

The small RNA universe of *Capitella teleta*

Sweta Khanal, Beatriz Zancanela, Jacob Oche Peter, and Alex Sutton Flynt*

Cellular and Molecular Biology

University of Southern Mississippi

Hattiesburg, MS

*Corresponding author

RNAi is an evolutionarily fluid mechanism with dramatically different activities across animal phyla. One major group where there has been little investigation is annelid worms. Here the small RNAs of the polychaete developmental model, *Capitella teleta*, are profiled across development. As is seen with nearly all animals, nearly 200 hundred microRNAs were found with 58 high confidence novel species. Greater miRNA diversity was associated with later stages consistent with differentiation of tissues. Outside miRNA, a distinct composition of other small RNA pathways was found. Unlike many invertebrates, an endogenous siRNA pathway was not observed, indicating pathway loss relative to basal planarians. No processively generated siRNA-class RNAs could be found arising from dsRNA precursors. This has significant impact on RNAi technology development for this group of animals. Unlike the apparent absence of siRNAs, a significant population of piRNAs were observed. For many piRNAs phasing and ping pong biogenesis pathways were identified. Interestingly, piRNAs were found to be highly expressed during early development, suggesting a potential role in regulation in metamorphosis. Critically, the configuration of RNAi factors in *C. teleta* is found in other annelids and mollusks, suggesting that similar biology is likely present in the wider clade. This study is the first providing comprehensive analysis of small RNAs in annelids.

INTRODUCTION

RNA interference (RNAi) is a widely applied genetic technology based on fundamental gene regulatory mechanisms where small RNAs induce complementary transcript degradation or destruction. There are also numerous examples of small RNAs directing epigenetic alterations. However, behavior of RNAi pathways vary from species to species making application of a single paradigm inappropriate (1). Among invertebrate animals, RNAi pathways appear to be especially plastic, indicating that at a minimum an order—if not family-level investigation of biogenesis mechanisms is necessary for effective development of gene silencing technology.

Animal small RNAs belong to three classes: microRNAs (miRNAs), piwi-associated RNAs (piRNAs), and small-interfering RNAs (siRNAs) (2). Among the three major classes of RNAi, miRNAs are the most conserved with sequences shared in nearly all animals (3). In comparison, both endogenously expressed siRNAs (endo-siRNAs) and piRNAs show almost no conservation even at the species level, likely due to their role in silencing invasive nucleic acids (4). In addition to these conservation patterns, each class is defined by loading into a distinct class of Argonaute/Piwi (Ago/Piwi) protein. Work from arthropods where each class is well-represented provides definitions where Ago1 loads miRNAs (miAgo), Ago2 siRNAs (siAgo), and Piwi piRNAs. Arthropod piwi proteins include the genes Aubergine (Aub) and Ago3. In contrast, even though vertebrate genomes encode multiple Agos, usually four, they do not have dedicated siAgos (5). Thus, RNAi/siRNA technology in vertebrates is based on miRNA mimicry and is distinct from siRNAs used for RNAi in flies and nematodes (6). Vertebrates do possess numerous Piwi proteins like arthropods. This highlights the lability of distinct siRNA pathways.

Indeed, analysis of selection in different *Drosophilid* pathway components shows miAgos are the most stable, followed by Piwi's, with siAgos being the most rapidly evolving (7).

In addition to association with distinct effectors, RNAi pathways are also defined by biogenesis. miRNA are derived from short hairpins that are initially cropped from heterogeneous transcripts by the RNase III enzyme Drosha followed by export and final maturation by a second RNase III, Dicer (8). miRNA hairpin precursors have features that include asymmetric bulges larger terminal loops of ~10nt depending on the species. Pre-miRNA stem sequences are deeply conserved, while the terminal loop sequence is less so (9). siRNAs are also produced by Dicer, are typically 20-24 nt, but are instead derived from long double-stranded RNA (dsRNA) molecules (10). Various sources yield siRNAs, that include viruses along with many endogenous species such as hairpin RNAs, cis-NATs, and transposable elements (11).

piRNA biogenesis occurs from two distinct processes phasing (phasi-piRNAs) and Ping-Pong cycle (pingpong-piRNAs), both are created independent of RNase III activity with processing driven by the "slicer" RNase activity intrinsic to Piwis (12). A consequence of this is a different size range (25-32nt) compared to miRNAs and siRNAs (20-24nt). phasi-piRNAs are produced from single-stranded RNAs (ssRNAs) that are initially cleaved by a Piwi protein, which leads to further processing by the endonuclease Zucchini (13, 14). Phasi-piRNAs are defined by 1U bias and close proximity (1-3nt) between the 3' end of an upstream piRNAs, and the 5' end of a downstream piRNA. Pingpong-piRNAs are created in an amplifying loop where partner Piwi proteins, Aub and Ago3 in *Drosophila*, slice transcripts that subsequently load into partners becoming new piRNAs (15). As pingpong-piRNAs are generated by slicing of complementary transcripts they can be identified by 10nt overhangs between piRNAs. Phasi-

68 piRNA and pingpong-piRNA biogenesis collaborate to generate piRNAs to suppress the
69 expression of unlicensed transcripts.

70 The paradigm for piRNA function is suppression of transposable elements (TEs) both in
71 arthropods and vertebrates. In these species large piRNA clusters serve as repositories of
72 forbidden elements (16). piRNA clusters are found both in uni-strand and dual-strand
73 arrangements that yield phasi-piRNAs that subsequently participate in the ping pong cycle (16).
74 There are also genic piRNAs produced from UTRs by phasing biogenesis, which are likewise
75 found in vertebrates and invertebrates (17). Initially, piRNAs were viewed as exclusive to
76 germ lines, however, for many invertebrates piRNAs are also present in soma where they not
77 only appear to suppress TEs, but also participate in gene regulatory networks (18). In contrast, *C.*
78 *elegans* piRNAs have a completely distinct biogenesis with each piRNA being produced from
79 short autonomous transcriptional units, defined by a specific motif ~ 50nt upstream of piRNA
80 transcriptional start site (19).

81 RNAi pathways are well-documented in model organisms such as *Drosophila*
82 *melanogaster*, and *Caenorhabditis elegans*, as well as vertebrates like mice and humans. In
83 comparison, there has been substantially less investigation of RNAi in spiralian, one of three
84 animal superphyla that includes flatworms, annelids, and mollusks; which branched ~700-850
85 million years from ecdysozoa. Research has mostly been in planarians which found all three small
86 RNA classes with a pronounced expansion of piRNAs (20). Other studies have investigated small
87 RNAs in mollusks that likewise noted substantial expansion of piRNAs (21). These leaves one
88 major spiralian group, annelids, where there has been no comprehensive investigation of small
89 RNA classes. Here we describe small RNA expression and biogenesis in the annelid
90 developmental model, *Capitella teleta*. From these efforts we find numerous novel miRNAs, and

as seen with other spiralian a substantial collection of somatic piRNAs. Interestingly, we do not find endogenous siRNAs suggesting that following the split from planarians, lophotrochozoans (annelids and mollusks) lost a distinct siRNA pathway. This greatly impacts RNAi approaches in these animals, informing gene silencing approaches that would be beneficial for manipulating a variety of economically significant organisms.

RESULTS

***C. teleta* Global Small RNA populations**

We began assessment of *C. teleta* RNAi pathways by examining this annelid's Ago proteins (Fig 1A). Surprisingly, only three Ago/Piwi proteins are encoded in the genome. To predict likely functions, homology to select Ago/Piwi proteins from *Drosophila melanogaster*, *Caenorhabditis elegans*, *Homo sapiens*, and *Schmidtea mediterranea* was assessed. In this analysis one *C. teleta* Ago clustered with miAgos while the other two fell in the piwi clade. *C. teleta* Piwi1 and Piwi2 group with *D. melanogaster* ping pong partners Ago3 and Aub, respectively, suggesting this biogenesis pathway is active. Missing from *C. teleta* was a siAgo, which are present in basal planarians. The absence of siAgo is correlated with *C. teleta* only having a single Dicer protein that is related to miRNA Dicers. In contrast, *S. mediterranea* possesses two, one of which groups with miRNA and other with siRNA Dicer (Sup Fig 1).

