

2-1-2022

## The Small RNA Universe of *Capitella teleta*

Sweta Khanal  
*University of Southern Mississippi*

Beatriz Schueng Zancanela  
*University of Southern Mississippi*

Jacob Oche Peter  
*University of Southern Mississippi, jacob.peter@usm.edu*

Alex Sutton Flynt  
*University of Southern Mississippi, alex.flynt@usm.edu*

Follow this and additional works at: [https://aquila.usm.edu/fac\\_pubs](https://aquila.usm.edu/fac_pubs)

---

### Recommended Citation

Khanal, S., Zancanela, B. S., Peter, J. O., Flynt, A. S. (2022). The Small RNA Universe of *Capitella teleta*. *Frontiers in Molecular Biosciences*, 9.  
Available at: [https://aquila.usm.edu/fac\\_pubs/20467](https://aquila.usm.edu/fac_pubs/20467)

This Article is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in Faculty Publications by an authorized administrator of The Aquila Digital Community. For more information, please contact [Joshua.Cromwell@usm.edu](mailto:Joshua.Cromwell@usm.edu).



# The Small RNA Universe of *Capitella teleta*

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

Cellular and Molecular Biology, University of Southern Mississippi, Hattiesburg, MS, United States

## OPEN ACCESS

### Edited by:

Taichiro Iki,  
Osaka University, Japan

### Reviewed by:

Bastian Fromm,  
UiT The Arctic University of Norway,  
Norway  
Anne Dueck,  
Technical University of Munich,  
Germany  
Hao Chen,  
Novartis Institutes for BioMedical  
Research, United States

### \*Correspondence:

Alex Sutton Flynt  
alex.flynt@usm.edu

### Specialty section:

This article was submitted to  
RNA Networks and Biology,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 27 October 2021

**Accepted:** 03 February 2022

**Published:** 25 February 2022

### Citation:

Khanal S, Zancanela BS, Peter JO and  
Flynt AS (2022) The Small RNA  
Universe of *Capitella teleta*.  
Front. Mol. Biosci. 9:802814.  
doi: 10.3389/fmolb.2022.802814

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 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 a significant impact on RNAi technology development for this group of animals. Unlike the apparent absence of siRNAs, a significant population of piRNAs was 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 to be present in the wider clade. This study is the first in providing comprehensive analysis of small RNAs in annelids.

**Keywords:** miRNA, piRNA, RNAi, annelid, development

## 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, the behavior of RNAi pathways varies from species to species, making application of a single paradigm inappropriate (Flynt, 2021). Among invertebrate animals, RNAi pathways appear to be mainly plastic, indicating that at a minimum an order—if not family-level investigation of biogenesis mechanisms is necessary for effective development of the gene-silencing technology.

Animal small RNAs belong to three classes: microRNAs (miRNAs), PIWI-interacting RNAs (piRNAs), and small-interfering RNAs (siRNAs) (Kim et al., 2009). Among the three major classes of RNAi, miRNAs are the most conserved with sequences shared in nearly all animals (Wienholds and Plasterk, 2005). In comparison, both endogenously expressed siRNAs (endo-siRNAs) and piRNAs show almost no conservation even at the species level, presumably due to their role in silencing invasive nucleic acids (Wynant et al., 2017). In addition to these conservation patterns, each class is defined by loading into a distinct class of Argonaute/PIWI (Ago/PIWI) proteins. Work from arthropods, where each class is well-represented, provides definitions where Ago1 (miAgo), Ago2 (siAgo), and PIWI proteins load miRNAs, siRNAs, and piRNAs, respectively. Arthropod PIWI

proteins include PIWI, Aubergine (Aub), and Ago3. In contrast, even though vertebrate genomes encode multiple Agos, usually four, they do not have dedicated siAgos (Müller et al., 2020). Thus, RNAi/siRNA technology in vertebrates is based on miRNA mimicry and is distinct from siRNAs used for RNAi in flies and nematodes (Doench et al., 2003). Vertebrates do possess numerous PIWI proteins as 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 proteins, with siAgos being the most rapidly evolving (Palmer et al., 2018).

In addition to association with distinct effectors, RNAi pathways are also defined by biogenesis. miRNAs 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 enzyme Dicer (Bartel, 2018). miRNA hairpin precursors have features that include asymmetric bulges with larger terminal loops of ~10 nt depending on the species. Pre-miRNA stem sequences are deeply conserved, while the terminal loop sequence is less so (Berezikov and Plasterk, 2005). siRNAs are also produced by Dicer and are typically 20–24 nt but are instead derived from long double-stranded RNA (dsRNA) molecules (Lee et al., 2004). Various sources yield siRNAs that include viruses along with may endogenous species such as hairpin RNAs, *cis*-NATs, and transposable elements (Okamura and Lai, 2008).

