which were classified as lower confidence miRNA candidates due to suboptimal features such as low expression or imprecise precursor cleavage patterns. The miRNA repertoire of the whitefly genome is similar in size to other insects (Kozomara, Birgaoanu et al. 2019). These results expand on prior miRNA annotations in whitefly due to the increased depth of datasets featured in this study.

Next we examined the size distribution of reads and found a bi-modal read size distribution with peaks at 22nt representing Dicer products (siRNAs and miRNAs) and 29-30nt (piRNAs) (Fig 1D). Among the Dicer products, roughly 56% derive from miRNAs. This shows piRNAs are more abundant relative to siRNAs in whitefly. To examine the modes of piRNA production we analyzed the abundance of read overlap pairs and the distance to 1U trailing reads (Fig 1D). During ping-pong biogenesis piRNA pairs are cleaved at the 10th base of the guiding RNA. Thus, when this mode is active piRNAs are found to overlap by 10 bases, which is clear in the dataset. Phasing piRNAs are biased to occur end to end and can be recognized by close proximity of trailing reads. Phasing is also evident in the alignment. The abundance of piRNAs is further reflected by the high proportion 1U reads in the size distribution. Simultaneously, a significant proportion of the reads also exhibit a “A” at the 10th base which would be found on ping pong pair reads due to pairing with 1U.

Non-miRNA, small RNA producing loci in whitefly

Using reads subtracted of miRNAs we annotated non-miRNA, small RNA-producing loci. 3873 regions were identified with a read depth greater than 40 and 500+ bp length (Fig 2A and Table S2). The ratio of the number of small (19-23nt) to long reads (25-30nt) was then calculated to distinguish whether the locus produced smaller siRNAs (19-23nt) or longer piRNAs (25-30nt). This showed the majority of loci appear to be piRNA generating. Only 50 loci had a ratio of small to long that was greater than “one”. Interestingly, the piRNA loci spanned regions

215 ranging in size from 500nt to 50kb, indicating diverse transcripts generate this
216 small RNA class (Yamanaka, Siomi et al. 2014). Apparent siRNA producing loci
217 tended to be shorter regions, of which the longest was about 4kb in size. We
218 then examined the distribution of read sizes at each locus (Fig 2B). Accumulation
219 was most clear in the piRNA range, with substantially less for siRNAs. However,
220 a minor signature of siRNA-sized reads could be seen at many loci. To confirm
221 that most loci were sites of piRNA production we examined read overlaps and
222 trailing 1U read distance, which shows evidence of ping-pong due to 10nt overlap
223 bias and phasing with juxtaposed trailing 1U reads, respectively. The exception
224 was ~100 loci that showed a greater accumulation of siRNAs.

225
226 To verify if these loci are sources of Dicer produced siRNAs we sought the
227 2nt overhang signature of RNase III processing (Fig 2C). Overlapping read pairs
228 between 15-31nt with this signature were quantified. Pairs were identified where
229 one strand (query read) of a potential duplex overlapped by less than two of its
230 entire length, which would occur with a 2nt overhang (Antoniewski 2014). All
231 potential combination of query and complementary target reads were quantified,
232 revealing that 22nt reads show the greatest evidence of Dicer processing, and
233 that this is likely size of Dicer2 products. The abundance of apparent Dicer
234 overlapping reads differed from the distribution of the reads in different size
235 ranges, validating this method of characterizing biogenesis. Interestingly, some
236 signal could also be seen in the 29-30nt sizes that likewise was not reflected in
237 the all read size distribution. This suggests a potential interaction between

siRNAs and piRNAs unlike what is reported in *Drosophila*, and is consistent with the frequent co-occurrence of siRNA sized and piRNA sized reads across all annotated loci (Fig 2B).

Next, we focused on the filter loci by expression to focus on the top 50 long read biased or short read biased loci (Fig S1, Table S1). The size distribution of reads for each locus was determined, which showed 28-30nt reads at long read loci, consistent with production of piRNAs (Fig S1). This contrasts with the short read loci, which show signal at 22nts and 29nts and 30nts, consistent with co-occurrence of piRNAs and siRNAs seen across all loci (Fig 2B). We also characterized the two groups of loci by length, expression, and 1U bias (Fig S1). This showed that long read loci are larger, have a greater bias towards 1U and greater expression, which are characteristics of piRNA clusters. Examining strand mapping showed that high expressing long read loci exhibit bias towards small RNA production from one strand indicating likely single-stranded precursor transcripts converted into phasing piRNAs (Fig S1) (Gainetdinov, Colpan et al. 2018). The short read loci are predominantly dual-stranded, which is suggestive of a double-stranded RNA precursor serving as a substrate for Dicer (Claycomb 2014). We also identified 22 hairpin RNA loci indicating that this variety of locus is present as a minority of the overall collection of whitefly siRNA-generating loci (Fig S2).

To predict the function of these 100 loci, reads aligning to these loci were mapped back to the whitefly genome permitting up to three mismatches. This alignment was then intersected to MEAM1v1.2 annotations (Fig 2D, E). The number of intersections was determined for each locus keeping mRNAs and TEs separate. Both long and short reads target mRNAs and TEs indicating possible roles for piRNAs and siRNAs not only in genome surveillance but also in gene regulatory networks. This is consistent with a proposed role for piRNAs in regulation of protein coding gene expression (Shamimuzzaman, Hasegawa et al. 2019). Taken together, this suggests that whitefly siRNAs and piRNAs are gene regulatory factors alongside miRNAs. These observations further reinforce the potential for exploiting these pathways for genetic technology that silences genic transcripts.

Whitefly endo-siRNA loci are also sources of piRNAs.

Prior work in whitefly has shown effectiveness of long dsRNA in gene silencing (de Paula, de Faria et al. 2015, Malik, Raza et al. 2016, Luo, Chen et al. 2017, Vyas, Raza et al. 2017, Grover, Jindal et al. 2019). The presumption is that these molecules are processed by Dicer into siRNAs. To better understand small RNAs simulated by fed dsRNA we used the computational approach described above that finds the 2nt overhang signature of RNase III cleavage in 20-23nt reads. Based on this, 76 loci exhibiting apparent Dicer processing were annotated (Fig 3A). When intersecting these Dicer loci with the high expressing long and short read loci (Fig S1), 42 short read loci and only one long read locus

283 have the Dicer processing signature (Fig 3A). Seqlogo analysis was also
284 performed on the 22nt Dicer reads showing a bias for 1U and a matched 20th
285 base A (20A). Among the other bases, 1G residues were disfavored along with
286 the paired 20C.

287
288 Next we inspected individual loci to understand their function and
289 biogenesis. The Dicer locus that overlapped with the one long read locus is an
290 interesting genomic site (Fig 3B). This region is a large phasing piRNA precursor
291 with an annotated, interior antisense transcript. The Dicer signature reads
292 coincide with this antisense transcript that seems to form a dsRNA with the
293 piRNA precursor. Other Dicer loci also arise from overlapping antisense
294 transcripts. Indeed, many cisNAT siRNAs are observed in the Dicer annotations,
295 with one such example shown from Scaffold1098 (Fig 3B).

296
297 Through curation of the annotations, loci were placed in five categories:
298 siRNA, cisNAT, No bias, piRNA, and piRNA cluster (Fig 3C) (Table S1). These
299 groupings were determined by evaluating dominant small RNA size and the
300 dominant processing signatures of read pairs—2nt overhangs for Dicer or 10nt
301 overlaps for ping-pong piRNA. The siRNA group are located in intergenic regions
302 and have a strong bias towards short reads that appear to be dicer processed.
303 The cisNATs were sites of siRNA production between opposing mRNAs as
304 showing in Fig 3B. In addition to these predominantly siRNA producing regions
305 that are similar to ones observed in *Drosophila* (Czech, Malone et al. 2008,

Ghildiyal, Seitz et al. 2008), many loci produce both siRNAs and piRNAs. These dual-identity loci could be grouped into one of three categories. One where there was equal production of siRNAs and piRNAs (No bias), a second for which some siRNAs were present, but piRNAs are dominant (piRNA), and the third group that harbors large piRNA clusters with only a minor production of siRNAs, the latter being similar to the locus shown in Fig 3B. These observations show that despite an Ago repertoire similar to *Drosophila* small RNA biogenesis in whitefly is distinct. This provides an opportunity to exploit these divergent activities for gene silencing and pest management.

Metabolism of exogenous dsRNA by whitefly

We then extended our evaluation of processing dsRNA transcripts to those introduced exogenously via feeding. Here we tested three off-target, synthetic dsRNAs dissolved in a sucrose solution fed through an artificial system. The RNAs cloned from genes of the potato psyllid *Bactericera cockerelli* (Sulc.) were fed to adult whiteflies from which small RNA and messenger RNA sequencing libraries were generated. Significant accumulation of reads arose exclusively from dsRNAs and not from other sections of the psyllid gene from which they were cloned (Fig 4A). However, only a fraction of the reads show a signature of Dicer processing based on 2nt overhangs. This suggests that most of the synthetic RNA was likely degraded with only a minority entering the siRNA pathway, which is reflected in the distribution of reads produced from the dsRNA sequences (Fig 4B). Only a modest peak was seen at 22nt with many more at

the 15nt size. The low efficiency is likely in part caused by dsRNA specific nucleases (dsRNases), which are common in hemipteran insects. Several dsRNases from gut and other tissues have been identified in whiteflies (Luo, Chen et al. 2017, Singh, Singh et al. 2017). Further, the alkaline pH of sternorrhynchan midgut is also compromised RNA integrity (Cristofolletti, Ribeiro et al. 2003, Molki, Ha et al. 2019).

Using these datasets, we sought to identify similarities between small RNAs derived from fed dsRNA and endogenously expressed siRNAs. Specifically reads were subsetted based on sequence content to find population where signatures of dicer processing were most evident (Fig 4C). This was guided by the seqlogo results of endogenous siRNAs that showed preference for 1U and depletion of 1G (Fig 3A). In unfiltered reads only a slight enrichment of 22nt RNAs was seen with no evidence of 2nt overhangs. Next, reads were extracted based on their 5' residue, which showed similar size distribution to the unfiltered library with the exception of 1G reads where there was no bias towards 22nt reads. However, for each subset no 2' overhang specific to 22nt reads was observed. The analysis was then extended to include not only the first base of the read but also the 20th base. When considering 1U/A/C(H) and 20A/U/G(D) greater abundance of 22nt reads was seen but still no substantial 2nt overhang Dicer signature. When 1U/A(W) and 20A/U(W) were examined an even greater enrichment of 22nt reads as well as 2nt overhang enrichment for this size was observed. For individual nucleotide pairs (1U-20A, 1A-20U, 1C-20G, 1G-20C)

read size and overhang enrichment increased, particularly for 1U-20A in size distribution and 1A-20U for 2nt overhang. This analysis provides a framework for computationally isolating siRNA processing signatures from degradation products, which is essential when considering exogenous dsRNA processing due to cloning of digestive contaminants.

To understand the physiological consequences of ingesting dsRNA we examined the effect on expression of the small RNA loci annotated in this study and protein coding genes. Studies in other animals suggest there may be competition between exogenous and endogenous small RNAs for biogenesis pathways, and that dsRNA may be recognized as a viral motif resulting in activation of defense pathways (Jelinek, Leonard et al. 2011). After feeding dsRNA, small RNA sequencing showed no significant change in endogenous small RNA expression compared to control (Fig 4D) (Table S3). For protein coding genes we observed about 500 transcripts that were differentially expressed based on a p-value ≤ 0.001 (Fig 4E) (Table S3). Only 20 of these genes exhibited a log(fold2) value greater than 2 or less than -2. All genes in this group have very low expression with 14 having unknown function. The genes having a known identity appear to be involved in basic metabolism or development. The one exception is a RNase H containing gene (Bta15726) that could be involved in an antiviral response. However, it is down regulated, which is inconsistent with being deployed to combat perceived viral infection. Thus, it would seem that whitefly does not mount an antiviral-type response to dsRNA. We

also observed no change in expression of RNAi factors such as the Ago and Dicer proteins, suggesting that whitefly can metabolize exogenous RNAi triggers without affecting its core RNAi processes. Taken together it appears that when ingested, the bulk of dsRNA is degraded with a small amount contributing to the siRNA pool, and that exposure to dsRNA has minimal impact on off-target gene expression in whitefly.

Exploiting somatic piRNAs in addition to siRNAs for gene silencing

In this study we found a significant population of piRNAs, which are more abundant than the endogenous siRNAs—the species exploited by existing RNAi approaches. The piRNAs also appear to be expressed in soma and show potential widespread control of mRNAs and not just a role in genome surveillance. This suggests that the piRNA pathway might be exploited to silence endogenous gene expression in whiteflies as an alternative method to the classic dsRNA-based siRNA strategy.

To trigger ectopic production, we engineered recombinant nucleic acids that take advantage of the major principle of piRNA biogenesis—recruitment of Piwi cleaved fragments into the pathway (Fig S3). We fused sequences from two loci annotated in this study, a piRNA bias locus (piRB-6) and siRNA-piRNA no bias locus (No bias-14) to target gene sequences. Both loci were among those that showed evidence of Dicer processing as well as piRNA production (Fig 3) (Fig S4, Table S1). We chose two different genes to target with these constructs:

Aquaporin1 (AQP1) and Alpha glucosidase1 (AGLU1), which were used in a previous study that yielded high gene knockdown via dsRNA (Vyas, Raza et al. 2017). To explore design principles, the positive strand of the locus was fused to AQP1 and the negative strand to AGLU1.