To understand the universe of small RNAs coordinated by these factors small RNA sequencing was performed on RNAs extracted from early embryo (1-4 cell), late embryo (mid gastrula), larvae, and adult anterior (mouth-segment 12) and posteriors (segment 12-anus) separately for males and females. *C. teleta* ovaries are in posterior segments while testes are anterior, thus the four adult libraries represent separately male and female gonad as well as soma

only anterior and posterior. In total, 187M small reads were acquired across the libraries, for which 84% mapped to the *C.teleta* genome excluded highly redundant sequences (Table 1). For initial assessment of small RNAs species, all libraries were combined followed by mapping of 15-32nt reads and application of a small RNA loci finding strategy based on read sites with coverage ≥ 200 and merging of features within 500bp. From this >20,000 loci expressing small RNAs were identified (Sup file 1). The size distribution of mapping reads to all loci showed two peaks, one at 22-23nt and another at 29-31 (Fig 1B). Based on observations from other invertebrates these two peaks likely represent Ago-loaded Dicer products (miRNAs/siRNAs) and piRNAs, respectively. Next size distribution per locus was calculated, which allowed segregation into 9 clusters (Fig 1B). A super majority (~90%) of loci represented by 6 of the 9 clusters showed mapping of 29-31 nt reads suggesting piRNA-producing regions are the most common small RNA loci in *C. teleta*. The remaining three loci seem to either represent miRNAs or the results of degradation (debris). Intersection of known miRNAs from public annotations found that ~70% were in cluster 2 and ~30% in cluster 4. A single known miRNA (bantam) was found in cluster 1.

When comparing the total alignments from each cluster the most significant were the two largest apparent piRNA representing clusters and the two containing most known miRNAs. The others that showed heterogenous sizes (clusters: 3, 7, 8, 9) or mapping of small, ≤ 20 nt sizes (cluster 1) comprised ~ 6% of all alignments. Next, we applied seqlogo analysis to further characterize small RNAs represented in each group (Sup Fig 2). Cluster one was predominately represented by the sequence of the miRNA bantam—the one miRNA found in this group. The other miRNA clusters had little bias, which 1T bias was seen for several putative piRNA groupings, such as cluster #4. Together these results show the expected miRNA but also abundant piRNA populations in *C. teleta* the datasets.

***C. teleta* miRNAs**

To further examine *C. teleta* small RNA classes we first sought to annotate miRNAs using the miRdeep2 algorithm (Sup File 2-7) (22). ~700 potential miRNAs were found by the method based on alignments from combined small RNA sequencing guided by existing annotations from *C. teleta*, *Eisenia fetida*, *Crassostrea gigas*, and *Lottia gigantea* (23-25). Of the 102 annotation in MirGeneDB, 99 were found (Sup File 3). The three missing miRNAs (miR-2-o36, miR-2690, and miR-33) were represented by reads in the combined sequencing libraries, and were likely overlooked by the algorithm due to duplicates in the case of miR-2 or other issues identifying hairpin folds (Sup File 4). Nevertheless, this shows the depth of sequencing described in this study is sufficient to uncover known *C. teleta* small RNAs. Alongside confirming known miRNAs, 19 “homology rescued” species were found either as novel duplicates of *C. teleta* miRNAs or having similarity to known miRNAs from *Eisenia fetida*, *Crassostrea gigas*, and *Lottia gigantea* (Sup File 5). In some cases, this homology rescued annotations appear to result from near identical duplication of *C. teleta* miRNAs or to have arisen earlier in the annelid lineage (Sup File 6).

Through this analysis we were able to annotate novel miRNAs, which were individually vetted based on expression, presence of Drosha and Dicer cleavage signature, and precision of 5' end processing (Fig 2A) (Sup File 6-7) (26). Out of 569 predicted novel miRNAs, 58 showed all features and were categorized as confident (Sup File 6). 117 were denoted as candidates due to suboptimal features that were not indicative of RNase III processing (Sup File 7). The last group of 390 was labeled as false positives and excluded from further analysis. Significant duplication was seen in novel *C. teleta* miRNAs with only 32 unique confident and 91 unique candidate species identified. Taking redundancy of known and homology rescued miRNAs into consideration, there

are 99 confident unique miRNA species and 190 if candidate miRNAs are included. Additional sequencing may promote the identity of candidate miRNAs to confident status and uncover additional miRNAs. Based on efforts in *Drosophila*, saturating sequencing depth for recovery of all miRNAs in the 333Mb genome of *C. teleta* would be achieved with ~600 million reads (26). This study only provides 30% of the requisite depth to exhaustively annotate miRNAs-particularly low abundance non-canonical species.

Examining the expression pattern of known miRNAs across development showed a general trend where a greater collection of known miRNAs increases over time comparing embryonic stages to larval and adult (Fig 2B). This has been observed in other species and reflects the greater cell type diversity that arises during differentiation (27). Greater known miRNA diversity is found in *C. teleta* anterior regions also likely due to the greater number of cell types found in tissues like brain and pharynx. Minor differences are seen between genders for both posterior and anterior. Expression of miRNAs is strikingly similar between male and female body, suggesting only a negligible role in ovary development. An opposite arrangement was seen with novel miRNAs. Here a substantial fraction was more abundant during post-zygotic through larval stages. They were not observed in the gravid female body sample suggesting they are either produced from embryonic transcription or as maternally deposited intact precursors. The discovery of these miRNAs is likely due to sampling of embryonic stages, which was not done for other annelids or mollusks. Considering their prolonged expression, these miRNAs may have a role in suppressing metamorphosis. Consistent with this, GO analysis of targets predicted by the targetscan algorithm found enrichment for response to oxygen, nitrogen, peptides, hormones, and insulin signaling (Sup Fig 3A) (28). Ingestion does not begin until swimming larva settle and metamorphose, therefore at early stages animals may have a modulated insulin signaling pathway.

Several of these early staged novel miRNAs were noted as having nearly identical expression during early embryogenesis (Fig 2D). Upon further inspection many highly abundant miRNAs were found to be encoded on a single scaffold (CAPTEscaffold_488). Many of these miRNAs appear to be tandem duplications reminiscent of the miR-430 cluster in zebrafish that eliminates maternal messages and the various miRNA clusters found expressed in *Drosophila* testis (29, 30). Targets of these miRNAs are enriched for cytoskeletal and histone regulators, which may be involved in the morphological and gene expression changes associated with metamorphosis (Sup Fig 3B). It will be intriguing if this miRNA cluster arrangement is present in the genome of other metamorphosing annelids. In addition to this highly expressed early-stage cluster, four other major clusters were annotated on CAPTEscaffold_6, CAPTEscaffold_324 and CAPTEscaffold_60 (Sup Fig 4).

During our curation of novel *C. teleta* miRNAs we noticed an abundance of unusual precursors. In the predicted miRNAs there were many species where the stem loop exhibited no bulges, and the loop was minimal with either a three or four base loop (Fig 2E). In the *C. teleta* genome we observed 13 species with a four-base loop and 6 with a three-base loop. Both were reminiscent of miR-451, which is processed in a Dicer independent manner (31, 32). However, for these miRNAs this does not appear to be the case. They exhibited strong evidence of dicer cleavage at their loops with reads aligning with precise 3'nt overhangs. To confirm the abundance of these hairpins was, in fact, unusual we assessed the abundance of similar precursor miRNAs reported for all species in miRbase (23). Out of 271 organisms only 28 had any three base loop miRNAs while only 65 had four base loops. For both configurations, the only species that had more of these tight loop hairpins was the Nile tilapia, *Oreochromis niloticus*, which appears to have an unusual miRNA biology (33). The next closest was humans, likely the most deeply sequenced species with

nearly 2000 reported miRNAs. In comparison with the 302 potential miRNAs (confident, known, homology-recovered, and candidate) we report for *C. teleta* we recover even more than the perhaps over-aggressively annotated human genome.