piRNA biogenesis occurs from two distinct processes: phasing and ping-pong amplification. Both phasing-dependent piRNAs (phasi-piRNAs) and ping-pong piRNAs are created independent of the RNase III activity, with processing driven by the “slicer” RNase activity intrinsic to PIWIs (Zamore, 2010). A consequence of this is a different size range (25–32 nt) compared to miRNAs and siRNAs (20–24 nt). 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 (Ross et al., 2014; Weick and Miska, 2014). Phasi-piRNAs are defined by 1U bias and close proximity (1–3 nt) between the 3' end of an upstream piRNA and the 5' end of a downstream piRNA. Ping-pong 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 (Brennecke et al., 2007). As ping-pong piRNAs are generated by slicing of complementary transcripts, they can be identified by 10 nt overhangs between piRNAs. Phasi-piRNA and ping-pong piRNA biogenesis collaborate to generate piRNAs to suppress the expression of unlicensed transcripts.

The paradigm for piRNA function is suppression of transposable elements (TEs) both in arthropods and vertebrates. In these species, large piRNA clusters serve as repositories of forbidden elements (Yamanaka et al., 2014). piRNA clusters are found both in uni-strand and dual-strand arrangements that yield phasi-piRNAs, which subsequently participate in the ping-pong cycle (Yamanaka et al., 2014). There are also genic piRNAs produced from UTRs by phasing biogenesis, which are likewise found in vertebrates and invertebrates (Robine et al., 2009). Initially, piRNAs were

viewed as exclusive to germlines; however, for many invertebrates, piRNAs are also present in soma where they not only appear to suppress TEs but also participate in gene regulatory networks (Lewis et al., 2018). In contrast, *C. elegans* piRNAs have a completely distinct biogenesis, with each piRNA being produced from short autonomous transcriptional units, defined by a specific motif ~ 50 nt upstream of the piRNA transcriptional start site (Billi et al., 2013).

RNAi pathways are well-documented in model organisms such as *Drosophila melanogaster* and *Caenorhabditis elegans*, as well as in vertebrates like mice and humans. In comparison, there has been substantially less investigation of RNAi in spiralian, one of three animal superphyla that include flatworms, annelids, and mollusks, which branched ~700–850 million years from Ecdysozoa. Research has mostly been in planarians which found all three small RNA classes with a pronounced expansion of piRNAs (Cao et al., 2020). Other studies have investigated small RNAs in mollusks that likewise noted substantial expansion of piRNAs (Jehn et al., 2018). These leave one major spiralian group, annelids, where there has been no comprehensive investigation of small RNA classes. Here, we describe small RNA expression and biogenesis in the annelid 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.

## MATERIALS AND METHODS

### *Capitella* Acquisition and Culture

*Capitella teleta* (juveniles and adults) were obtained from Dr. Elaine Seaver's lab at the Whitney Laboratory for Marine Bioscience at the 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 500 ml container with a tablespoon full of mud and 200 ml sea water. Adults were fed once a week, and juveniles were fed once every 2 weeks, by adding a scoop of new mud to the containers. Worms were kept in a growth chamber maintained at 20°C. Adult containers were routinely checked for the presence of brood tubes and embryos. Once found, about 25 larvae were transferred into new containers and placed in a 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 h. Containers were then checked for the brood tubes containing new laid eggs. Early embryos (2-cell stage) from