Using these constructs, both synthetic dsRNAs and ssRNAs (single-stranded RNA) were generated and fed to whiteflies in the artificial system described above. The concentration of RNAs (30 ng/ul) used was similar to what was previously fed to whitefly in dsRNA experiments (Vyas, Raza et al. 2017). Luciferase sequences fused to piRB-6/No_bias-14 were used as off-target controls. Following feeding access for 6 days, expression of target genes was assessed by qPCR (Fig 5A). As previously reported, dsRNAs elicited gene knockdown of 68-80%. Satisfyingly, the piRNA triggers showed a similar degree of gene silencing with reduction in target expression of 60-80%. This result was observed for both ssRNA and dsRNA triggers with both piRNA sequences (piRB-6, No_bias-14) and targets (AQP1, AGLU1). For AQP1, piRNA triggers were equal to dsRNA (conventional dsRNA), while AGLU1 was not as well down-regulated by the piRNA triggers relative to dsRNA suggesting inclusion of positive strand sequence might lead to superior knockdown. However, by combining ssRNA and dsRNA piRNA triggers for either sequence, gene silencing became comparable to conventional dsRNA for AGLU1. These results provide robust evidence that piRNA triggers, even those comprised of ssRNA, are capable of gene silencing in organisms that share RNAi biology with whiteflies.

421

422 Small RNAs were then sequenced to characterize the processing of the
423 piRNA triggers. Small RNAs were sequenced from animals fed piRB-6 dsRNAs
424 and ssRNAs targeted to both AQP1 and AGLU1 (Fig 5B-F, Fig 6). Reads
425 mapping to these triggers showed significant heterogeneity in read size with no
426 accumulation of a specific size, indicating the bulk of fed RNAs were degraded.
427 To identify potential small RNAs among the detritus we determined the relative
428 abundance of 1U reads, a characteristic of piRNAs as well as siRNAs (Fig 5B).
429 From this we found significantly more 26-30nt piRNA-sized reads. In the double-
430 stranded treatment a small peak possibly corresponding to 22nt siRNAs could be
431 observed, but not for the single-stranded piRNA triggers.

432

433 Next we focused on the identity of small RNAs produced against the target
434 gene. piRNA biogenesis could be observed for both triggers but more so for the
435 single-stranded versions (Fig 5C). Ping-pong processing was observable when
436 comparing the number of overlaps for different nucleotide combinations that can
437 form pairs: 1U/10A, 1A/10U, 1C/10G, and 1G/10C. Read pairs were determined
438 for all ranges of reads, and for those in piRNA sizes (28-30nt). The greatest
439 enrichment for 28-30nt reads was seen for those with the signature of ping-pong
440 piRNAs: 1U/10A. Phasing was also assessed for each strand of the piRNA
441 triggers (Fig 5 D&E). This biogenesis mechanism was evident for the transcribed
442 strand of single-stranded triggers, which is complementary to the target gene
443 (AQP1, AGLU1). For both strands of the double-stranded trigger and the

potential target derived reads in the single-stranded fed condition, less phasing was evident though a noticeable trend toward close proximity of 1U reads was seen. Next, we investigated whether siRNAs were processed from the triggers by examining 2nt overhangs in read populations as in Fig 4C. When reads with 1U/A/C-20U/A/G or 1U/A-20A/U were examined, the double-stranded trigger showed a greater number of 22nt Dicer signature reads (Fig 5F). Together these results show that regardless of whether the trigger is double-stranded or single-stranded piRNAs are produced. However, there is less piRNA production from dsRNAs. Presumably, accessing the piRNA pathway requires an unwinding step for dsRNAs mediated by gut or cellular helicases, reducing the entry of double-stranded triggers. In comparison, the double-stranded trigger give rise more production of siRNAs.

To understand the differences in target knockdown by the different piRNA trigger configurations we investigated the biogenesis of small RNAs from each. Before examining the exogenous triggers, we investigated more deeply small RNA production from the endogenous piRB-6 locus used to make the piRNA triggers (Fig 6A-D). This region shows clear piRNA phasing, ping-pong, and siRNA biogenesis (Fig 6A-C). Read alignments of each biogenesis mode were visualized at the locus (Fig 6D). 28-30nt 1U-10A reads overlapping by 10nt represent ping-pong reads. Phasing piRNAs are reads 28-30nt long with a 1U that didn't show a 10nt overlap. siRNAs are reads that start with 1U/1A and a 20A/20U also showing a 2nt overhang. At this locus the positive strand of the

locus shows nearly 20-fold accumulation of small RNAs. This is clearly due to phasing piRNAs on the positive strand, and only modest accumulation of ping-pong piRNAs and siRNAs on the negative strand.

The asymmetry of read expression at the piRB-6 locus appears to cause the difference in gene silencing for the two configurations of piRNA triggers (Fig 5A). This is apparent when the accumulations of small RNA types are examined for the gene targeted region of each trigger (Fig 6E-G). As with the endogenous loci we quantified 1U-10A ping-pong piRNAs, 1U non-ping-pong piRNAs (phasing), and siRNAs with 1U/A-20A/U. The AGLU1 trigger is composed of the antisense of piRB-6, and could be targeted by sense phasing piRNAs, ping-pong piRNAs, and siRNAs. For the single-stranded version of the AGLU1 trigger we observe significant accumulation of ping-pong piRNAs and phasing on the strand synthesized and fed. The ping-pong piRNAs complementary to the trigger are likely derived from the target gene, which is robustly silenced by this trigger. This contrasts with the double-stranded AGLU1 trigger, which shows that the off-target strand is much more robustly converted into small RNAs, particularly presumptive phased piRNAs. This explains the lower silencing efficiency for double-stranded AGLU1 trigger (Fig 5A). The off-target strand of the AGLU1 dsRNA trigger duplex is the strand that is phased in the endogenous locus.

This same phenomenon is seen in AQP1 triggers which sport the sense strand of piRB-6 for the on-target strand. For single-stranded AQP1 nearly all the

490 RNAs appear to be phasing piRNAs, and for the double-stranded version most of
491 the phasing piRNAs are on target. Both of these trigger versions lead to robust
492 gene silencing. These results indicate that a superior choice for piRNA trigger
493 design is to select the phased strand of piRNA loci to fuse with gene targeting
494 sequences. It also shows that the small population of endogenous antisense
495 ping-pong piRNAs or possibly even the siRNAs have a heightened role in
496 promoting phasing. This is an intriguing departure from *Drosophila* where trailing
497 piRNAs are produced downstream of a site of piwi protein-initiated cleavage.
498 Here it seems phasing of piRB-6 can be initiated internally because the region
499 cloned for these triggers only includes an interior section of the locus (Fig 6D). It
500 is also clear from these results that double-stranded triggers, expectedly, lead to
501 greater production of siRNAs.

503 Discussion

504 This study provides an in-depth analysis of the RNAi pathways in *B.*
505 *tabaci*, a hemipteran insect pest and plant virus vector, and offers a rationale
506 design of piRNA-based gene silencing biotechnology. Most significantly, we
507 show ingested RNAs can enter piRNA pathways, which opens up the possibility
508 for an entirely new strategy for gene silencing and potentially commercial
509 products. On a superficial level whitefly small RNAs seem similar to *Drosophila*.
510 There are three distinct types of small RNAs (miRNAs, siRNAs, and piRNAs), as
511 occur in fruit flies. However, upon close inspection, the biogenesis and function
512 of the endogenous small RNAs in whitefly are quite different. Our work reinforces

the consistent observation that non-miRNA RNAi pathways are fluid; clade specific duplication of the RNAi factors is common, even loss of an entire class of small RNA has occurred in several metazoan clades (Sarkies, Selkirk et al. 2015, Calcino, Fernandez-Valverde et al. 2018, Mondal, Klimov et al. 2018). Further, these findings illustrate the benefits of in-depth dissection of the RNAi biology for an evolutionarily and biologically different organisms, beyond those examined in model study systems, for developing genetic technology.

Through this comprehensive annotation of whitefly small RNA loci over 200 novel microRNAs are described, as well as 3878 siRNA or piRNA loci. Previously described configurations whitefly siRNA and piRNA loci were observed such as large single-stranded, phased piRNA loci and siRNA expressing cis-NAT and hpRNA loci (Fig 3) (Table S1, Fig S2). However, curation of loci found extensive evidence of siRNA and piRNA biogenesis occurring simultaneously at many loci. In fact, this appeared to be the rule for the majority of endogenous siRNA and piRNA genes, and seemingly, is related to a different biogenesis and function for whitefly siRNAs or piRNAs. In *Drosophila* and vertebrates, piRNAs mainly control TEs in germline, however, many of the piRNA pathway accessory proteins such as Rhino, Deadlock, Cutoff, Moonshiner from the Drosophilids are not conserved indicating that piRNAs are shaped to individual organism's biology in an evolutionary arm race between the piRNAs and their targets (Ozata, Gainetdinov et al. 2019). Indeed, it is predicted that abundant somatic piRNAs engage gene regulatory networks in many basal

arthropods, such as hemipterans, suggesting that this is the ancestral piRNA biology (Lewis, Quarles et al. 2018). This combined with observations that whitefly piRNAs respond to viral infection suggest diverse roles for these small RNAs in this insect (Shamimuzzaman, Hasegawa et al. 2019). Moreover, we find that phasing biogenesis can be initiated in the interior of loci which is distinct from the trigger/responder/trailing piRNA arrangement seen for phased *Drosophila* piRNAs.

While RNAi has been successful for controlling some pests like coleopterans (beetles), many other pests such as some lepidopterans (moths and butterflies) are unresponsive to exogenous RNAi trigger (Shukla, Kalsi et al. 2016, Parsons, Mondal et al. 2018). Penetrance of RNAi in hemipteran insects is moderate and higher dosage of dsRNA is required (Joga, Zotti et al. 2016). pH in the gut of the hemipteran insects is basic and presence of the nucleases in gut has been reported in whiteflies, aphids, and other hemipteran insects (Luo, Chen et al. 2017, Singh, Singh et al. 2017). We have noticed in this study that only a minority of the reads produced from the dsRNA trigger are siRNAs (Fig 4A&B). This could be attributed to low abundance of the intact dsRNA for uptake by the gut epithelium cells. We see a similar accumulation of degradation products with piRNA triggers. Interestingly, even with single-stranded RNA triggers we see significant accumulation of small RNA reads from synthetic RNA along with some antisense reads. The antisense reads we observe have dominant ping-pong and minor siRNA signature. How the single-stranded RNAs trigger production of

these molecules is not clear but could involve the recruitment of target mRNA into small RNA biogenesis. While this is standard behavior for piRNAs, it is not typical for siRNAs in organisms that don't possess Rdrp activity (Sarkies, Selkirk et al. 2015, Almeida, Andrade-Navarro et al. 2019, Pinzon, Bertrand et al. 2019). We view this result as a first glimpse at a heretofore unappreciated small RNA biogenesis mechanism that involves interaction between siRNA and piRNA biogenesis, consistent with the widespread co-occurrence at endogenous loci. Encountering such an unknown interaction is not entirely surprising as this study represents the first effort to characterize small RNA biogenesis on a per locus level in a non-holometabolous insect. This further reinforces the value of knowledge-based RNAi design gleaned from investigating exogenous trigger processing. In this study we provide clear rules for maximizing piRNA production, which could be fundamental to potent gene silencing technology aimed at aphids, mealybugs, psyllids, whiteflies, and other hemipterans.

As hemipteran insects respond to exogenous long dsRNA-mediated RNAi trigger only moderately, utilizing the gene silencing function of the piRNA pathway is exciting. These results show that in whitefly while there is significant sensitivity to dsRNA, there is very little physiological response to dsRNA feeding. Even the secreted gut dsRNases do not become transcriptionally activated by feeding. This will likely apply to other hemipteran herbivores with similar composition of RNAi pathways and dsRNases. We expect that piRNA triggers, single-stranded or double-stranded, will likewise be physiologically neutral. The

most promising result we report is that exogenous piRNA triggers are as effective as the siRNA versions. This study provides the first report of the exploitation of piRNAs as a feeding-based insect pest control strategy. Thus, this approach could become key for designing effective RNAi approaches against many insect pests that are found to be resistant to dsRNA-mediated RNAi. Finally, dsRNAs are capable of activating interferon response in humans and other vertebrates through binding of TLR3 receptors (Zhang, Xiang et al. 2016). Deploying ssRNA piRNA triggers as a pest control approach would avoid activating this pathway. As a result, beneficial, non-pest organisms in the field would also be spared from off-target effects of dsRNAs as piRNA triggers rely on the specific genomic sequence of the target species and would not be converted into siRNAs as would happened with dsRNA-based triggers. Taken together, these findings demonstrate the benefit of in-depth studies of non-model organismal RNAi biology, and demonstrate somatic piRNAs can be used for environmental RNAi.

Materials and Methods

Whitefly colony maintenance

Insects in this study came from the type *B. tabaci* Arizona B biotype (AZ-B) whitefly colony established in Brown lab in 1988 following its discovery on poinsettia plants in Tucson, Arizona (Vyas, Raza et al. 2017). For this study, AZ-B adult whiteflies were serially transferred to and reared on cotton *Gossypium hirsutum* L. cv Deltapine 5415) plants at the 8–10 leaf stage.

Phylogenetic tree construction

T. castaneum sequences of the Argonaute proteins were downloaded from NCBI (EFA09197.2, Ago1; EFA11590.1, Ago2a; EFA04626.2, Ago2b; EFA02921.1, Ago3; EFA07425.1, Piwi). Whitefly sequences were downloaded from *B. tabaci* MEMA1 genome database: <ftp://www.whiteflygenomics.org/pub/whitefly> and the Argonaute sequences were curated using blast and protein domain search tools InterPro and ScanProsite. The final Argonaute genes are: Bta01840, BtAgo1; Bta00938, BtAgo2a; Bta12142, BtAgo2b; Bta04637, BtAgo3; Bta00007, BtPiwi1; Bta00198, BtPiwi2; Bta08949, BtPiwi3. Annotated *D. melanogaster* and *C. elegans* sequences were also obtained from NCBI. The phylogenetic tree shown in Fig 1A was reconstructed in <http://www.phylogeny.fr> suite. Multiple sequence alignment was carried out using MUSCLE, phylogenetic tree was constructed by Maximum Likelihood (ML) method, and the ML tree was visualized by TreeDyn.