***C. teleta* piRNAs**

Even more abundant than miRNAs in datasets, were apparent piRNAs in *C. teleta*, both in terms of number of putative piRNA loci and abundance of piRNA molecules (Fig 1B). To analyze all possible piRNAs in *C. teleta* we identified all alignments of 15-31 nt reads that exhibit 10 nt overlaps, which is indicative of ping pong processing as well as trigger piRNA-mediated cleavage that initiate phasi-piRNAs (34). This read mapping profile was overlaid with all loci with 15-31 nt aligning reads that were over 1000 bp long. From this, 976 loci were found that had both significant read accumulation and exhibited read mapping with 10 overlaps. To assess these piRNA loci, we first compared size and expression of each (Fig 3A). Loci run the gamut from 1000 to 10,000 bp with relatively low expression for any given locus. This was somewhat unexpected. In other invertebrates massive loci are present that generate extensive amounts of piRNAs. An example of this is the ~180kb “piRNA cluster” flamenco locus, which in *Drosophila* serves as a repository of unlicensed transcripts (35). No such locus was found in *C. teleta*. The most extensive piRNA loci found in our efforts were less than 20kb (Fig 3B). This could be a consequence of the relatively poor assembly of the *C. teleta* genome, however, even in significantly poorer assembled genomes multiple such small RNA loci can be identified (36).

Another unexpected aspect of piRNA expression in *C. teleta* is the greater expression of piRNA in early developmental stages vs later stages (Fig 3C). Comparing expression across development found the highest level of expression in embryo stages, which was not seen in gravid

females. Thus, like the cluster miRNAs, piRNAs are either the product of early transcription or post-fertilization processing. Here more so than cluster miRNAs expression is higher in embryonic stages suggesting the later-post-fertilization processing. This is similar to what is seen in *Drosophila* where maternally deposited piRNA precursors serve to propagate piRNAs in the germline (37). Even so, some of the most highly expressed piRNA loci are expressed throughout development in all conditions. We also observe that gonad containing tissues (male head and female body) have higher expression of piRNAs, consistent with the presences of Piwi1 positive cells in germ cells (38). However, even the non-gonad containing tissues has abundant piRNAs, which might be associated with regeneration processes based on piwi1 expression in the posterior growth zone, for example.

To further characterize biogenesis of putative piRNAs we compared parameters previously described for piRNA processing. First, we simultaneously calculated the percent of reads aligning to each locus where the first base was a “U”, the average length of mapping reads, and locus length. By intersecting these parameters, we find that longer loci and reads with higher average read length exhibit greater 1U bias. Thus, like other species, piRNAs are characterized by longer reads (26+ bp) and 1U bias. Several short read loci were recovered that did not follow this trend. They appear to be loci where degradation fragments are abundant, but nevertheless may be the targets of piRNAs subject to turnover by piRNA-mediated cleavage. In addition to first base bias, we also examined ping pong and phasing biogenesis (Fig 3 E, F). Using a similar computational approach to find 10nt overlap loci we sought to visualize ping pong biogenesis in all loci. Unsurprisingly, in these loci we observe that most have a very clear signature of 10 nt overlap (Fig 3 E). However, a handful of loci are present that don't have a clear signature. Next, we also assessed the presence of piRNA phasing (Fig 3F). Here we found a strong phasing signature both at the 1-3 nt position

of 1U reads and at the following 30 nt position. Just like the ping pong signature there are some clear exceptions. Thus, while piRNA biogenesis described in systems such as *Drosophila* predominates, a handful of exceptions remain.

Comparing piRNA biogenesis patterns in *C. teleta* with Arthropods found ~98% of loci had shared characteristics such as clear phasing and pingpong signatures (Fig 4A). For these loci, there is seemingly random accumulation of reads in the region. The remaining 2%, however, showed a highly distinct expression pattern (Fig 4A,B) (Sup File 9). Unlike classic arthropod-type piRNA loci these atypical small RNA regions present as cluster-like repeating sites. Despite this unusual configuration these loci share some features with canonical piRNA loci such as partial ping pong overlaps, however, phasing is not shared. This is likely due to the distinct peaks of piRNAs that are clearly not the product of processive cleavages mediated by Zuc (Sup Fig 5).

To further characterize these loci, we curated features of both arthropod and non-arthropod varieties (Fig 4B). For both locus types they are found within intergenic, genic, and in intronic regions. Proportionally many non-arthropod types are encoded within gene annotations indicating they are associated with *bona fide* transcripts. Classic piRNA loci are found in all configurations where piRNAs are produced from both strands strand, biased towards one strand, or only expressed from a single strand. In contrast the atypical loci are majority uni-strand. A major concern with annotation of piRNAs is multi-mapping of piRNA-derived reads. We find that for both types unique mapping hits are present in roughly half of the non-arthropod loci. This further suggests that the atypical loci are confident sources of piRNAs, and not found due to mapping artifacts. Finally, we find that both types of loci have similar sizes.

Inspection of sequence identity at non-arthropod loci revealed that for some they consist of short repeat elements, while others have more complex sequences (Fig 4D). Even for highly

repetitive loci, polymorphisms are present within different repeat elements that when perfect mapping is enforced the expression profile is retained, all 16 loci retained their profile. Thus, it would further appear that the expression of these unusual piRNAs is not a mapping artifact but represent an alternate biogenesis mechanism. Moreover, similar loci have been noted in the pacific oyster, *C. gigas*, suggesting that this undescribed piRNA biogenesis mechanism may be present in multiple organisms (39).

A distinct siRNA pathway is absent in *C. teleta*

Unlike miRNAs and piRNAs our investigation did not uncover apparent endogenous siRNAs. This is apparent when *C. teleta* small RNA loci are compared to those of the planarian, *Schmidtea mediteranea*, a related bilateral animal where RNAi induced by long dsRNA is confidently validated (40). Small RNA studies in this planarian report all three classes of small RNAs (miRNAs, siRNAs and piRNAs). There are substantial differences in *S. mediteranea* RNAi machinery where there are 2 Dicers and 3 Agos compared to *C. teleta* 1 Dicer and 1 Ago (Fig 1A & Sup Fig 1). To further probe the differences in small RNA biogenesis between annelids and platyhelminths, we compared read size patterns of high expressing small RNA loci found with the method described for Fig 1B in *S. mediteranea* using a public small RNA dataset (41) (Fig 5A, Fig 1B). In the planarian analysis, most loci have mapping in the 19-23nt size range and are likely siRNAs or miRNA. Only a quarter of the small RNA loci were 29-30nt piRNAs, a striking difference to *C. teleta*.

Next, we sought small RNA loci with Dicer processing signatures based on the presence of reads with 2nt overhangs that range from 20-23nt long (42). These alignments were intersected with high expressed loci (Fig 1B, Fig 5A). Comparison of the 500 loci with highest small RNA

expression by size and Dicer overhang mapping revealing a substantially different distribution of loci when compared to *S. mediterranea* (Sup Fig 6A). *C. teleta* loci were substantially shorter in comparison with some examples in *S. mediterranea* that were longer than 40kb. The *C. teleta* loci typically corresponded to miRNAs while in *S. mediterranea* they are larger encoding a processively cleaved dsRNA. The largest locus from each species was compared to assess Dicer signatures (Sup Fig 6 B, C). In the locus from *C. teleta* a couple isolated mappings were seen whereas the *S. mediterranea* locus shows 21nt abundant reads that overlap with 2nt 3' overhangs. Thus, siRNA biology in annelids is, at a minimum, significantly reduced after diverging from planarians, which have impressive siRNA generating loci.

To further probe the apparent absence of siRNA class small RNA in *C. teleta* we assessed all possible sources involving dsRNA formed from dual strand transcription. Arthropods have 100's of such endogenous siRNA-loci that correspond to cis-NAT transcripts and transposable elements as well as cryptic dsRNA-producing loci (11). Due to the incomplete nature of *C. teleta* gene annotations only 122 instances of overlapping gene annotation are reported, and among those only 22 are overlaps greater than 40 bases (Sup Fig 6D). Alignments to the overlapping gene regions are limited with very few showing alignments of small RNAs in the size range of Dicer products (20-23nt). Even if these reads are present, they do not coincide with opposite strand alignments that would be expected from Dicer cleavage (Sup Fig 6E). To circumvent this limitation, we identified all regions of the *C. teleta* genome with mapping of 20-23nt reads and greater than 30 read coverage that did not correspond to miRNAs (Fig 5B). For each of the 5191 loci the ratio of forward to reverse reads was calculated and binned in to three categories: "bias" where there was 2-fold greater alignments on one strand, "mixed" where the ratio fell between 1.2-1.9 times greater on one strand, and "Equal" with the ratio being 1-1.1 (43). Alignment of 20-23nt

reads at these loci were simultaneously assessed for read overlaps such as 10 nucleotide overlaps seen for ping pong processing and 2nt less than full overlaps associated with Dicer cleavage. Loci were sorted by log strand bias ratio and visualized as a bar plot for bias and heatmaps for read overlap Z-scores (Fig 5B). Combined mixed and equal loci were roughly a quarter (24%) of the loci. In the overlap heatmaps, a strong pingpong signature was observed, suggesting that these 20-23nt reads are predominantly piRNA-type possibly truncated by trimming. Indeed, ~60% of the loci overlap with cluster 4 from Fig 1B. Simultaneously, a very minor signal was observable at the Dicer overlap position, particularly for 21nt reads (Fig 5B). However, these overlaps were not more prevalent in mixed or equal strand alignments with only 21% of loci with positive Dicer overlap Z-score values in these categories. Enrichment would be observed if the portion with Dicer signature exceeded 24%, suggesting no correlation between Dicer signatures and sites of potential dual strand transcription.

Next, we focused on loci from figure 5B that had mixed or equal strand mapping and a positive Z-score for Dicer overlaps in 21nt reads (19 nt overlaps) (Fig 5C). For this subset we sought to understand identity and expression of small RNAs by comparing RPKM values, presence of ping pong signature, and the relative abundance of short (20-23nt) relative to long (26-32nt) reads. Nearly all loci had greater alignment of longer reads compared to short reads with only 5 (1.8%) with more short vs long. Consistent with the bias towards longer reads ~80% of loci also had positive Z-scores for ping pong processing. Further highlighting the bias towards piRNAs, expression per locus tapered significantly away from those with a greater portion of piRNA-sized read alignments. Next, we quantified the number of 21nt read pairs that overlap by 19. Calculation of Z-scores in Fig 5B reports overlap biases of the indicated read size paired with a read of any other size, thus we sought to quantify 21nt reads paired with other 21nts (Fig 5D). From this we

found that over 75% of loci in Fig 5C have less than 10 unique 21nt reads (5 pairs). To further probe these loci for evidence of Dicer processing we inspected the 25% of loci that had the highest number of 21-21nt pairs overlapping by 19. Nearly all pair alignments were isolated suggesting distributive processing (Fig 5E, Sup Fig 7). Indeed, for the 47 loci examined only 6 had phased reads that would occur from processive cleavage. Together there is little evidence for siRNAs processed from long dsRNA. Instead, our results suggest *C. teleta* Dicer may be involved with cleavage of some substrates but may only be engaging in distributive processing.

To examine potential siRNA-related enzymatic activities in greater detail we characterized domains and catalytic residues in *C. teleta* Dicer and Ago (Sup Fig 8). A key domain involved in processive Dicer activity is the N-terminal Helicase domain that hydrolyzes ATP as a part of substrate engagement (44). While *C. teleta* Dicer has a recognized Helicase domain several of the key residues involved in ATP hydrolysis are altered in comparison to *D. melanogaster* Dicer2 and *H. sapiens* Dicer, which both show processive behavior. *C. teleta* Ago possess the same slicer residues as *D. melanogaster* Ago1 and Ago2, however, this is expected as miRNAs when pairing extensively with a target can direct slicing.

DISCUSSION

This comprehensive analysis of small RNA populations in the marine annelid *C. teleta* shows the presence of only two RNA classes, miRNAs and piRNAs, with an apparent loss of endogenous siRNAs. This is consistent with absence of siAgo proteins and was borne out by lack of compelling signatures of Dicer processing. Similar configurations of RNAi pathways components are found in other lophotrochozoans such as gastropods, cephalopods, and brachiopods, and leeches (21). The exception might be bivalves where there are two Ago

proteins, however, they seem to be result of duplication of a miAgo, and not a distinct miAgo/siAgo pair. The benefit of losing the siRNA pathway in these animals is not clear. In ecdysozoans, siRNAs have a significant role in anti-viral defense (45). It is curious that a useful mechanism would be jettisoned by *C. teleta* and possibly other lophotrochozoans. However, a similar event occurred independently in the deuterostomes, including basal echinoderms (46). In the place of siRNAs, in both lophotrochozoans and echinoderms piRNAs appear to take the place of viral siRNAs. Poriferans, in contrast seem to mount an siRNA response to viruses indicating that piRNA-mediated anti-viral defense has evolved independently.

In addition to the loss of siRNAs we observe unusual miRNA hairpins that do not exhibit unpaired stem bases with extremely short hairpins. More so than nearly all other species, they are abundant in *C. teleta*. The exception is the Nile tilapia, which has an unusual collection of miRNAs. Further research is needed to establish whether this is coincidental or a product of changes in Dicer enzymatic activity. Interestingly, these hairpins would be excellent candidates for a miR-451-like Dicer independent biogenesis, however, they do not appear to mature through this pathway. Reads align to hairpins in a pattern expected for Dicer processing, suggesting a competency for *C. teleta* Dicer to processes this type of hairpin that is not present in vertebrate Dicer proteins. It will also be interesting as to whether other annelids have these same tight hairpins found in *C. teleta* or if this worm is an outlier like the Nile Tilapia.

As we found with other small RNA classes, we also observe some unexpected features of *C. teleta* piRNAs. First, we did not observe a massive piRNA cluster which are observed in a variety of animals. Instead, we find loci that reach a maximum of 10kb, none of which exhibit high expression. This does not suggest, however, that piRNA function has diverged significantly from arthropods. Recent efforts to delete large piRNA clusters from *Drosophila* found that they

were dispensable for fertility and transposon control (47). Thus, in *C. teleta*, even though there is no large piRNA cluster, piRNAs can be expected to retain their role in transposon control. However, when this might happen during *C. teleta* development is questionable. Unlike many animals we do not observe the piRNA expression correlated to gonads. In this worm we find greater expression in the early embryos. Thus, germline defense may happen post-fertilization vs during gametogenesis, and that the collection of inherited piRNAs peak at a different point in this animal's life cycle.

In this species we also observe unusual piRNA loci that are produced by an undescribed mechanism. Unlike many piRNAs these are derived from simple repeat sequences. Tests that involve examining perfect and unique read mapping found that these piRNA loci were not annotated as a result of read mapping artifacts. Also this type of loci are found in related species (39). Processing mechanisms of these piRNAs are unknown and will require genetic studies that define both processing but also the function of piRNAs derived from these loci. It will also be intriguing to see if they have a distinct activity such as the trigger, responder and trailer piRNAs described in *Drosophila* (48). Alternately, these piRNAs may be like *C. elegans* piRNAs produced by discrete transcriptional units.

Likely, the most impactful outcome of this report is the guidance offered for developing RNAi approaches in animals that share *C. teleta* small RNA biology. The lack of a dedicated siRNA pathway suggests that long dsRNA approaches used in ecdysozoans and planarians may not be advisable. It would unsurprising that dsRNA molecules have become a pathogen pattern and could lead to an antiviral response and not specific gene knockdown. As an alternative, technology that exploits miRNA biology as it is deployed in vertebrates should be effective. The single Ago in *C. teleta* has slicer residues thus potent gene silencing should be possible. Further,

the unusual miRNA structures we observe may provide additional configurations for exploiting the miRNA pathway. Lastly, it may also be possible to exploit piRNA mechanisms by introducing synthetic RNAs that possess complementarity to known piRNAs (42). This species has abundant piRNAs so a variety of cell types could be targeted, also there are two piRNA types that could be modeled for gene silencing technology.

MATERIALS and METHODS

Capitella acquisition and culture:

Capitella teleta (juveniles and adults) were obtained from Dr. Elaine Seaver's lab at Whitney Laboratory for Marine Bioscience at University of Florida. They were grown in organically enriched mud from Biloxi bay, Mississippi. Around 20 adult worms (~10 of each male and females) were placed in a 500ml container with a tablespoon full of mud and 200ml sea water. Adults were fed once a week and juveniles were fed once every two weeks by adding a scoop of new mud to the containers. Worms were kept in growth chamber maintained at 20°C. Adult containers were routinely checked for the presence of brood tubes and embryos. Once found about 25 larva were transferred into new containers and placed in growth chamber and fed weekly with new mud.