Cloning of whitefly sequences and *in vitro* transcription of ssRNA and dsRNA

AQP1 (KF377800.1) and AGLU1 (KF377803.1) sequences from a previous study (Vyas, Raza et al. 2017) were cloned in pGEMT-easy vector. The cloned plasmids were used as templates for PCRs, which were used in ssRNA and dsRNA synthesis reactions. For creating the fusion constructs (adding piRNA/siRNA sequences to the gene of interest (GOI): AQP1, AGLU1, and Luciferase sequences), SOEing PCR method was followed (S1 Text). 238 nt and

199 nt long region from No_bias-14 locus (Scaffold40734: 1537-1774, 1811-2009) were fused to left and right sites of the GOI respectively. From piRB-6 locus the left and right flanking sequences were 342 and 366 nt respectively (Scaffold185: 15168-15509, 15616-15981) (S1 Text) All Six phusion constructs were cloned into pGEMT-easy plasmid for double-stranded and single-stranded RNA synthesis. 231 nt luciferase gene sequence from psiCHECK™-2 (Promega, catalog number C8021) vector was cloned into the pGEMT-easy vector. ssRNA and dsRNA from the luciferase sequence was used as control RNA.

Each of the piRNA trigger constructs consisted of three parts, which were PCR amplified from whitefly cDNA using Phire Plant Direct PCR Master Mix (Catalog number: F160S) following manufacturer's instruction. During these PCRs 30 nt sequence from the left and right flanking regions were added to the gene of interest (AQP1, AGLU1, Luciferase) sequences by adding the sequences in the forward and reverse primers of the GOI. Gel extracted PCR products (GeneJET Gel Extraction Kit, Catalog number: K0691) were then ligated using two separate SOEing PCRs. First, the left flanking sequenced was attached to the GOI and gel extracted. In the second step the fusion product from the first step was ligated to the right flanking sequence. These sequences are provided in the S1 Text.

PCR products with T7 promoter sites on both strands were used for dsRNA synthesis while for ssRNA, PCR was carried by allowing the T7 promoter

site in one strand. PCR products were directly used to synthesize the synthetic RNAs using MEGAscript™ T7 Transcription Kit (Thermo Fisher Scientific, Catalog number AM1334) following manufacturer's protocol.

Oral delivery of the synthetic RNAs to whitefly, RNA extraction and qPCR

Using a hand-held aspirator 100 adult whiteflies were collected for each biological replicate from the colony and transferred to a plastic feeding chamber. 200 µl, 30 ng/µl RNA in 20% sucrose solution was sandwiched between two sterile Parafilm M® layers and feeding access to the solution was given to the insects for six days. On day 6, insects were collected for RNA extraction.

Total RNA was extracted following the standard Trizol RNA extraction method. The extracted RNAs were DNase I-treated (Invitrogen, DNA-Free™ kit, Lot 00522653) and 2 µg RNA was used for cDNA synthesis using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Lot 00692533). The TaqMan qPCR master mix (Applied Biosystems, Universal PCR Master Mix, Lot #1908161) was used for quantitative gene expression analysis using standard protocol. Whitefly 18S rRNA gene was used for normalizing the expression of the target genes. All qPCR primer sequences from a previous study were used in this study and can be found in S1 Text (Vyas, Raza et al. 2017). Each treatment and control groups of the synthetic RNA feeding was carried out using at least three independent biological replicates. $\Delta\Delta C_t$ method was used for gene

knockdown analysis. Student t-test and one-way ANOVA were used for statistical analysis in CFX Maestro software v1.1

mRNA library preparation, sequencing, and gene expression

Total RNAs were extracted using conventional Trizol RNA extraction method from different manually dissected tissues of whiteflies (gut, salivary gland, and whole body) (Cicero and Brown 2011). RNA integrity was confirmed using an Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Santa Clara, CA). Sequencing libraries were constructed using Illumina's TruSeq RNA Sample Preparation Kit v2, Catalog # RS-122-2002 (Set B). Using magnetic oligo (dT) beads, only poly(A) tail containing RNAs were separated from total RNA. Next, the mRNAs were fragmented by zinc treatment, and first-strand cDNA was synthesized from the fragmented RNAs using SuperScript II reverse transcriptase and random primers from Invitrogen. Then second-strand cDNA was synthesized, and Illumina multiple indexing adapters were ligated to the fragments. The remaining library construction steps were carried out following manufacturer's protocol. Quality filter and processing of the sequenced reads were performed using Illumina CASAVA v1.7.0, FastQC, and Trimmomatic. For each of the RNAi factors analyzed (Fig 1B), reads were mapped with bowtie2 to transcripts sequences from the whitefly genome database www.whiteflygenomics.org (Langdon 2015, Chen, Hasegawa et al. 2016). Bedtools was used to count read alignments to each transcript (Quinlan and Hall 2010).

696

697 **Small RNA library preparation and sequencing**

698 Total RNA was extracted from adult whiteflies using standard Trizol RNA
699 extraction protocol. Following DNase treatment, small RNA-seq libraries were
700 constructed using NEXTflex Small RNA-Seq Kit v3 (NOVA-5132-06). First, A 3'
701 4N Adenylated adapter was ligated to the 3' end and 5' standard illumina adapter
702 was ligated to the 5' end of the RNAs. Reverse Transcription was carried out on
703 the adapter ligated RNAs. Synthesized cDNAs were PCR amplified, and each
704 sample was barcoded with 17 Illumina-compatible in-line barcode. PCR products
705 were cleaned up by NEXTflex cleanup beads, and size selection of the DNAs
706 was performed on a Sage Scientific Blue Pippin. Sequencing was carried out on
707 a 1x75 flow cell on the NextSeq 500 platform (Illumina) at the Arizona State
708 University's genomics core and on a 2x150 flow cell NovaSeq platform at the
709 genomics core, University of Colorado, Denver.

710

711 **Bioinformatics pipelines used for small RNA analysis**

712 Small RNA reads were quality checked using FastQC and the adapter
713 sequences were cleaved and trimmed using fastx toolkit. Next, 15-35 nt size
714 reads were mapped to whitefly genome (MEAM1 genome v1.2) using bowtie with
715 default parameters (Chen, Hasegawa et al. 2016). The genome mapped reads
716 were isolated for the downstream analysis. mirDeep2 was used to annotate the
717 miRNAs (Friedlander, Chen et al. 2008). Initial calls by the algorithm were
718 manually inspected for recognized features of miRNAs (Berezikov, Liu et al.

2010). Annotations that showed evidence of mature and star strands in the appropriate Dicer cleavage register as well as significant expression were placed in the confident category. Deviation from these characteristics resulted in placement of annotation in the candidate category.

For non-miRNA annotations, small RNA reads, either taking all reads, 19-23nt sized reads, and 25-30nt reads were aligned using bowtie multimapping (-a -m 100) options. Bowtie was also used to identify the targets by allowing three mismatches. Size distributions were calculated with basic unix commands: awk, sort, uniq etc. Using bowtie alignments ping-pong overlap, piRNA phasing, and Dicer siRNA overhangs signatures were calculated as previously reported (Antoniewski 2014, Han, Wang et al. 2015). Samtools and Bedtools were used to count read alignments and identify high expressing regions and bias towards short and long read loci, as well as determine potential targets (Quinlan and Hall 2010). The R packages Scatterplot3d, sushi, heatmap2, pheatmap, and ggplot2 were used to draw the read density graphs (Kolde 2012, Phanstiel, Boyle et al. 2014, Warnes, Bolker et al. 2016, Wickham 2016, Ligges, Maechler et al. 2018). The seqlogo program was used to visualize nucleotide biases (Crooks, Hon et al. 2004). Read subsetting based on sequence content was carried out using standard linux tools (grep, awk, etc).

Acknowledgments

JKB and MM support is provided from Cotton Incorporated Project # 06-829

and USDA-NIFA Project ARZT-3026620-G25-574, GRANT#12469271. AF is supported by NSF MCB 1845978. Computational resources were provided by NSF: ACI 1626217

References

- Almeida, M. V., M. A. Andrade-Navarro and R. F. Ketting (2019). "Function and Evolution of Nematode RNAi Pathways." Noncoding RNA **5**(1).
- Antoniewski, C. (2014). "Computing siRNA and piRNA overlap signatures." Methods Mol Biol **1173**: 135-146.
- Antoniewski, C. (2014). Computing siRNA and piRNA Overlap Signatures. Animal Endo-SiRNAs: Methods and Protocols. A. Werner. New York, NY, Springer New York: 135-146.
- Bartel, D. P. (2018). "Metazoan MicroRNAs." Cell **173**(1): 20-51.
- Berezikov, E., N. Liu, A. S. Flynt, E. Hodges, M. Rooks, G. J. Hannon and E. C. Lai (2010). "Evolutionary flux of canonical microRNAs and mirtrons in *Drosophila*." Nat Genet **42**(1): 6-9; author reply 9-10.
- Brito, T., A. Julio, M. Berni, L. de Castro Poncio, E. S. Bernardes, H. Araujo, M. Sammeth and A. Pane (2018). "Transcriptomic and functional analyses of the piRNA pathway in the Chagas disease vector *Rhodnius prolixus*." PLoS Negl Trop Dis **12**(10): e0006760.
- Brown, J. K. (2010). Phylogenetic Biology of the Bemisia tabaci Sibling Species Group. Bemisia: Bionomics and Management of a Global Pest. P. A. Stansly and S. E. Naranjo. Dordrecht, Springer Netherlands: 31-67.
- Calcino, A. D., S. L. Fernandez-Valverde, R. J. Taft and B. M. Degnan (2018). "Diverse RNA interference strategies in early-branching metazoans." BMC Evol Biol **18**(1): 160.
- Carthew, R. W. and E. J. Sontheimer (2009). "Origins and Mechanisms of miRNAs and siRNAs." Cell **136**(4): 642-655.
- Chen, W., D. K. Hasegawa, N. Kaur, A. Klot, P. V. Pinheiro, J. Luan, M. C. Stensmyr, Y. Zheng, W. Liu, H. Sun, Y. Xu, Y. Luo, A. Kruse, X. Yang, S. Kontsedalov, G. Lebedev, T. W. Fisher, D. R. Nelson, W. B. Hunter, J. K. Brown, G. Jander, M. Cilia, A. E. Douglas, M. Ghanim, A. M. Simmons, W. M. Wintermantel, K.-S. Ling and Z. Fei (2016). "The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance." BMC Biology **14**(1): 110.

776 Chen, W., D. K. Hasegawa, N. Kaur, A. Kliot, P. V. Pinheiro, J. Luan, M. C. Stensmyr, Y.
 777 Zheng, W. Liu, H. Sun, Y. Xu, Y. Luo, A. Kruse, X. Yang, S. Kontsedalov, G. Lebedev, T. W.
 778 Fisher, D. R. Nelson, W. B. Hunter, J. K. Brown, G. Jander, M. Cilia, A. E. Douglas, M.
 779 Ghanim, A. M. Simmons, W. M. Wintermantel, K. S. Ling and Z. Fei (2016). "The draft
 780 genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights
 781 into virus transmission, host adaptation, and insecticide resistance." BMC Biol **14**(1):
 782 110.

783 Cicero, J. M. and J. K. Brown (2011). "Functional Anatomy of Whitefly Organs Associated
 784 with Squash Leaf Curl Virus (Geminiviridae: Begomovirus) Transmission by the B Biotype
 785 of *Bemisia tabaci* (Hemiptera: Aleyrodidae)." Annals of the Entomological Society of
 786 America **104**(2): 261-279.

787 Claycomb, J. M. (2014). "Ancient endo-siRNA pathways reveal new tricks." Curr Biol
 788 **24**(15): R703-715.

789 Cristofolletti, P. T., A. F. Ribeiro, C. Deraison, Y. Rahbe and W. R. Terra (2003). "Midgut
 790 adaptation and digestive enzyme distribution in a phloem feeding insect, the pea aphid
 791 *Acyrtosiphon pisum*." J Insect Physiol **49**(1): 11-24.

792 Crooks, G. E., G. Hon, J.-M. Chandonia and S. E. Brenner (2004). "WebLogo: a sequence
 793 logo generator." Genome research **14**(6): 1188-1190.

794 Czech, B., C. D. Malone, R. Zhou, A. Stark, C. Schlingehayde, M. Dus, N. Perrimon, M.
 795 Kellis, J. A. Wohlschlegel, R. Sachidanandam, G. J. Hannon and J. Brennecke (2008). "An
 796 endogenous small interfering RNA pathway in *Drosophila*." Nature **453**(7196): 798-802.

797 de Moya, R. S., J. K. Brown, A. D. Sweet, K. K. O. Walden, J. R. Paredes-Montero, R. M.
 798 Waterhouse and K. P. Johnson (2019). "Nuclear Orthologs Derived from Whole Genome
 799 Sequencing Indicate Cryptic Diversity in the *Bemisia tabaci* (Insecta: Aleyrodidae)
 800 Complex of Whiteflies." Diversity **11**(9): 151.

801 de Paula, N. T., J. C. de Faria and F. J. Aragao (2015). "Reduction of viral load in whitefly
 802 (*Bemisia tabaci* Gen.) feeding on RNAi-mediated bean golden mosaic virus resistant
 803 transgenic bean plants." Virus Res **210**: 245-247.

804 El-Shesheny, I., S. Hajeri, I. El-Hawary, S. Gowda and N. Killiny (2013). "Silencing
 805 abnormal wing disc gene of the Asian citrus psyllid, *Diaphorina citri* disrupts adult wing
 806 development and increases nymph mortality." PLoS One **8**(5): e65392.

807 Fagegaltier, D., A.-L. Bougé, B. Berry, É. Poisot, O. Sismeiro, J.-Y. Coppée, L. Théodore, O.
 808 Voinnet and C. Antoniewski (2009). "The endogenous siRNA pathway is involved in
 809 heterochromatin formation in *Drosophila* species." Proceedings of the National Academy of Sciences
 810 **106**(50): 21258-21263.