RNA extraction from developmental stages of capitella

Early embryos were acquired by separating sexually mature males and females for 5-7 days and then keeping them together for 10-12 hours. Containers were then checked for the brood tubes containing new laid eggs. Early embryos (2 cell stage) from two brood tubes were collected for RNA extraction using 100 µl of Tri-Reagent LS. Similarly, late-stage embryo, larva (swimmers), 3 adult male anteriors (containing sperm sac) and 3 adult female anteriors were collected in 100

μl of Tri-Reagent LS. After grinding, 100 μl of deionized water and 800 μl of trizol LS was added and the extraction was completed following manufacturer protocols. Addition of deionized water was necessary to mitigate excess salts for efficient purification of RNAs. For posterior RNA extractions an additional purification using mirVana™ miRNA Isolation Kit was performed to address contaminants likely from fecal matter that interfered with nucleic acid manipulating enzymes and possibly gel electrophoresis. RNAs extracted from the 7 different tissues were then subjected to small RNA sequencing utilizing Illumina NextSeq500 after library construction with the Illumina TruSeq small RNA cloning kit. Two rounds of library construction and sequencing was used for adult posterior libraries. Quality of the datasets was validated using the mirTrace tool (Sup File 11) (49). All libraries were positively identified as lophotrochozoan with very little contamination by rRNA. For both embryo stages where piRNA are dominant, fewer miRNA sized reads were recovered as percent of the library. All clipped, unfiltered data are available through the NCBI SRA database under bioproject #PRJNA777269. *S. mediteranea* dataset was acquired from NCBI SRA under bioproject # PRJNA117181. Planarian small RNA libraries were generated using similar protocols to what we used for *C. teleta* (TRIzol extraction (Invitrogen) and T4 RNA ligase 2 (Rnl2(1–249)K227Q) for library preparation)

Small RNA analysis pipeline

Genome sequence and genome annotation files for *C. teleta* were acquired from Ensembl metazoan (*Capitella_teleta_v1.0*). Small RNA analysis was carried out using pipelines diagrammed in supplement (Sup Fig 9). Small RNA loci were identified using bowtie alignments converted to bedgraph format and filtered based on coverage using awk. Likewise, size distribution was determined using awk to quantify length of reads extracted from alignments to single loci.

miRNAs in *C. teleta* were investigated using mirDeep2 and standard parameters (22). Annotations from MirGeneDB were used to guide annotations (24). Novel miRNAs were assessed by manual curation. mirDeep2 calls were evaluated for presence of RNase III cleavage (2nt 3' overhangs) between mature and star strands. Potential miRNAs were also assessed for 5' processing precision where a >90% of reads aligning to a hairpin arm share a 5' base. Loci that met these criteria were classified as confident. If a miRNA did not meet both standards or if star reads did not exceed a coverage of 8 the miRNA was considered candidate. If the potential miRNA failed both criteria it was labeled a false positive and not reported. mirDeep2 annotation outputs provided for Known, Confident, and Candidate miRNAs (Sup File 3,5,7,8)

For heatmaps, libraries were normalized for number of reads mapped to the genome. A python-based algorithm was used to find the overlapping read pairs that represent Dicer and pingpong signatures (50). Small RNA of 15-31 nt were used to find targets (same length) having 10nt overlap indicating pingpong signature while overlap of 2nt less than query indicates Dicer signatures. piRNA loci were analyzed for phasing using piPipes (51). Graphics and visualizations were obtained using ggplot2, gplot, sushi plot, and pheatmap.

Acknowledgements

This work was supported by NSF 1845978, and Mississippi INBRE: funded by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM103476. Computational resources were provided by NSF: ACI 1626217.

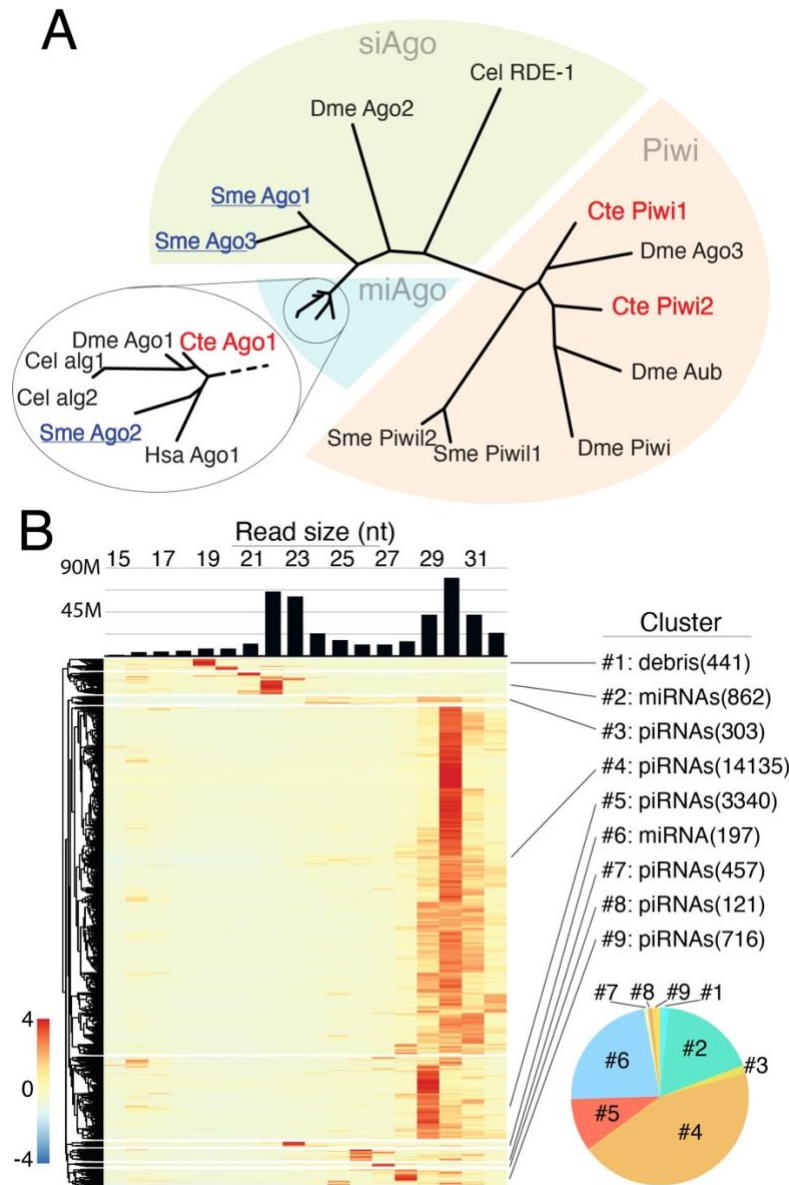


Figure 1: Small RNA populations of *C. teleta* A) Phylogenetic comparison of Argonaute protein from various organism (*C. elegans*, *S. mediterranea*, *D. melanogaster* and *H. sapiens*) with *C. teleta*. *C. teleta* Argonaute protein clusters together with miAgos. B) Size distribution of small RNA from >20000 small RNA producing loci. Bar graph shows bulk reads across all libraries. Heatmap is row normalized small RNA expression per locus. Loci were clustered into groups (#1-9) depending on the length of the RNA expressed at loci. Clusters of miRNAs contain 21-23nt reads while piRNA clusters contain 28-31nt reads. The cluster #1 was characterized as debris (degradation products) based on the small length of the RNAs, however a single annotated miRNA species (bantam) was found within the cluster. The pie chart (bottom right) shows the relative abundances of piRNAs (warm colors) and miRNAs (cool colors) based per cluster.