812 Friedlander, M. R., W. Chen, C. Adamidi, J. Maaskola, R. Einspanier, S. Knespel and N.
813 Rajewsky (2008). "Discovering microRNAs from deep sequencing data using miRDeep."
814 Nat Biotechnol **26**(4): 407-415.

815 Friedlander, M. R., S. D. Mackowiak, N. Li, W. Chen and N. Rajewsky (2012). "miRDeep2
816 accurately identifies known and hundreds of novel microRNA genes in seven animal
817 clades." Nucleic Acids Res **40**(1): 37-52.

818 Gainetdinov, I., C. Colpan, A. Arif, K. Cecchini and P. D. Zamore (2018). "A Single
819 Mechanism of Biogenesis, Initiated and Directed by PIWI Proteins, Explains piRNA
820 Production in Most Animals." Molecular cell **71**(5): 775-790.e775.

821 Gainetdinov, I., C. Colpan, A. Arif, K. Cecchini and P. D. Zamore (2018). "A Single
822 Mechanism of Biogenesis, Initiated and Directed by PIWI Proteins, Explains piRNA
823 Production in Most Animals." Mol Cell **71**(5): 775-790 e775.

824 Ghildiyal, M., H. Seitz, M. D. Horwich, C. Li, T. Du, S. Lee, J. Xu, E. L. Kittler, M. L. Zapp, Z.
825 Weng and P. D. Zamore (2008). "Endogenous siRNAs derived from transposons and
826 mRNAs in Drosophila somatic cells." Science **320**(5879): 1077-1081.

827 Grover, S., V. Jindal, G. Banta, C. N. T. Taning, G. Smagghe and O. Christiaens (2019).
828 "Potential of RNA interference in the study and management of the whitefly, Bemisia
829 tabaci." Arch Insect Biochem Physiol **100**(2): e21522.

830 Han, B. W., W. Wang, P. D. Zamore and Z. Weng (2015). "piPipes: a set of pipelines for
831 piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq,
832 ChIP-seq and genomic DNA sequencing." Bioinformatics **31**(4): 593-595.

833 Head, G. P., M. W. Carroll, S. P. Evans, D. M. Rule, A. R. Willse, T. L. Clark, N. P. Storer, R.
834 D. Flannagan, L. W. Samuel and L. J. Meinke (2017). "Evaluation of SmartStax and
835 SmartStax PRO maize against western corn rootworm and northern corn rootworm:
836 efficacy and resistance management." Pest Manag Sci **73**(9): 1883-1899.

837 Huang, X., K. Fejes Toth and A. A. Aravin (2017). "piRNA Biogenesis in Drosophila
838 melanogaster." Trends Genet **33**(11): 882-894.

839 Ishizu, H., T. Kinoshita, S. Hirakata, C. Komatsuzaki and M. C. Siomi (2019). "Distinct and
840 Collaborative Functions of Yb and Armitage in Transposon-Targeting piRNA Biogenesis."
841 Cell Rep **27**(6): 1822-1835 e1828.

842 J K Brown, a. D R Frohlich and R. C. Rosell (1995). "The Sweetpotato or Silverleaf
843 Whiteflies: Biotypes of Bemisia tabaci or a Species Complex?" Annual Review of
844 Entomology **40**(1): 511-534.

845 Jaubert-Possamai, S., G. Le Trionnaire, J. Bonhomme, G. K. Christophides, C. Rispe and D.
846 Tagu (2007). "Gene knockdown by RNAi in the pea aphid Acyrthosiphon pisum." BMC
847 Biotechnol **7**: 63.

848 Jelinek, I., J. N. Leonard, G. E. Price, K. N. Brown, A. Meyer-Manlapat, P. K. Goldsmith, Y.
849 Wang, D. Venzon, S. L. Epstein and D. M. Segal (2011). "TLR3-specific double-stranded
850 RNA oligonucleotide adjuvants induce dendritic cell cross-presentation, CTL responses,
851 and antiviral protection." *J Immunol* **186**(4): 2422-2429.

852 Joga, M. R., M. J. Zotti, G. Smagghe and O. Christiaens (2016). "RNAi Efficiency, Systemic
853 Properties, and Novel Delivery Methods for Pest Insect Control: What We Know So Far."
854 *Front Physiol* **7**: 553.

855 Kanakala, S., S. Kontsedalov, G. Lebedev and M. Ghanim (2019). "Plant-Mediated
856 Silencing of the Whitefly *Bemisia tabaci* Cyclophilin B and Heat Shock Protein 70 Impairs
857 Insect Development and Virus Transmission." *Front Physiol* **10**: 557.

858 Knorr, E., E. Fishilevich, L. Tenbusch, M. L. F. Frey, M. Rangasamy, A. Billion, S. E.
859 Worden, P. Gandra, K. Arora, W. Lo, G. Schulenberg, P. Valverde-Garcia, A. Vilcinskis
860 and K. E. Narva (2018). "Gene silencing in *Tribolium castaneum* as a tool for the targeted
861 identification of candidate RNAi targets in crop pests." *Sci Rep* **8**(1): 2061.

862 Kolde, R. (2012). "Pheatmap: pretty heatmaps." **61**(926): 915.

863 Kozomara, A., M. Birgaoanu and S. Griffiths-Jones (2019). "miRBase: from microRNA
864 sequences to function." *Nucleic Acids Res* **47**(D1): D155-D162.

865 Langdon, W. B. (2015). "Performance of genetic programming optimised Bowtie2 on
866 genome comparison and analytic testing (GCAT) benchmarks." *BioData Min* **8**(1): 1.

867 Lau, N. C., N. Robine, R. Martin, W. J. Chung, Y. Niki, E. Berezikov and E. C. Lai (2009).
868 "Abundant primary piRNAs, endo-siRNAs, and microRNAs in a *Drosophila* ovary cell
869 line." *Genome Res* **19**(10): 1776-1785.

870 Lewis, S. H., K. A. Quarles, Y. Yang, M. Tanguy, L. Frézal, S. A. Smith, P. P. Sharma, R.
871 Cordaux, C. Gilbert, I. Giraud, D. H. Collins, P. D. Zamore, E. A. Miska, P. Sarkies and F. M.
872 Jiggins (2018). "Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence
873 against transposable elements." *Nat Ecol Evol* **2**(1): 174-181.

874 Li, X., F. Zhang, B. Coates, Y. Zhang, X. Zhou and D. Cheng (2016). "Comparative profiling
875 of microRNAs in the winged and wingless English grain aphid, *Sitobion avenae* (F.)
876 (Homoptera: Aphididae)." *Sci Rep* **6**: 35668.

877 Ligges, U., M. Maechler, S. Schnackenberg and M. U. Ligges (2018). "Package
878 'scatterplot3d'." Recuperado de [https://cran.rproject.](https://cran.rproject.org/web/packages/scatterplot3d/scatterplot3d.pdf)
879 [org/web/packages/scatterplot3d/scatterplot3d.pdf](https://cran.rproject.org/web/packages/scatterplot3d/scatterplot3d.pdf).

880 Luo, Y., Q. Chen, J. Luan, S. H. Chung, J. Van Eck, R. Turgeon and A. E. Douglas (2017).
881 "Towards an understanding of the molecular basis of effective RNAi against a global
882 insect pest, the whitefly *Bemisia tabaci*." *Insect Biochem Mol Biol* **88**: 21-29.

883 Malik, H. J., A. Raza, I. Amin, J. A. Scheffler, B. E. Scheffler, J. K. Brown and S. Mansoor
884 (2016). "RNAi-mediated mortality of the whitefly through transgenic expression of

double-stranded RNA homologous to acetylcholinesterase and ecdysone receptor in tobacco plants." *Sci Rep* **6**: 38469.

Molki, B., P. T. Ha, A. L. Cohen, D. W. Crowder, D. R. Gang, A. Omsland, J. K. Brown and H. Beyenal (2019). "The infection of its insect vector by bacterial plant pathogen "Candidatus Liberibacter solanacearum" is associated with altered vector physiology." *Enzyme Microb Technol* **129**: 109358.

Mondal, M., P. Klimov and A. S. Flynt (2018). "Rewired RNAi-mediated genome surveillance in house dust mites." *PLoS Genet* **14**(1): e1007183.

Okamura, K. (2012). "Diversity of animal small RNA pathways and their biological utility." *Wiley Interdiscip Rev RNA* **3**(3): 351-368.

Ozata, D. M., I. Gainetdinov, A. Zoch, D. O'Carroll and P. D. Zamore (2019). "PIWI-interacting RNAs: small RNAs with big functions." *Nat Rev Genet* **20**(2): 89-108.

Parsons, K. H., M. H. Mondal, C. L. McCormick and A. S. Flynt (2018). "Guanidinium-Functionalized Interpolyelectrolyte Complexes Enabling RNAi in Resistant Insect Pests." *Biomacromolecules* **19**(4): 1111-1117.

Phanstiel, D. H., A. P. Boyle, C. L. Araya and M. P. Snyder (2014). "Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures." *Bioinformatics* **30**(19): 2808-2810.

Pinzon, N., S. Bertrand, L. Subirana, I. Busseau, H. Escriva and H. Seitz (2019). "Functional lability of RNA-dependent RNA polymerases in animals." *PLoS Genet* **15**(2): e1007915.

Quinlan, A. R. and I. M. Hall (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." *Bioinformatics* **26**(6): 841-842.

Sarkies, P., M. E. Selkirk, J. T. Jones, V. Blok, T. Boothby, B. Goldstein, B. Hanelt, A. Ardila-Garcia, N. M. Fast, P. M. Schiffer, C. Kraus, M. J. Taylor, G. Koutsovoulos, M. L. Blaxter and E. A. Miska (2015). "Ancient and novel small RNA pathways compensate for the loss of piRNAs in multiple independent nematode lineages." *PLoS Biol* **13**(2): e1002061.

Shamimuzzaman, M., D. K. Hasegawa, W. Chen, A. M. Simmons, Z. Fei and K. S. Ling (2019). "Genome-wide profiling of piRNAs in the whitefly Bemisia tabaci reveals cluster distribution and association with begomovirus transmission." *PLoS One* **14**(3): e0213149.

Shukla, J. N., M. Kalsi, A. Sethi, K. E. Narva, E. Fishilevich, S. Singh, K. Mogilicherla and S. R. Palli (2016). "Reduced stability and intracellular transport of dsRNA contribute to poor RNAi response in lepidopteran insects." *RNA Biol* **13**(7): 656-669.

920 Singh, I. K., S. Singh, K. Mogilicherla, J. N. Shukla and S. R. Palli (2017). "Comparative
921 analysis of double-stranded RNA degradation and processing in insects." Sci Rep **7**(1):
922 17059.

923 Thakur, N., S. K. Upadhyay, P. C. Verma, K. Chandrashekar, R. Tuli and P. K. Singh (2014).
924 "Enhanced whitefly resistance in transgenic tobacco plants expressing double stranded
925 RNA of v-ATPase A gene." PLoS One **9**(3): e87235.

926 Vyas, M., A. Raza, M. Y. Ali, M. A. Ashraf, S. Mansoor, A. A. Shahid and J. K. Brown
927 (2017). "Knock down of Whitefly Gut Gene Expression and Mortality by Orally Delivered
928 Gut Gene-Specific dsRNAs." PLoS One **12**(1): e0168921.

929 Wang, B., L. Wang, F. Chen, X. Yang, M. Ding, Z. Zhang, S. S. Liu, X. W. Wang and X. Zhou
930 (2016). "MicroRNA profiling of the whitefly Bemisia tabaci Middle East-Aisa Minor I
931 following the acquisition of Tomato yellow leaf curl China virus." Virology **13**: 20.

932 Warnes, M. G. R., B. Bolker, L. Bonebakker, R. Gentleman and W. Huber (2016).
933 "Package 'gplots'." Various R Programming Tools for Plotting Data.

934 Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New
935 York.

936 Yamaguchi, S., A. Oe, K. M. Nishida, K. Yamashita, A. Kajiya, S. Hirano, N. Matsumoto, N.
937 Dohmae, R. Ishitani, K. Saito, H. Siomi, H. Nishimasu, M. C. Siomi and O. Nureki (2020).
938 "Crystal structure of Drosophila Piwi." Nat Commun **11**(1): 858.

939 Yamanaka, S., M. C. Siomi and H. Siomi (2014). "piRNA clusters and open chromatin
940 structure." Mob DNA **5**: 22.

941 Yu, N., O. Christiaens, J. Liu, J. Niu, K. Cappelle, S. Caccia, H. Huvenne and G. Smagghe
942 (2013). "Delivery of dsRNA for RNAi in insects: an overview and future directions." Insect
943 Sci **20**(1): 4-14.

944 Zha, W., X. Peng, R. Chen, B. Du, L. Zhu and G. He (2011). "Knockdown of midgut genes
945 by dsRNA-transgenic plant-mediated RNA interference in the hemipteran insect
946 Nilaparvata lugens." PLoS One **6**(5): e20504.

947 Zhang, L., W. Xiang, G. Wang, Z. Yan, Z. Zhu, Z. Guo, R. Sengupta, A. F. Chen, P. A.
948 Loughran, B. Lu, Q. Wang and T. R. Billiar (2016). "Interferon beta (IFN-beta) Production
949 during the Double-stranded RNA (dsRNA) Response in Hepatocytes Involves
950 Coordinated and Feedforward Signaling through Toll-like Receptor 3 (TLR3), RNA-
951 dependent Protein Kinase (PKR), Inducible Nitric Oxide Synthase (iNOS), and Src
952 Protein." J Biol Chem **291**(29): 15093-15107.

953 Zhu, K. Y. and S. R. Palli (2020). "Mechanisms, Applications, and Challenges of Insect RNA
954 Interference." Annu Rev Entomol **65**: 293-311.

955 Zotti, M. J. and G. Smagghe (2015). "RNAi Technology for Insect Management and
956 Protection of Beneficial Insects from Diseases: Lessons, Challenges and Risk
957 Assessments." Neotrop Entomol **44**(3): 197-213.

958

959

960

961

Figures

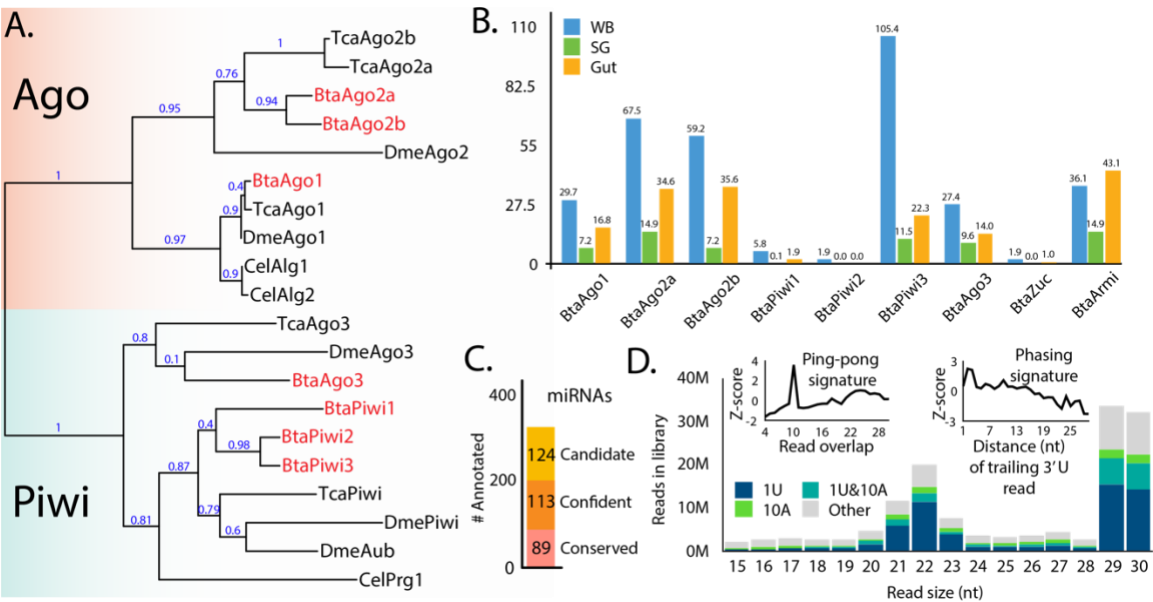


Figure 1. RNAi pathways in whitefly (*B. tabaci*).

A) Relatedness of Argonaute and Piwi (Ago/Piwi) proteins from whitefly (*Bta*) to orthologs in *Drosophila* (*Dme*), *Tripterygium* (*Tca*), and select family members from *C. elegans* (*Cel*). Ago and Piwi clades highlighted by colored boxes, and whitefly genes in red text. Phylogenetic tree was constructed using maximum likelihood method. Branch support values shown at nodes. **B)** Expression determined by RPKM of whitefly Ago/Piwi proteins in whole body (WB), gut, and salivary gland (SG). **C)** Numbers of microRNAs (miRNAs) annotated in this study. Loci are categorized into those conserved with *Drosophila*, novel highly confident, and lower confidence candidates. **D)** Distribution of small RNA read sizes mapping to the whitefly genome (MEAM v1.2), and piRNA biogenesis modes. Left inset shows read overlap Z-scores to demonstrate the ping-pong piRNA signature of 10nt overlaps, and right panel distance to 3' 1U reads showing the phasing signature with proximal trailing reads. Bars in the size distribution are colored based on the portion of reads with the sequence identity indicated in the inset legend.

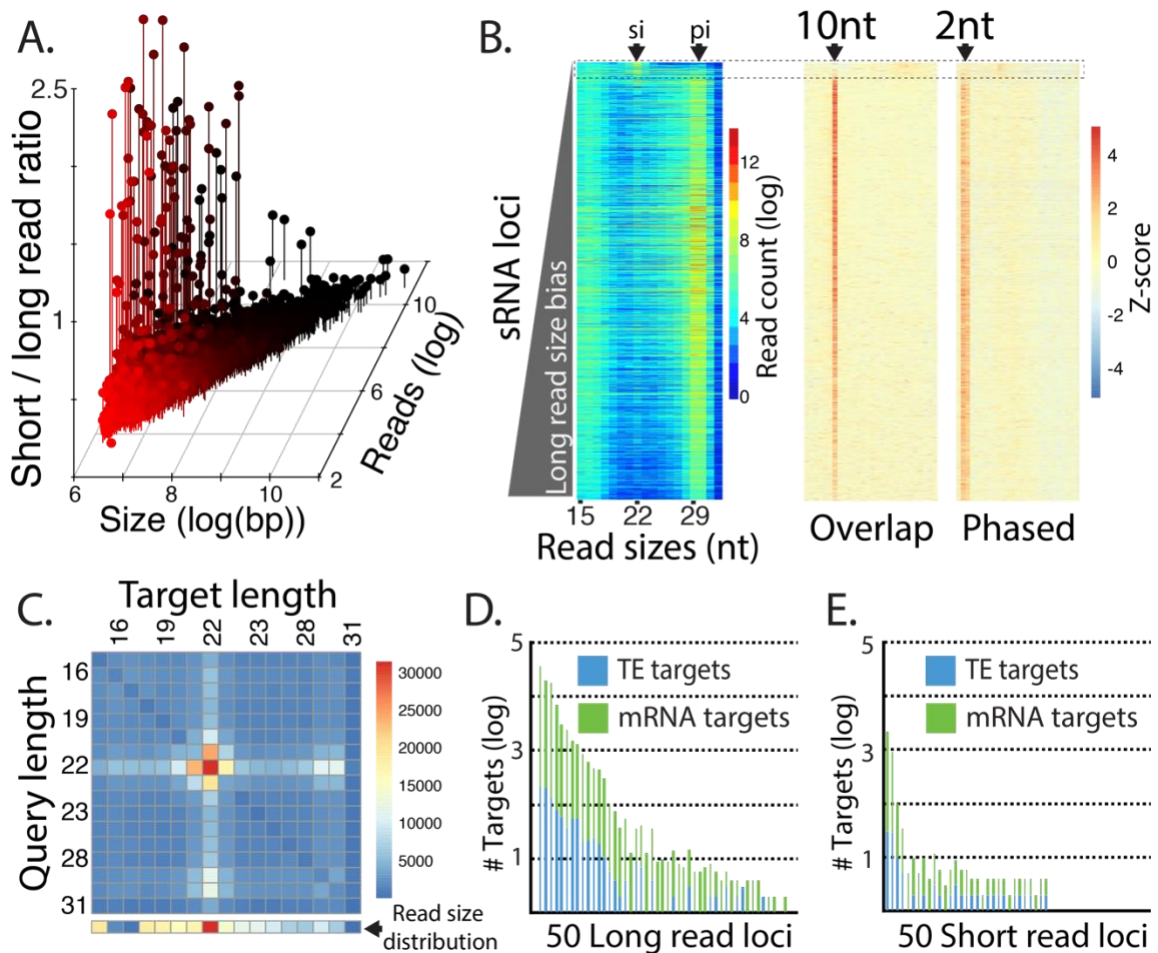


Figure 2. Whitefly small RNA expressing loci.

A) Comparison of 3878 small RNA loci in annotated in whitefly by locus size, number of mapped reads, and the ratio of short (19-23nt) to long (25-30nt) mapping reads. **B)** Analysis of small RNA sizes and piRNA biogenesis signatures for all 3878 loci. Each row of the heatmap represents a locus, which are arranged by read size bias with short read bias at the top and long at the bottom. Left panel shows size distribution. Nucleotide sizes below. Arrows at top show sizes expected to represent siRNAs (si) and piRNAs (pi). Middle panel shows read overlaps quantified by Z-score, arrow shows the 10nt overlap size. Right panel shows distance of trailing 1U read, arrow shows the 2nt proximal read distance. Dashed line box highlights the ~100 loci that do not have piRNA signatures in terms of read size, overlaps, or phasing. This group of loci have more reads at the 22nt (siRNA) size. **C)** Matrix of dicer 2nt overhang signature calculated for loci in the dashed box in panel B. Read pairs where the query read overlapped by 2 minus its total length were quantified and plotted in the heatmap. Line of boxes below the matrix show the read size distribution for reads mapping to the analyzed loci (dashed box in part B) **D-E)** Number of mRNA and transposable

element targets for the 50 most high expressing **D)** loci biased to long reads or **E)** loci biased toward short reads.

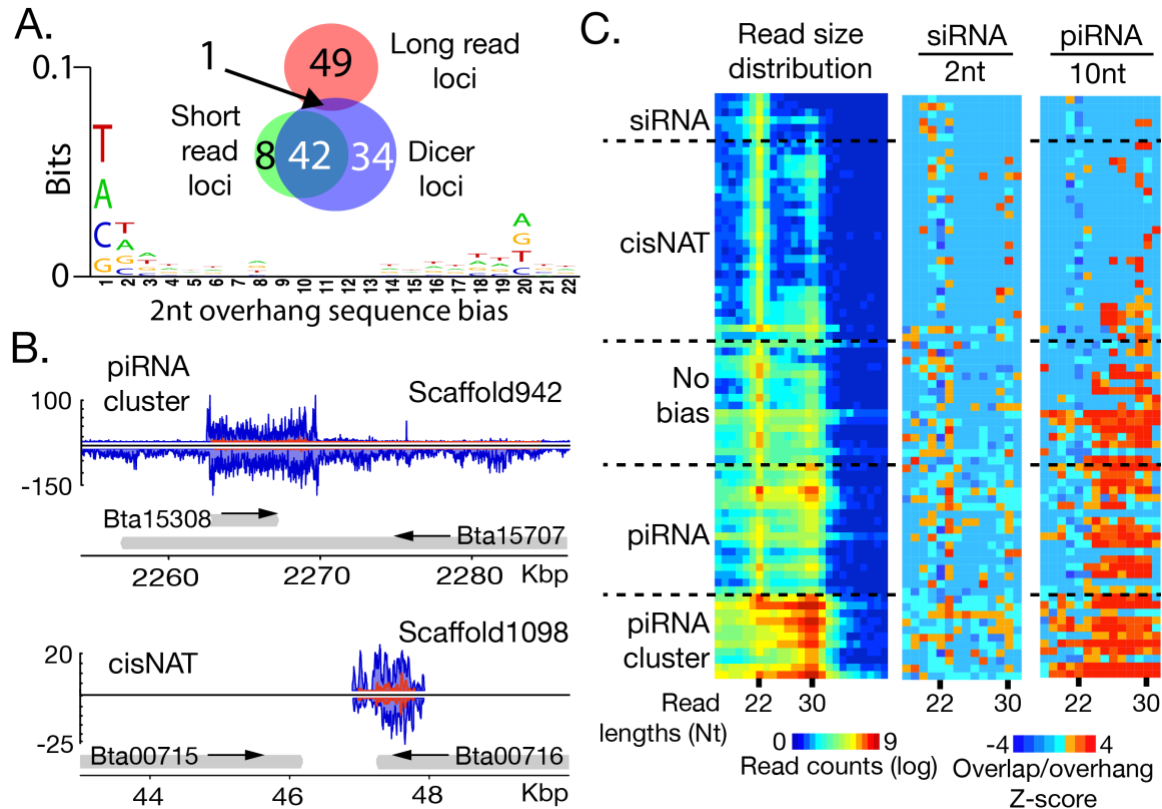


Figure 3. Identification of whitefly loci expressing small RNAs with characteristics of Dicer cleavage and ping-pong piRNAs.

A) Intersection of Dicer processing loci showing 2nt overhangs for reads sized 20-23nt with long and short read loci. The sequence biases of Dicer read loci are shown below in the seqlogo graphic **B)** Appearance of Dicer produced small RNAs (siRNAs) at sites of convergent transcription. Top panel shows expression of siRNAs in a piRNA cluster. Bottom panel is a cis-natural antisense transcript (cisNAT). Blue trace shows all reads mapping to locus. Read trace shows reads with Dicer 2nt overhang cleavage pattern. **C)** Read size distribution and biogenesis pattern of small RNAs produced at 76 Dicer signature loci. Length of reads in heatmaps indicated below. Curated identities shown on the left. Leftmost heatmap shows the distribution of reads sizes. Middle shows z-scores for 2nt overhangs (siRNAs), and right heatmaps showing z-scores for 10nt overlaps (piRNAs).

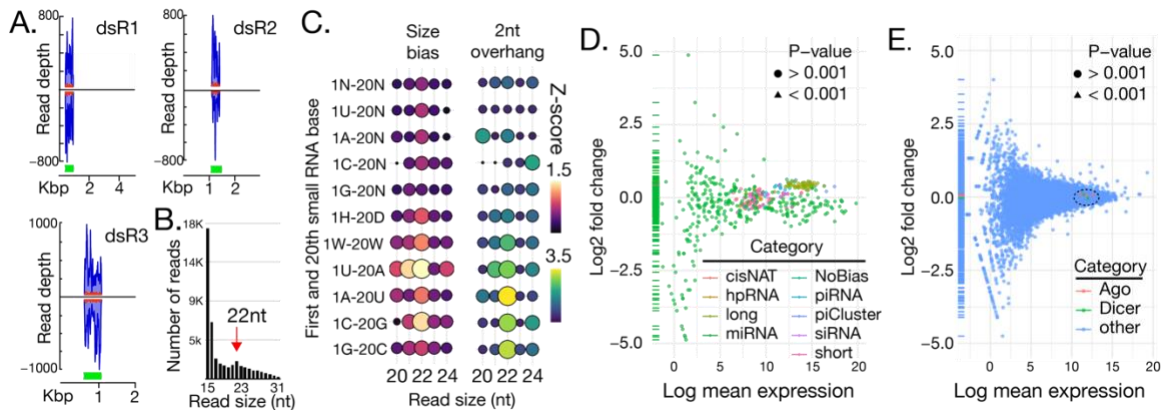


Figure 4. Metabolism of exogenous dsRNAs in whitefly.