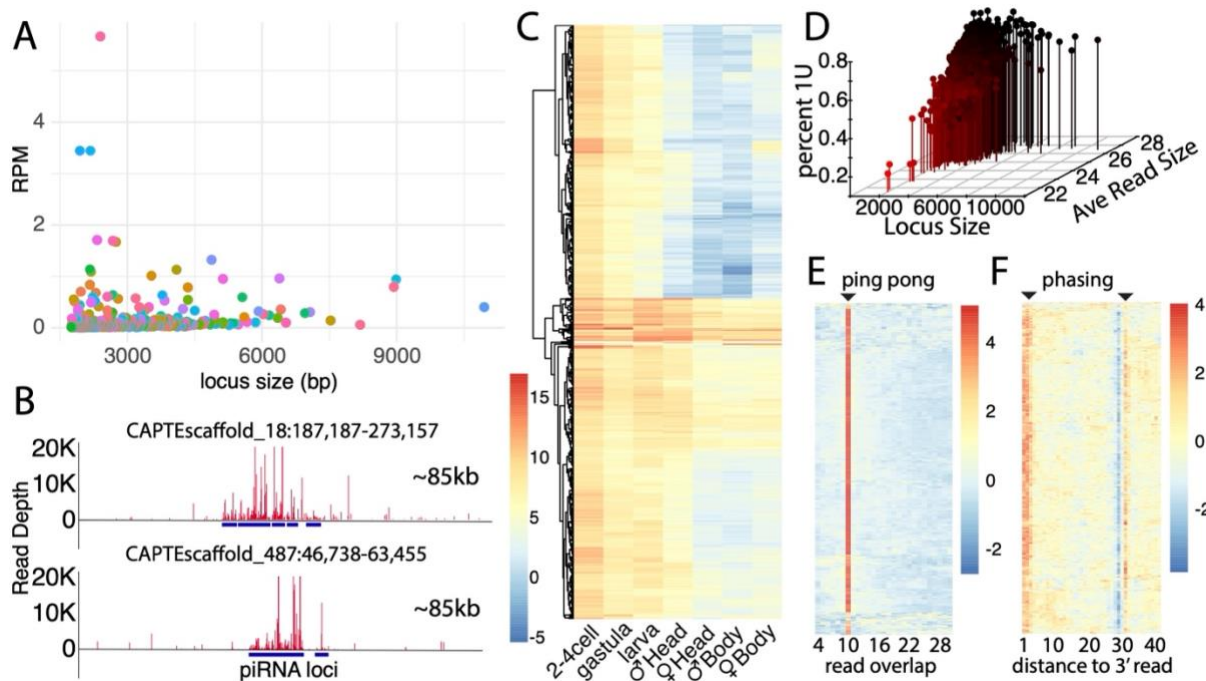


Figure 3: Capitella genome contains extensive piRNA loci. A) Size distribution of piRNA loci plotted against read expression in each locus (Reads per million). Majority of loci are smaller in length with not as much depth. Alternating colors are applied to assist datapoint visualization. B) Graphic representation of two of the largest piRNA loci located in scaffold 18 and 487. The entire window represents 85Kb and highest depth of reads of 20k each. C) piRNA expression across developmental stages. Adult male and female samples were further divided into anterior and posterior end (Head and Body, respectively) to record expression in gonads (male gonads are present in the anterior end and female gonads are present in the posterior end). Scores are based on normalized $\log_2(\text{RPM})$ values. D) Percentage of Uridine bias at first position for each piRNA locus. Percentage of bias is plotted against piRNA locus size and average read size. E) Ping pong signature in piRNA loci. Horizontal axis shows number of nucleotides overlap between reads in each locus. There are few loci exceptions (bottom) that do express 10 nt overlap. Values are based on z-score. F) Phasing signature in piRNA loci based on expression of U bias at each nt position. There is a strong presence of U bias at first position. Low expression at position 29 marks the splicing site for the RNA read and strong U bias at position 31 is indicative of the first nucleotide of the next read.

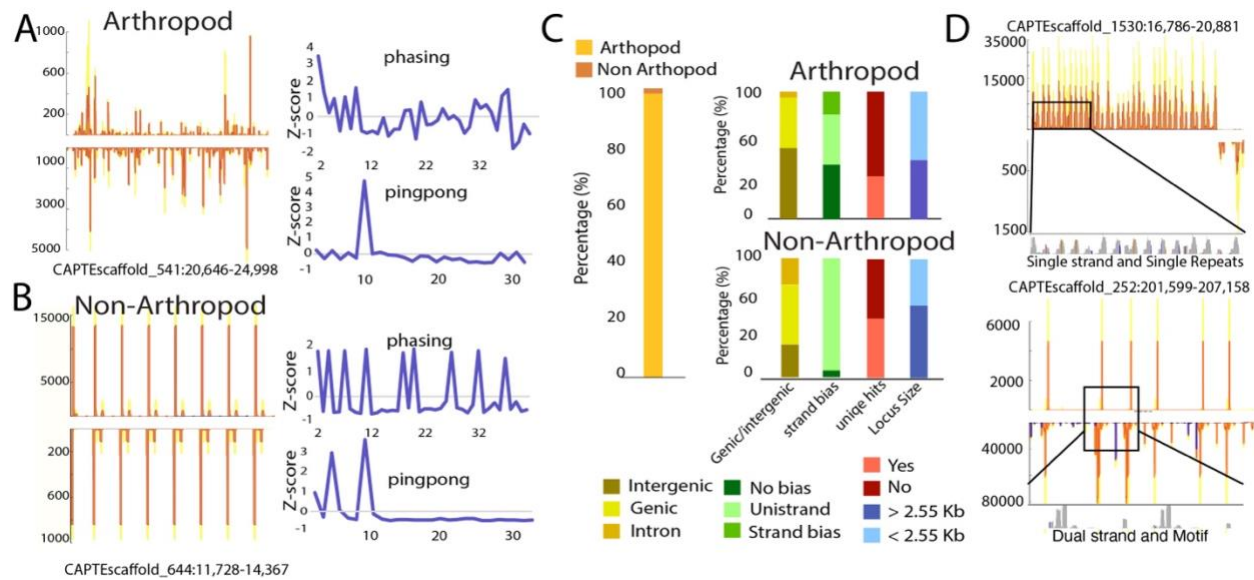


Figure 4: Non canonical piRNA loci consist of repetitive sequence. A-B) Graphic representation of Arthropod like and non-Arthropod like loci, respectively. Non-Arthropod like loci are characterized by the presence and expression of repetitive patterns with reads containing the same sequence. C) Characterization of Arthropod and non-Arthropod like loci. Percentage of piRNA loci that are Arthropod or non-Arthropod like (left). For each category, they were subdivided based on the physical features of each locus (right). Genomic location divided into intergenic regions, genic (loci spanned through exons and introns) or intronic (loci was located inside a large intronic region). Transcriptional bias was divided into no bias (total transcripts for one strand < 5x total transcript of the opposite strand), strand bias (total transcripts of one strand $\geq$ 5x but $\leq$ 10x total transcript of the opposite strand), and unistrand (total transcripts of one strand > 10x total transcripts of the opposite strand). Loci were also divided based on the presence of novel mapped reads (unique hits) and by locus size (larger or smaller than 2.55 Kb). D) Non-Arthropod loci were also divided into type of repetition present at each locus (single read repeats or motifs with multiple repetitive reads) and into the presence of repeats on one strand or both strands (single or dual strand).