A) Accumulation of reads mapping to dsRNA sites (green boxes) in the context of the originating transcript from potato psyllid (*B. cockerelli*). Blue trace shows all reads mapping to locus. Read trace shows reads with Dicer cleavage pattern. **B)** Size distribution of read derived from the three off-target dsRNAs (shown in A). Red arrow shows the expected size of siRNAs (22nt) **C)** Balloon plot showing characterization of sequence biases in exogenous siRNAs. Read sizes indicated below. Color and diameter of circle scale with Z-scores quantifying the reads in different sizes. On left the sequence identities of the small RNA population is indicated for first base of the read and the 20th base of the read. N = any residue, H = U/A/C, D = A/U/G, and W = A/U. The left group of circles show the abundance of reads, and the right group of circles abundance of reads with 2nt overhangs. **D-E)** Differential expression of **D)** small RNA loci and **E)** mRNAs between whiteflies treated with water or the three off-target dsRNAs. Data points colored by identity. Circles represent non-significant change in expression, triangles significant. Dashed circle shows location of Dicer and Ago proteins in scatterplot.

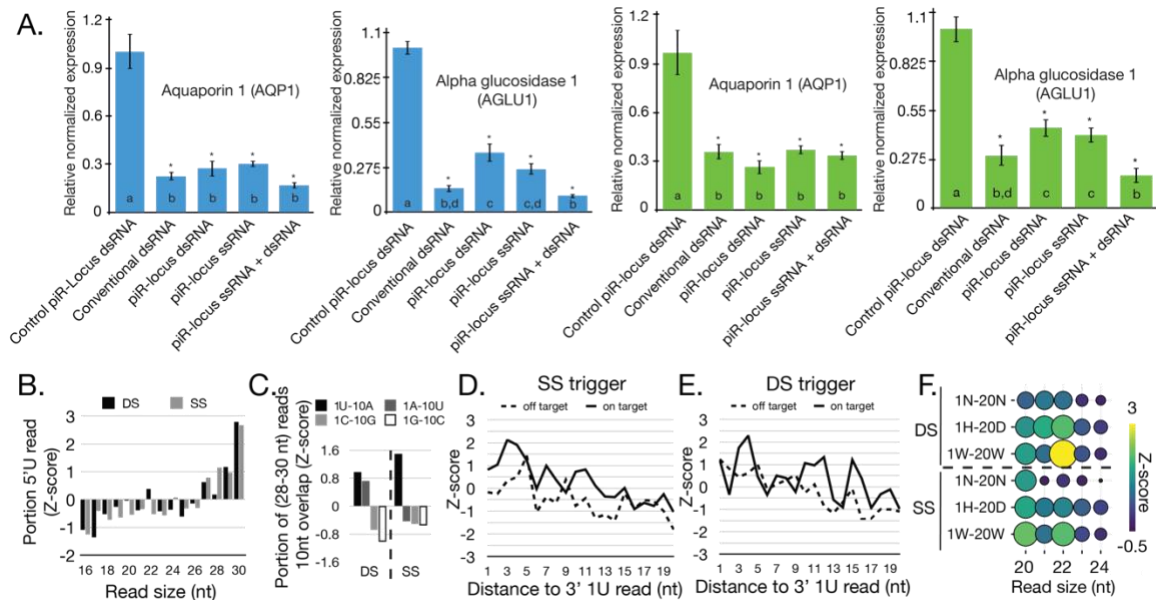


Figure 5. Exogenous piRNA mediated gene silencing in whitefly.

A) Relative expression of AQP1 and AGLU1 genes determined by qPCR after feeding with synthetic RNAs generated from piRNA triggers. Blue bar graphs are results when target gene sequences are fused to sequence from piRNA biased locus 6 (piRB-6) sequences. Green graphs are when they are fused to NoBias-14 sequences. At least three independent biological replicates were used for each type of feeding. Error bars show standard error, and letters indicate significance groups determined by Tukey HSD test. *P ≤ 0.05. **B-F)** Analysis of small RNA sequencing data from animals fed piRB-6 based piRNA triggers that map to the synthetic RNAs. **B)** Portion of small RNA sequencing reads with 1U residues shows biased to long (piRNA) sized reads. Black bars are from double-stranded (DS) triggers and grey from single-stranded (SS) versions. **C).** Enrichment of ping-pong piRNA pairs in longer sized RNAs (28-30nt) in the target gene region of the piRNA triggers. Sequence identities indicated in the legend. DS = double-stranded triggers, and SS = single-stranded triggers. 1U-10A reads, which are characteristic of bona fide ping-pong piRNAs show the greatest abundance. **D-E)** Phasing signature plots separated by off target and on target strands for **D)** single-stranded piRNA triggers and **E)** double-stranded triggers. **F)** Balloon plot showing dicer 2nt overhangs the DS (double-stranded) and SS (single-stranded) triggers. Color and size of circles scale with the abundance of 2nt overhang pairs. Left shows the sequence identities of small RNAs analyzed (N = any residue, H = U/A/C, D = A/T/G, W = A/T).

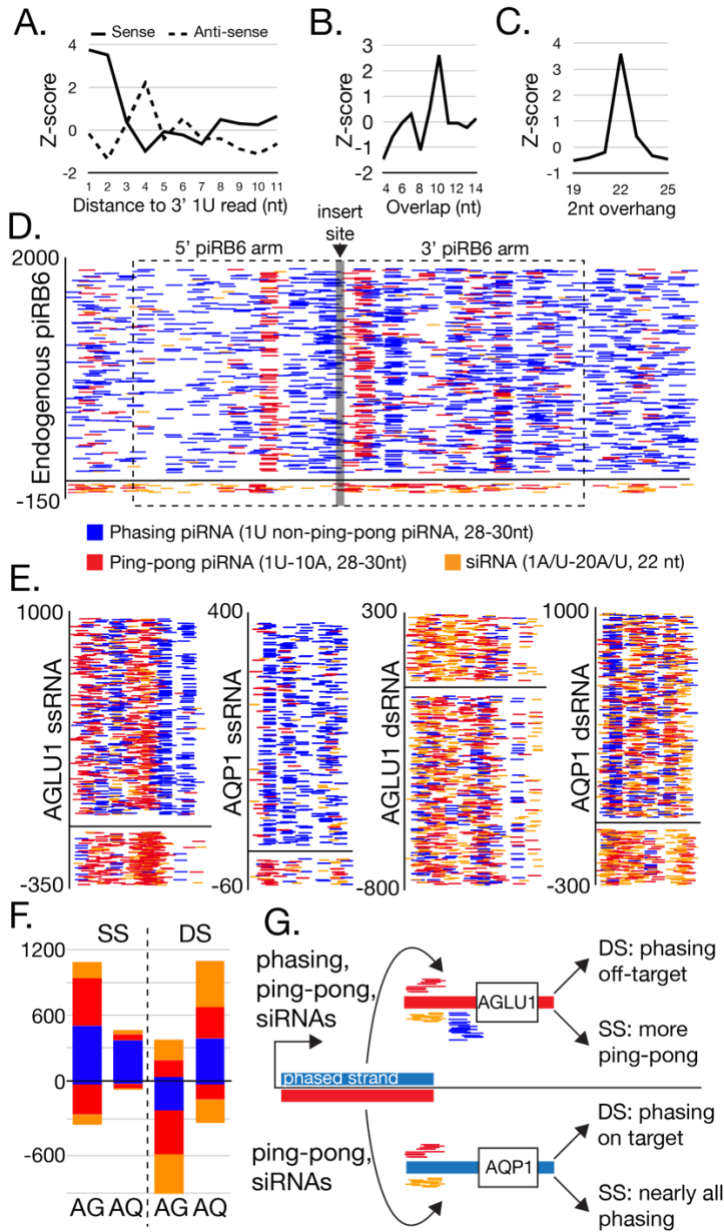


Figure 6. Processing of piRNA triggers.

A-D) Characterization of the piRB-6 locus. **A)** Phasing analysis of trailing 1U reads shows greater phasing signature on the plus strand of the locus compared to the antisense strand. **B)** Overlap analysis for the piRB-6 locus showing a peak at 10nt overlaps. **C)** Enrichment of 22nt reads that overlap by 2nt at piRB-6. **D)** Read accumulation at piRB-6. Alignments are colored by identity. Blue represents phasing piRNAs characterized by long 28-20nt 1U reads that do not overlap by 10 with antisense reads and therefore unlikely to be involved in ping-pong. Red are ping-pong piRNAs being 28-30nt reads that have 1U/10A sequences that also overlap by 10. Orange is siRNAs being 22nt reads that have

2nt overhangs with a 1U/A and 10A/U. The region cloned for the piRNA triggers indicated by dashed line box. The site of inserting target sequences show by grey line. Y-axis shows read density. **E)** Read accumulation using the color coding in part D at the sequence target region of piRNA triggers. Similarly, y-axis represents read density. **F)** Quantification of read identities by strand for plots shown in part E. Color scheme same as used in D&E. AG = AGLU1, AQ = AQP. **G)** Diagram showing the consequences of using different piRNA trigger configuration. Blue represents phasing strand of piRB-6, red the complementary. Same color scheme in D used to represent reads.

1134
1135 **Supplementary Information for**
1136

1137 **Exploiting somatic piRNAs in the whitefly, *Bemisia tabaci***
1138 **enables a new mode of gene silencing through synthetic RNA**
1139 **feeding**
1140

1141 Mosharrof Mondal¹, Judith K. Brown¹, Alex Flynt^{2,*}

1142 ¹School of Plant Sciences, University of Arizona, Tucson, Arizona, 85721, United
1143 States of America

1144 ²Cellular and Molecular Biology, University of Southern Mississippi, Hattiesburg,
1145 Mississippi, 39406, United States of America

1146
1147 *Corresponding author: Alex Flynt

1148 Email: alex.flynt@usm.edu
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159

Supplementary Text

Whitefly phusion constructs synthesis and cloning

PCR recipe for SOEing PCR

Phire Plant Direct PCR Master Mix	25 uL
Phire dilution buffer	1.5 ul
Fragment A (25ng)	
Fragment B (25ng)	
ddH ₂ O	make the reaction to 50 uL

Thermocycler steps

98°C	5 min	
98°C	7 sec	} 10 cycle
53 °C	10 sec	
72°C	7 sec	
72°C	5 min	

After the first 10 cycles, end primers were added and the thermocycler was ran for another 25 cycles following these steps:

98°C	5 min	
98°C	7 sec	} 25 cycle
63 °C (Annealing T _m)	10 sec	
72°C	15 sec	
72°C	5 min	

Phire DNA polymerase doesn't create any 'A' overhang. For T-A cloning into pGEM-T easy vector, the 'A' nt was added to the final fusion products using *Taq* DNA polymerase.

1188 **Sequences of the constructs**

1189 Capital letter sequences are piRNA trigger, sandwiched lower case sequences
1190 are from gene of interest (AQP1, AGLU1, Luciferase)

1191 **AQP1-piRB-6**

1192 AGCAGCTTCTTGCCCTCTGATTCCACGGTTTCTTCTTAAAGGGCCCCGACGACTGCTGCGGGCCTT
1193 GATAAGGCGCGCTCCTGTTATTTGCCTCACGGAACGTCTTTTCCGCGGCCATCATTGCGTCCATT
1194 GATCGGATCAAATCTTGCCCTCATTGCATCCACGGCTCGAGTATTCTATCCGTATCCGCACGATT
1195 TAGATCAACTGCGTGTACCAAAGTCGCTAGGGCGTTCTCATTGGCCTTCACCCGGGATTCTAAGG
1196 ATGATTCCCTGCCCCGTATAGTGATTACGGCCAAAATAGCGCCCCCTCCTTTGCTGGTTCGCGGCT
1197 ACTGCTAGCTTCGCATTtgcacaaatgccttggagccatctgtggagcaatcattctgaatgaaa
1198 tcacgcaaaaaacagggttacacggctgctggtaatctgggagtaacgacactgtctacaggagtt
1199 tccgacctgcaggggtgtggcgatagaagcactaatcacatttgtgctgcttttagttgtccagtc
1200 cgtctgcgatgggaagcggaccgacatcaaaggatctatcggcggttgcgataggattcgcaattg
1201 ctTCCGTCGAGTTAACTTTAGCCAAGCCCGCTAGTTTTCTCTTCGCTTGAACGTAATCCAACGGG
1202 TCCTCATTTTTCTCCCTGCGTTTCGCGCCGAGAATTTTCGTGAGGGCATCCTCGTCGCTGTCAAAGTA
1203 TTGGATCAATTTCTTCTTTACTTCTCAAAAGTCCTGCAGTTACCGAACGCTACCTCTTCATTGT
1204 CGTAGTACTGGATGGCACGTTTCGCTAAGTGATTCTGAGTTGGTCCCGCTTTTCTTGATCCGAA
1205 CATTTCTTATAGAAATTTTCAAATCTTTTAGAAATCTCTAACGTCGTAGTCAGCTTCTCCTTT
1206 GAATAGTTTTCTAAACGGCGGTGCCTTAATCGTCACCGTAGGT