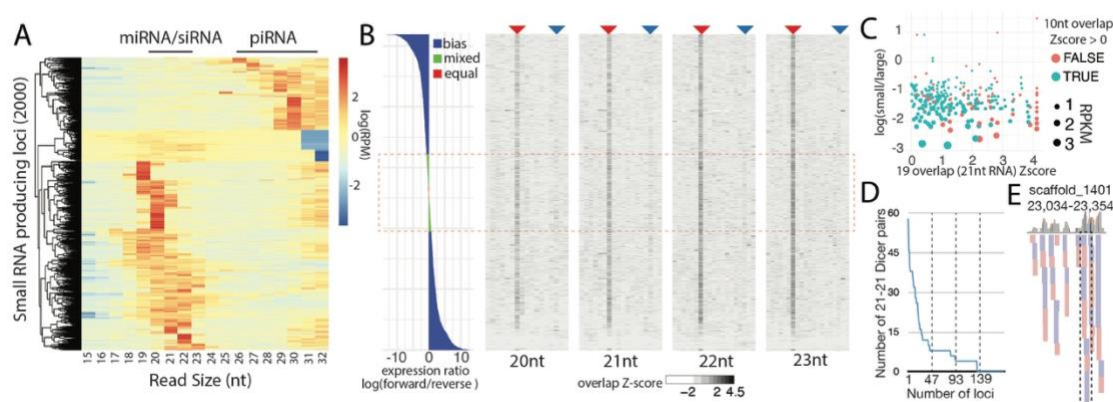


Figure 5: Little evidence for distinct siRNA biogenesis in *C. teleta*. A) Size distribution of small RNA species in *S. mediterranea*. Compared to a similar analysis in Fig1B there is a substantial greater fraction of small RNA loci with a distribution suggesting miRNA/siRNA identity. B) Characterization of all *C. teleta* loci with alignment of small 20-23nt reads with coverage > 30. Left panel is strand bias expressed as the log ratio of forward and reverse mapping reads. Right four panels show the Z score for read overlaps starting from 4 base overlaps to full read overlaps. The analysis is performed separately for reads of different lengths (20nt, 21nt, 22nt, and 23nt). The red arrow shows 10nt ping pong overlap, and blue arrow dicer overlaps (2 less than full overlap). All panels represent the same sorted loci top to bottom. Red dashed box indicates reads with mixed or equal number of alignments on both strands. C) Scatterplot of loci with 21nt reads with Dicer processing signature and Mixed/Equal read mapping. The Z-score of Dicer overlap compared to the log of small (20-23nt) to large (26-32nt). Points are colored based on whether a positive ping pong Z-score was also observed and sized scaled by RPKM for locus. D) Quantification of 21nt reads overlapping with Dicer overhangs at loci with a Dicer signature Z-score > 1. E) Alignment of 21nt reads overlapping by 19 bases at the locus measured in part D with the highest number of 21nt-21nt pairs. The dashed box shows the only example of a phased set of small RNAs.

- 564 1. A. S. Flynt: Insecticidal RNA interference, thinking beyond long dsRNA. *Pest*
565 *Management Science*, 77(5), 2179-2187 (2021) doi:<https://doi.org/10.1002/ps.6147>
- 566 2. V. N. Kim, J. Han and M. C. Siomi: Biogenesis of small RNAs in animals. *Nature Reviews*
567 *Molecular Cell Biology*, 10(2), 126-139 (2009) doi:10.1038/nrm2632
- 568 3. E. Wienholds and R. H. A. Plasterk: MicroRNA function in animal development. *FEBS*
569 *Letters*, 579(26), 5911-5922 (2005) doi:<https://doi.org/10.1016/j.febslet.2005.07.070>
- 570 4. N. Wynant, D. Santos and J. Vanden Broeck: The evolution of animal Argonautes:
571 evidence for the absence of antiviral AGO Argonautes in vertebrates. *Scientific Reports*, 7(1),
572 9230-9230 (2017) doi:10.1038/s41598-017-08043-5
- 573 5. M. Müller, F. Fazi and C. Ciaudo: Argonaute Proteins: From Structure to Function in
574 Development and Pathological Cell Fate Determination. *Frontiers in Cell and Developmental*
575 *Biology*, 7(360) (2020) doi:10.3389/fcell.2019.00360
- 576 6. J. G. Doench, C. P. Petersen and P. A. Sharp: siRNAs can function as miRNAs. *Genes &*
577 *development*, 17(4), 438-442 (2003) doi:10.1101/gad.1064703
- 578 7. W. H. Palmer, J. D. Hadfield and D. J. Obbard: RNA-Interference Pathways Display High
579 Rates of Adaptive Protein Evolution in Multiple Invertebrates. *Genetics*, 208(4), 1585-1599 (2018)
580 doi:10.1534/genetics.117.300567
- 581 8. D. P. Bartel: Metazoan MicroRNAs. *Cell*, 173(1), 20-51 (2018)
582 doi:10.1016/j.cell.2018.03.006
- 583 9. E. Berezikov and R. H. A. Plasterk: Camels and zebrafish, viruses and cancer: a microRNA
584 update. *Human Molecular Genetics*, 14(suppl_2), R183-R190 (2005) doi:10.1093/hmg/ddi271
- 585 10. Y. S. Lee, K. Nakahara, J. W. Pham, K. Kim, Z. He, E. J. Sontheimer and R. W. Carthew:
586 Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell*,
587 117(1), 69-81 (2004) doi:10.1016/s0092-8674(04)00261-2
- 588 11. K. Okamura and E. C. Lai: Endogenous small interfering RNAs in animals. *Nature*
589 *reviews. Molecular cell biology*, 9(9), 673-678 (2008) doi:10.1038/nrm2479
- 590 12. P. D. Zamore: Somatic piRNA biogenesis. *EMBO Journal*, 29(19), 3219-3221 (2010)
591 doi:10.1038/emboj.2010.232
- 592 13. E.-M. Weick and E. A. Miska: piRNAs: from biogenesis to function. *Development*,
593 141(18), 3458 LP-3471 (2014) doi:10.1242/dev.094037
- 594 14. R. J. Ross, M. M. Weiner and H. Lin: PIWI proteins and PIWI-interacting RNAs in the
595 soma. *Nature*, 505(7483), 353-359 (2014) doi:10.1038/nature12987
- 596 15. J. Brennecke, A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam and G. J.
597 Hannon: Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in
598 Drosophila. *Cell*, 128(6), 1089-1103 (2007) doi:10.1016/j.cell.2007.01.043
- 599 16. S. Yamanaka, M. C. Siomi and H. Siomi: piRNA clusters and open chromatin structure.
600 *Mobile DNA*, 5(1), 22-22 (2014) doi:10.1186/1759-8753-5-22
- 601 17. N. Robine, N. C. Lau, S. Balla, Z. Jin, K. Okamura, S. Kuramochi-Miyagawa, M. D.
602 Blower and E. C. Lai: A broadly conserved pathway generates 3'UTR-directed primary piRNAs.
603 *Current biology : CB*, 19(24), 2066-2076 (2009) doi:10.1016/j.cub.2009.11.064
- 604 18. S. H. Lewis, K. A. Quarles, Y. Yang, M. Tanguy, L. Frézal, S. A. Smith, P. P. Sharma, R.
605 Cordaux, C. Gilbert, I. Giraud, D. H. Collins, P. D. Zamore, E. A. Miska, P. Sarkies and F. M.