1207

1208 **AQP1-No_bias-14**

1209 TTGCGTTCCTGCTCCCTTTGCCCTTACCGCGCTCAATTATCTCTATTAGAACCGGAGATATTTCG
1210 GTTTACAAAAATTTTTTGGGGCCAGCCCCCTTAATCCTTTCCCTATGGACTTCCTATATGGCC
1211 CCAGAGGTAGCCCCGGGGTTAGGCAAATAATCCCAAAAATTTCCCAAATTCTAACGGAAATGT
1212 GGCCTACCGCCCCCTACGTCACTCTGGCTATGACGTAGTTGATtgcacaaatgccttggagccat
1213 ctgtggagcaatcattctgaatgaaatcacgcaaaaaacagggttacacggctgctggtaatctgg
1214 gagtaacgacactgtctacaggagtttccgacctgcaggggtgtggcgatagaagcactaatcaca
1215 tttgtgctgcttttagttgtccagtcctgtcgcgatgggaagcggaccgacatcaaaggatctat
1216 cggcggtgcgataggattcgcaattgctTTACGTGCCGTTACACCGGTTACCGACATCAGGTTCC
1217 TTCAAATCGGACACGGGCGCCCCCTCCCCGAGGGGATGCCAATGGGGGGAGGTCCAGGCCGAAGC
1218 CTGACTTTCTACTACCTCCGGAGCTGTGCCCTTCTCTGCACGTCCAGTTGAGCACTGGTGGGCT
1219 GACCTCGGGGACAAGGTCGCCTTAACCTTACCG

1220

1221 **AGLU1-piRB-6**

1222 AGCAGCTTCTTGCCCTCTGATTCCACGGTTTCTTCTTAAAGGGCCCCGACGACTGCTGCGGGCCTT
1223 GATAAGGCGCGCTCCTGTTATTTGCCTCACGGAACGTCTTTTCCGCGGCCATCATTGCGTCCATT
1224 GATCGGATCAAATCTTGCCCTCATTGCATCCACGGCTCGAGTATTCTATCCGTATCCGCACGATT
1225 TAGATCAACTGCGTGTACCAAAGTCGCTAGGGCGTTCTCATTGGCCTTCACCCGGGATTCTAAGG
1226 ATGATTCCCTGCCCCGTATAGTGATTACGGCCAAAATAGCGCCCCCTCCTTTGCTGGTTCGCGGCT
1227 ACTGCTAGCTTCGCATTctgtccatccaaccctggattgccttttggtaatctttggcgggagag
1228 cgaccgctcacgtgcgtaataaagaagaattgaatggcatatgggctcctgggtttccctcaaa
1229 ctgataatagtccattgttctatcgagagttgtatatgcttctgtcattagtagtctttgttttcc

1230 cctctttctttgtgtagaaatcgaagacctccctgaaccttggttatcaatctatatgtattcggt
1231 tgggccatcgccgcgaccggttgtaattccagtagttcgtcgggtcgagatcgggagagcaa
1232 ttcttgggtctcgccattTCCGTCGAGTTAACTTTAGCCAAGCCCGCTAGTTTTCTCTTCGCTTGA
1233 ACGTAATCCAACGGGTCCTCATTTTCTCCCTGCGTTCGCGCCGAGAATTCGTGAGGGCATCCTC
1234 GTCGCTGTCAAAGTATTGGATCAATTTCTTCTTTACTTCCTCAAAAGTCCTGCAGTTACCGAACG
1235 CTACCTCTTCATTGTCGTAGTACTGGATGGCACGTTTCGCTAAGTGATTCTGAGTTGGTCCCGC
1236 TTTTCTTGATCCGAACATTTCTTATAGAAATTTTCAAAATCTTTTAGAAATTCTCTAACGTCGTA
1237 GTCAGCTTCTCCTTTGAATAGTTTTCTAAACGGCGGTGCCTTAATCGTCACCGTAGGT

1238

1239 **AGLU1-No_bias-14**

1240 TTGCGTTCCTGCTCCCTTTGCCCTTTACCGCGCTCAATTATCTCTATTAGAACCGGAGATATTCG
1241 GTTTACAAAAATTTTTTGGGGCCAGCCCCCTTAATCCTTTCCCTATGGACTTCCTATATGGCC
1242 CCAGAGGTAGCCCCGGGGTTAGGCAAATAATCCCAAAAAATTCCAAATTCTAACGGAAATGT
1243 GGCACCTACGCCCCCTACGTCACCTCTGGCTATGACGTAGTTGATctgtccatccaaccctggattg
1244 ccttttggtaatctttggcgaggagacgaccgctcacgtgcgtaataaagaagaaattgaatggc
1245 atatgggctcctgggttccccctcaaactgataatagtcattgttctatcgagagttgtatatgc
1246 ttctgtcattagtagtacttttgttttccccctctttctttgtgtagaaatcgaagacctccctgaacc
1247 ttgttatcaatctatatgtattcggttggtccatcgccgcgaccggttgtaattccagtagttc
1248 gtccgggtcgagatcgggagagagcaattcttgggtctcgccattTTACGTGCCGTTACACCGGTTA
1249 CCGACATCAGGTTTCCTTCAAATCGGACACGGGCGCCCCCTCCCCGAGGGGATGCCAATGGGGGGAG
1250 GTCCCAGGCCGAAGCCTGACTTTCTACTACCTCCGGAGCTGTGCCCTTCTCTGCACGTCCCAGTT
1251 GAGCACTGGTGGGCTGACCTCGGGGACAAGGTCGCCTTAACCTACCG

1252

1253 **Luciferase-piRB-6**

1254 AGCAGCTTCTTGCTCTGATTCCACGGTTTCTTCTTAAAGGGCCCCGACGACTGCTGCGGGCCTT
1255 GATAAGGCGCGCTCCTGTTATTTGCCTCACGGAACGTCTTTTCCGCGGCCATCATTGCGTCCATT
1256 GATCGGATCAAATCTTGCTCATTGCATCCACGGCTCGAGTATTCTATCCGTATCCGCACGATT
1257 TAGATCAACTGCGTGACCAAAGTCGCTAGGGCGTTCTCATTGGCCTTCACCCGGGATTCTAAGG
1258 ATGATTCTGCCCCGTATAGTGATTTACGGCCAAAATAGCGCCCCCTTCCTTTGCTGGTCGCGGCT
1259 ACTGCTAGCTTCGCATTTtcgtgccagagtctttcgacagggacaaaaccattgccctgatcatg
1260 aacagctctgggtctaccggcctgcctaagggcggtggcctgcctcatcgcaccgcctgtgtgcg
1261 cttctctcacgcccgcgaccctattttcggcaaccagatcatccccgacacgcctattctgagcg
1262 tggtgccattccaccacggcttcggcatgttcaccacctgggctacctgattTCCGTCGAGTTA
1263 ACTTTAGCCAAGCCCGCTAGTTTTCTCTTCGCTTGAACGTAATCCAACGGGTCTCATTTTCTCC
1264 CTGCGTTCGCGCCGAGAATTCGTGAGGGCATCCTCGTCGCTGTCAAAGTATTGGATCAATTTCT
1265 TCTTTACTTCCTCAAAAGTCCTGCAGTTACCGAACGCTACCTCTTCATTGTCGTAGTACTGGATG
1266 GCACGTTTCGCTAAGTGATTTCTGAGTTGGTCCCGCTTTTCTTGATCCGAACATTTCTTATAGAA
1267 ATTTTCAAAATCTTTTAGAAATTCTCTAACGTCGTAGTCAGCTTCTCCTTTGAATAGTTTTCTAA
1268 ACGGCGGTGCCTTAATCGTCACCGTAGGT

1269

1270 **Luciferase-No_bias-14**

1271 TTGCGTTCCTGCTCCCTTTGCCCTTTACCGCGCTCAATTATCTCTATTAGAACCGGAGATATTCG
1272 GTTTACAAAAATTTTTTGGGGCCAGCCCCCTTAATCCTTTCCCTATGGACTTCCTATATGGCC

1273 CCAGAGGTAGCCCCGGGGGTAGGCAAATAATCCCAAAAATTCCCAAATTCTAACGGAAATGT
1274 GGCCTACCGCCCCCTACGTCACTCTGGCTATGACGTAGTTGATttcgtgccagagtctttcgaca
1275 gggacaaaaccattgccctgatcatgaacagctctgggtctaccggcctgcctaagggcgtggcc
1276 ctgctcatcgcaccgcctgtgtgcgcttctctcacgcccgcgaccctattttcggcaaccagat
1277 catccccgacaccgctattctgagcgtggtgccattccaccacggcttcggcatgttcaccacc
1278 tgggctacctgattTTACGTGCCGTTACACCGGTTACCGACATCAGGTTCTTCAAATCGGACAC
1279 GGGCGCCCCCTCCCCGAGGGGATGCCAATGGGGGGAGGTCCCAGGCCGAAGCCTGACTTTCTACTA
1280 CCTCCGGAGCTGTGCCCTTCTCTGCACGTCCCAGTTGAGCACTGGTGGGCTGACCTCGGGGACAA
1281 GGTGCGCTTAACTTACCG

1282

1283 **AQP1 sequence used in this study to synthesize dsRNA (from accession #**
1284 **KF377800.1)**

1285 tcgcacaatgccttgagccatctgtggagcaatcattctgaatgaaatcacgcaaaaaacaggt
1286 tacacggctgctggtaatctgggagtaacgacactgtctacaggagtctccgacctgcagggtgt
1287 ggcgatagaagcactaatcacatttgtgctgcttttagttgtccagtcctgtcgatgggaagc
1288 ggaccgacatcaaaggatctatcggcgttgcataggttcgcaattgct

1289

1290 **AGLU1 sequence used in this study to synthesize dsRNA (from accession #**
1291 **KF377803.1)**

1292 ctgtccatccaaccctggattgccttttggtaatctttggcgggagagcgaccgctcacgtgcgt
1293 aataaagaagaaattgaatggcataatgggctcctggtttccctcaaactgataatagtccattg
1294 ttctatcgagagttgtatatgcttctgtcatttagtacttttgttttccctctttctttgtgtag
1295 aaatcgaagacctccctgaaccttggtatcaatctatatgtattcggttggtccatcgctccgcga
1296 ccggttgtaattccagtagttcgtcgggtcgagatcgggcgagagcaattcttggtctcgccatt

1297

1298 **Luciferase sequence was cloned from psiCHECK™-2 plasmid (Promega, catalog #**
1299 **C8021)**

1300 ttcgtgccagagtctttcgacagggacaaaaccattgccctgatcatgaacagctctgggtctac
1301 cggcctgcctaagggcgtggccctgcctcatcgcaccgcctgtgtgcgcttctctcacgcccgcg
1302 accctattttcggcaaccagatcatccccgacaccgctattctgagcgtggtgccattccaccac
1303 ggcttcggcatgttcaccaccctgggctacctgatt

1304

1305

1306

1307

1308

1309

1310 **Primer sequences used in this study**

1311 (Underlined regions are overlap to the genes of interest)

1312 **piB6_A F**

1313 AGCAGCTTCTTGCCCTGATTCCAC

1314 **piB6_B R**

1315 AATGCGAAGCTAGCAGTAGCCGC

1316 **piB6_C F**

1317 TCCGTCGAGTTAACTTTAGCCAAGCC

1318 **piB6_D R**

1319 ACCTACGGTGACGATTAAGGCACC

1320 **Eq14_A F**

1321 TTGCGTTCCTGCTCCCTTTGCC

1322 **Eq14_B R**

1323 ATCAACTACGTCATAGCCAGAGTGACG

1324 **Eq14_C F**

1325 TTACGTGCCGTTACACCGGTTACC

1326 **Eq14_D R**

1327 CGGTAAGTTAAGGCGACCTTGTC

1328 **AQP1-piB6 F**

1329 GCTGGTCGCGCTACTGCTAGCTTCGCATTTTCGCACAATGCCTTGGAGCCATC

1330 **AQP1-piB6 R**

1331 AGCGGGCTTGGCTAAAGTTAACTCGACGGAAGCAATTGCGAATCCTATCGCAACG

1332 **AQP1-Eq14 F**

1333 CTACGTCACCTCTGGCTATGACGTAGTTGATTCGCACAATGCCTTGGAGCCATC

1334 **AQP1-Eq14 R**

1335 GATGTCGGTAACCGGTGTAACGGCACGTAAAGCAATTGCGAATCCTATCGCAACG

1336 **AGLU1-piB6 F**

1337 GCTGGTCGCGGCTACTGCTAGCTTCGCATTCTGTCCATCCAACCCTGGATTGCC
 1338 **AGLU1-piB6 R**
 1339 AGCGGGCTTGGCTAAAGTTAACTCGACGGAAATGGCGAGACCAAGAATTGCTCTCG
 1340 **AGLU1-Eq14 F**
 1341 CTACGTCACTCTGGCTATGACGTAGTTGATCTGTCCATCCAACCCTGGATTGCC
 1342 **AGLU1-Eq14 R**
 1343 GATGTCGGTAACCGGTGTAACGGCACGTAAATGGCGAGACCAAGAATTGCTCTCG
 1344 **AGLU1_dsRNA F**
 1345 CTGTCCATCCAACCCTGGATTGCC
 1346 **AGLU1_dsRNA R**
 1347 AATGGCGAGACCAAGAATTGCTCTCG
 1348 **Aqp1_dsRNA F**
 1349 TCGCACAATGCCTTGGAGCCATC
 1350 **Aqp1_dsRNA R**
 1351 AGCAATTGCGAATCCTATCGCAACG
 1352
 1353
 1354
 1355
 1356
 1357
 1358

Supplementary Figures

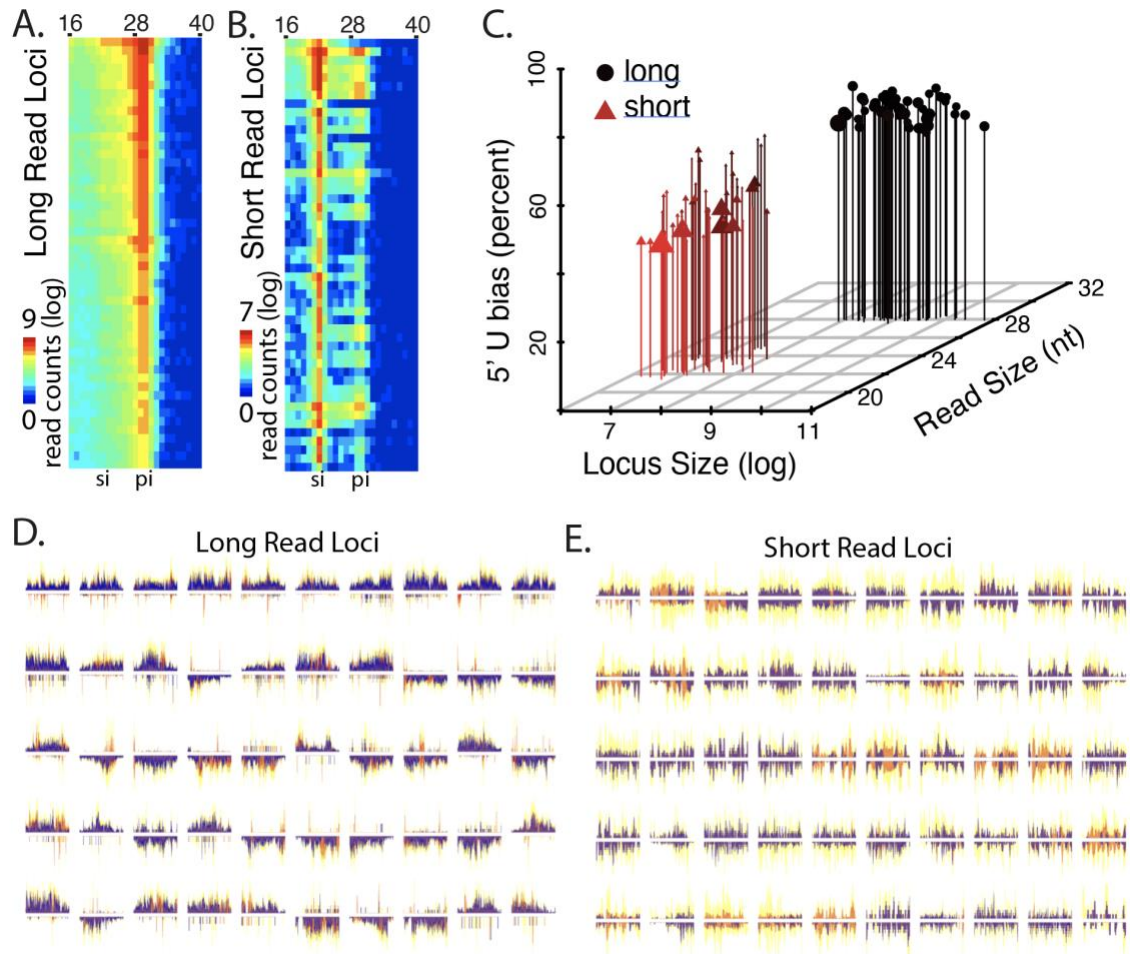
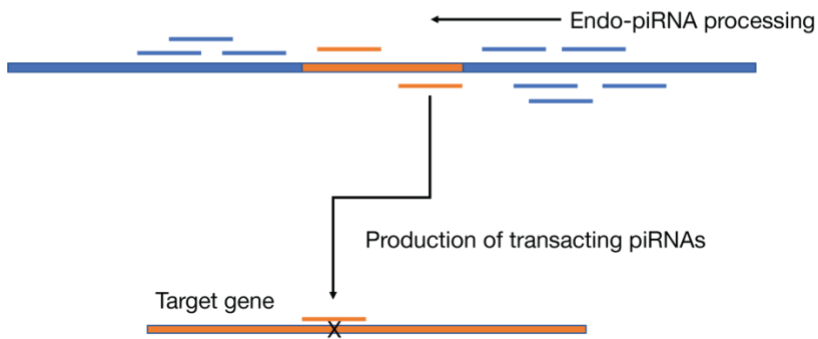


Figure S1. Characterization of the 50 most highly expressed small read biases and long read biased loci.

A) 50 most high expressing long read (25-30nt) biased-loci. Heatmap shows distribution of reads between 16nt and 40nt. **B)** 50 most high expressing short read (19-23nt). si = siRNA sized, pi = piRNA sized. biased-loci Heatmap shows distribution of reads between 16nt and 40nt Yellow shows density of all reads. **C)** 3D scatterplot for the loci in A&B assessed by locus length, expression, and 1U bias. **D)** 50 top long (25-31nt) read loci where orange shows multi-mapping long reads, and violet uniquely-mapping long reads. **E)** 50 top short (19-23nt) read loci where orange shows multi-mapping short reads, and violet uniquely-mapping short reads.

A. Proposed piRNA trigger



B. Proposed piRNA/siRNA co-triggers

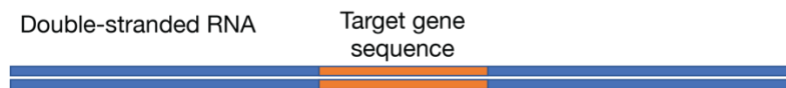
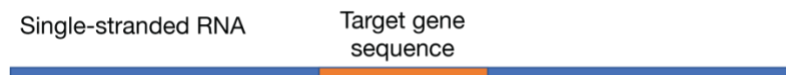
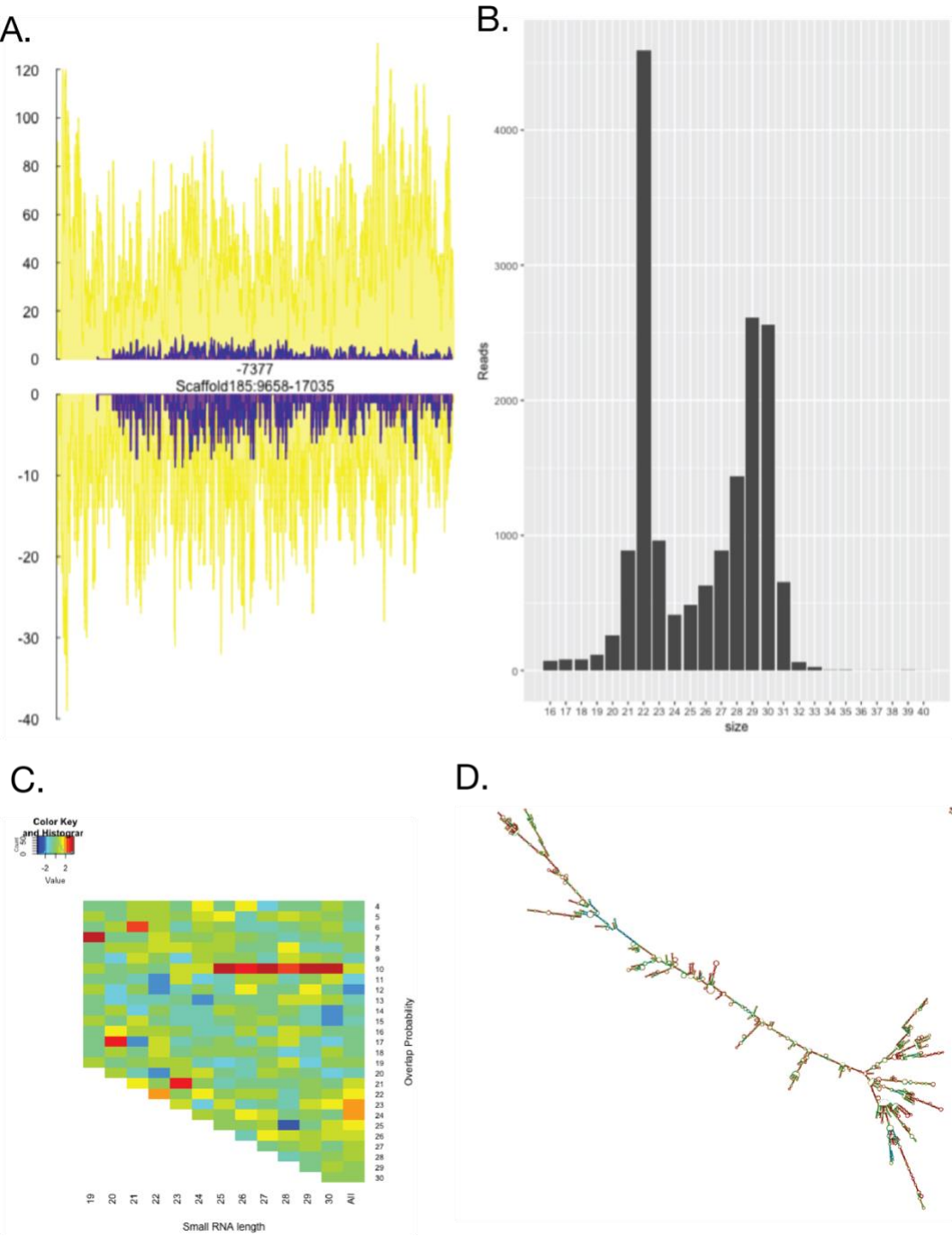


Figure S2. Proposed mechanism for piRNA/siRNA trigger

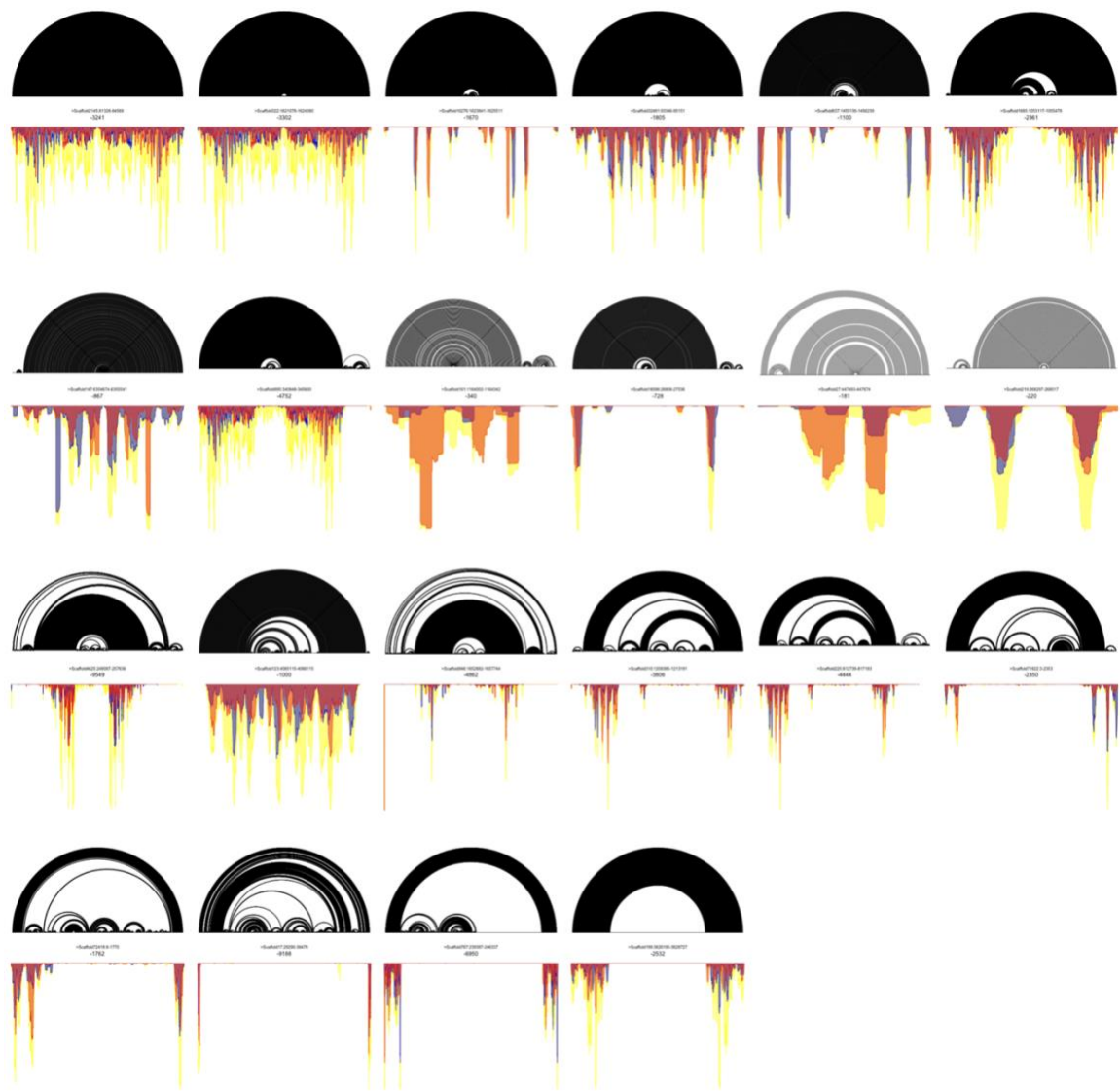


1378

1379 **Figure S3. Closer inspection of piRB-6 locus.**

1380 A) Read density of the mapped reads. Yellow bars show all read mapping, violet
1381 bars represent reads that mapped uniquely to the locus. B) Size distribution of
1382 the reads mapped to the locus. C) Overlap probability z-score of the mapped
1383 reads. D) Local secondary structures of the entire locus

1384



1385

1386

1387 **Figure S4. Visualization of RNA structure and small RNA expression at**
1388 **curated hpRNA loci.**

1389 Top part of each panel depicts RNA structure with lines connecting one or more
1390 bases indicating pairing. Bottom panel is a density plot showing relative read
1391 depth across the locus. Red color indicates accumulation of 20-23nt reads that
1392 map to more than one position in the genome. Blue indicates 20-23nt reads that
1393 map uniquely to the locus. Yellow shows density of all read sizes.

1394

1395 **Supplementary Datasets:**

1396

1397 **S1 Table (separate data file):** Annotated small RNA loci in this study (miRNA,
1398 cisNTAs, hpRNA, etc.)

1399 **S2 Table (separate data file):** All small RNA loci identified in this study (total
1400 3873 loci)

1401 **S3 Table (separate data file):** Differential expression of the mRNA and small
1402 RNA loci (dsRNA fed vs control)

1403