- Jiggins: Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. *Nat Ecol Evol*, 2(1), 174-181 (2018) doi:10.1038/s41559-017-0403-4
19. A. C. Billi, M. A. Freeberg, A. M. Day, S. Y. Chun, V. Khivansara and J. K. Kim: A Conserved Upstream Motif Orchestrates Autonomous, Germline-Enriched Expression of *Caenorhabditis elegans* piRNAs. *PLOS Genetics*, 9(3), e1003392-e1003392 (2013)
20. Z. Cao, D. Rosenkranz, S. Wu, H. Liu, Q. Pang, X. Zhang, B. Liu and B. Zhao: Different classes of small RNAs are essential for head regeneration in the planarian *Dugesia japonica*. *BMC Genomics*, 21(1), 876 (2020) doi:10.1186/s12864-020-07234-1
21. J. Jehn, D. Gebert, F. Pipilescu, S. Stern, J. S. T. Kiefer, C. Hewel and D. Rosenkranz: PIWI genes and piRNAs are ubiquitously expressed in mollusks and show patterns of lineage-specific adaptation. *Communications Biology*, 1(1), 137 (2018) doi:10.1038/s42003-018-0141-4
22. M. R. Friedlander, W. Chen, C. Adamidi, J. Maaskola, R. Einspanier, S. Knespel and N. Rajewsky: Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol*, 26(4), 407-15 (2008) doi:10.1038/nbt1394
23. A. Kozomara, M. Birgaoanu and S. Griffiths-Jones: miRBase: from microRNA sequences to function. *Nucleic Acids Research*, 47(D1), D155-D162 (2018) doi:10.1093/nar/gky1141
24. B. Fromm, E. Høye, D. Domanska, X. Zhong, E. Aparicio-Puerta, V. Ovchinnikov, Sinan U. Umu, Peter J. Chabot, W. Kang, M. Aslanzadeh, M. Tarbier, E. Mármol-Sánchez, G. Urgese, M. Johansen, E. Hovig, M. Hackenberg, Marc R. Friedländer and Kevin J. Peterson: MirGeneDB 2.1: toward a complete sampling of all major animal phyla. *Nucleic Acids Research* (2021) doi:10.1093/nar/gkab1101
25. A. Bhambri, N. Dhaunta, S. S. Patel, M. Hardikar, A. Bhatt, N. Srikakulam, S. Shridhar, S. Vellarikkal, R. Pandey, R. Jayarajan, A. Verma, V. Kumar, P. Gautam, Y. Khanna, J. A. Khan, B. Fromm, K. J. Peterson, V. Scaria, S. Sivasubbu and B. Pillai: Large scale changes in the transcriptome of *Eisenia fetida* during regeneration. *PLOS ONE*, 13(9), e0204234 (2018) doi:10.1371/journal.pone.0204234
26. J. Mohammed, A. S. Flynt, A. M. Panzarino, M. M. H. Mondal, M. DeCruz, A. Siepel and E. C. Lai: Deep experimental profiling of microRNA diversity, deployment, and evolution across the *Drosophila* genus. *Genome Res*, 28(1), 52-65 (2018) doi:10.1101/gr.226068.117
27. A. S. Flynt, N. Li, E. J. Thatcher, L. Solnica-Krezel and J. G. Patton: Zebrafish miR-214 modulates Hedgehog signaling to specify muscle cell fate. *Nature genetics*, 39(2), 259-263 (2007) doi:10.1038/ng1953
28. V. Agarwal, G. W. Bell, J.-W. Nam and D. P. Bartel: Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, 4, e05005 (2015) doi:10.7554/eLife.05005
29. Y. Liu, Z. Zhu, I. H. T. Ho, Y. Shi, J. Li, X. Wang, M. T. V. Chan and C. H. K. Cheng: Genetic Deletion of miR-430 Disrupts Maternal-Zygotic Transition and Embryonic Body Plan. *Frontiers in Genetics*, 11(853) (2020) doi:10.3389/fgene.2020.00853
30. J. Mohammed, D. Bortolamiol-Becet, A. S. Flynt, I. Gronau, A. Siepel and E. C. Lai: Adaptive evolution of testis-specific, recently evolved, clustered miRNAs in *Drosophila*. *RNA (New York, N.Y.)*, 20(8), 1195-1209 (2014) doi:10.1261/rna.044644.114
31. J. S. Yang, T. Maurin, N. Robine, K. D. Rasmussen, K. L. Jeffrey, R. Chandwani, E. P. Papapetrou, M. Sadelain, D. O'Carroll and E. C. Lai: Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. *Proc Natl Acad Sci U S A*, 107(34), 15163-8 (2010) doi:10.1073/pnas.1006432107
32. D. Cifuentes, H. Xue, W. Taylor David, H. Patnode, Y. Mishima, S. Cheloufi, E. Ma, S. Mane, J. Hannon Gregory, D. Lawson Nathan, A. Wolfe Scot and J. Giraldez Antonio: A Novel

- miRNA Processing Pathway Independent of Dicer Requires Argonaute2 Catalytic Activity. *Science*, 328(5986), 1694-1698 (2010) doi:10.1126/science.1190809
33. D. Pinhal, L. A. Bovolenta, S. Moxon, A. C. Oliveira, P. G. Nachtigall, M. L. Acencio, J. G. Patton, A. W. S. Hilsdorf, N. Lemke and C. Martins: Genome-wide microRNA screening in Nile tilapia reveals pervasive isomiRs' transcription, sex-biased arm switching and increasing complexity of expression throughout development. *Scientific Reports*, 8(1), 8248 (2018) doi:10.1038/s41598-018-26607-x
34. F. Mohn, D. Handler and J. Brennecke: piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science*, 348(6236), 812-817 (2015) doi:10.1126/science.aaa1039
35. C. Goriaux, E. Théron, E. Brasset and C. Vaury: History of the discovery of a master locus producing piRNAs: the flamenco/COM locus in *Drosophila melanogaster*. *Frontiers in genetics*, 5, 257-257 (2014) doi:10.3389/fgene.2014.00257
36. M. Mondal, P. Klimov and A. S. Flynt: Rewired RNAi-mediated genome surveillance in house dust mites. *PLoS Genet*, 14(1), e1007183 (2018) doi:10.1371/journal.pgen.1007183
37. M. H. Fabry, F. A. Falconio, F. Joud, E. K. Lythgoe, B. Czech and G. J. Hannon: Maternally inherited piRNAs direct transient heterochromatin formation at active transposons during early *Drosophila* embryogenesis. *eLife*, 10, e68573 (2021) doi:10.7554/eLife.68573
38. V. C. Giani, E. Yamaguchi, M. J. Boyle and E. C. Seaver: Somatic and germline expression of piwi during development and regeneration in the marine polychaete annelid *Capitella teleta*. *EvoDevo*, 2(1), 10-10 (2011) doi:10.1186/2041-9139-2-10
39. S. Huang, Y. Ichikawa, Y. Igarashi, K. Yoshitake, S. Kinoshita, F. Omori, K. Maeyama, K. Nagai, S. Watabe and S. Asakawa: Piwi-interacting RNA (piRNA) expression patterns in pearl oyster (*Pinctada fucata*) somatic tissues. *Scientific Reports*, 9(1), 247 (2019) doi:10.1038/s41598-018-36726-0
40. L. Rouhana, J. A. Weiss, D. J. Forsthoefel, H. Lee, R. S. King, T. Inoue, N. Shibata, K. Agata and P. A. Newmark: RNA interference by feeding in vitro-synthesized double-stranded RNA to planarians: methodology and dynamics. *Developmental dynamics : an official publication of the American Association of Anatomists*, 242(6), 718-730 (2013) doi:10.1002/dvdy.23950
41. M. R. Friedländer, C. Adamidi, T. Han, S. Lebedeva, T. A. Isenbarger, M. Hirst, M. Marra, C. Nusbaum, W. L. Lee, J. C. Jenkin, A. Sánchez Alvarado, J. K. Kim and N. Rajewsky: High-resolution profiling and discovery of planarian small RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 106(28), 11546-11551 (2009) doi:10.1073/pnas.0905222106
42. M. Mondal, J. K. Brown and A. Flynt: Exploiting somatic piRNAs in *Bemisia tabaci* enables novel gene silencing through RNA feeding. *Life Science Alliance*, 3(10), e202000731 (2020) doi:10.26508/lsa.202000731
43. C. Zhang and G. Ruvkun: New insights into siRNA amplification and RNAi. *RNA Biol*, 9(8), 1045-9 (2012) doi:10.4161/rna.21246
44. N. K. Sinha, J. Iwasa, P. S. Shen and B. L. Bass: Dicer uses distinct modules for recognizing dsRNA termini. *Science (New York, N.Y.)*, 359(6373), 329-334 (2018) doi:10.1126/science.aag0921
45. D. B. Gammon and C. C. Mello: RNA interference-mediated antiviral defense in insects. *Current opinion in insect science*, 8, 111-120 (2015) doi:10.1016/j.cois.2015.01.006

46. F. M. Waldron, G. N. Stone and D. J. Obbard: Metagenomic sequencing suggests a diversity of RNA interference-like responses to viruses across multicellular eukaryotes. *PLoS genetics*, 14(7), e1007533-e1007533 (2018) doi:10.1371/journal.pgen.1007533
47. D. Gebert, L. K. Neubert, C. Lloyd, J. Gui, R. Lehmann and F. K. Teixeira: Large *Drosophila* germline piRNA clusters are evolutionarily labile and dispensable for transposon regulation. *Molecular Cell*, 81(19), 3965-3978.e5 (2021) doi:<https://doi.org/10.1016/j.molcel.2021.07.011>
48. F. Mohn, D. Handler and J. Brennecke: Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science*, 348(6236), 812-817 (2015) doi:10.1126/science.aaa1039
49. W. Kang, Y. Eldfjell, B. Fromm, X. Estivill, I. Biryukova and M. R. Friedländer: miRTrace reveals the organismal origins of microRNA sequencing data. *Genome Biology*, 19(1), 213 (2018) doi:10.1186/s13059-018-1588-9
50. C. Antoniewski: Computing siRNA and piRNA Overlap Signatures. In: *Animal Endo-SiRNAs: Methods and Protocols*. Ed A. Werner. Springer New York, New York, NY (2014) doi:10.1007/978-1-4939-0931-5_12
51. B. W. Han, W. Wang, P. D. Zamore and Z. Weng: piPipes: a set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq, ChIP-seq and genomic DNA sequencing. *Bioinformatics*, 31(4), 593-5 (2015) doi:10.1093/bioinformatics/btu647