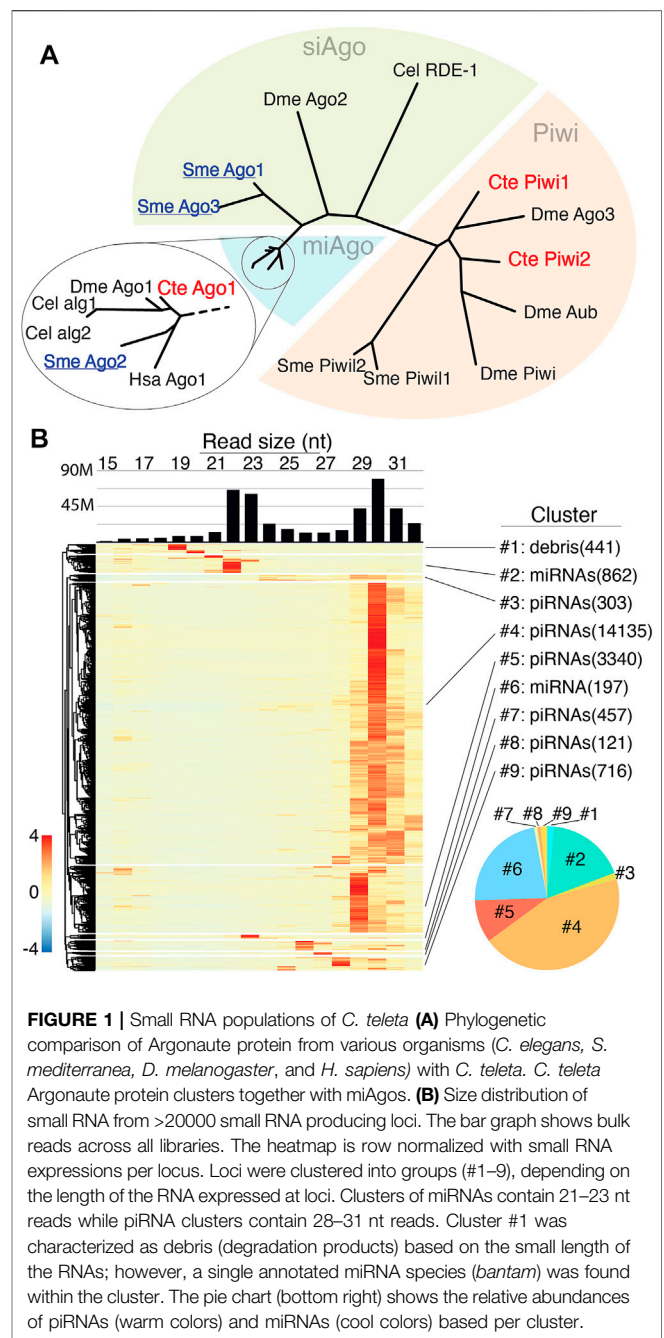
two brood tubes were collected for RNA extraction using 100  $\mu$ l of TRI Reagent LS. Similarly, late-stage embryo, larva (swimmers), three adult male anteriors (containing sperm sac), and three adult female anteriors were collected in 100  $\mu$ l of TRI Reagent LS. After grinding, 100  $\mu$ l of deionized water and 800  $\mu$ l of TRIzol LS were 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, additional purification using the 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 seven 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 were used for adult posterior libraries. The quality of the datasets was validated using the miRTrace tool (Supplementary File S10) (Kang et al., 2018). All libraries were positively identified as lophotrochozoan with very little contamination by rRNA. For both embryo stages, where piRNAs are dominant, fewer miRNA-sized reads were recovered as a percent of the library. All clipped, unfiltered data are available through the NCBI SRA database under the bioproject #PRJNA777269. The *S. mediterranea* dataset was acquired from NCBI SRA under the bioproject # PRJNA117181. Planarian small RNA libraries were generated using protocols similar 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 Metazoa (Capitella\_teleta\_v1.0). Small RNA analysis was carried out using pipelines diagrammed in supplement (Supplementary Figure S9). Small RNA loci were identified using bowtie alignments converted to the bedgraph format and filtered based on coverage using awk. Similarly, size distribution was determined using awk to quantify the length of reads extracted from alignments to single loci.

miRNAs in *C. teleta* were investigated using miRDeep2 and standard parameters (Friedländer et al., 2008). Annotations from MirGeneDB were used to guide annotations (Fromm et al., 2021). Novel miRNAs were assessed by manual curation. miRDeep2 calls were evaluated for presence of RNase III cleavage (2 nt 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 an miRNA did not meet both standards or if star reads did not exceed a coverage of eight, the miRNA was considered candidate. If the potential miRNA failed both criteria, it was labeled a false positive and not reported. miRDeep2 annotation provided outputs for known, confident, and candidate miRNAs (Supplementary Files S3, S5, S7, S8).

For heatmaps, libraries were normalized for the number of reads mapped to the genome. A python-based algorithm was used to find the overlapping read pairs that represent Dicer and ping-pong signatures (Antoniewski and Werner, 2014). Small RNAs of 15–31 nt were used to find targets (same length) having a 10 nt overlap, indicating



**FIGURE 1 |** Small RNA populations of *C. teleta* (A) Phylogenetic comparison of Argonaute protein from various organisms (*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. The bar graph shows bulk reads across all libraries. The heatmap is row normalized with small RNA expressions per locus. Loci were clustered into groups (#1–9), depending on the length of the RNA expressed at loci. Clusters of miRNAs contain 21–23 nt reads while piRNA clusters contain 28–31 nt reads. 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.

the ping-pong signature, while the overlap of 2 nt less than a query indicates Dicer signatures. piRNA loci were analyzed for phasing using piPipes (Han et al., 2015). Graphics and visualizations were obtained using ggplot2, gplot, sushi plot, and pheatmap.

## RESULTS

### *C. teleta* Global Small RNA Populations

We began assessment of *C. teleta* RNAi pathways by examining this annelid's Ago proteins (Figure 1A). Surprisingly, only three



Ago/PIWI proteins are encoded in the genome. To predict likely functions, a 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 piRNA biogenesis is active. Missing from *C. teleta* was a siAgo, which is present in basal planarians. The absence of siAgo is correlated with *C. teleta* only having a single Dicer protein that is related to miRNA processing rather than siRNA processing (Figure 1A). In contrast, *S. mediterranea* possesses two, one which groups with miRNA-processing Dicer and the other with siRNA-processing Dicer (Supplementary Figure S1).

To understand the Universe of small RNAs coordinated by these factors, small RNA sequencing was performed on RNAs extracted from early embryos (1–4 cell), late embryos (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 in anterior segments; thus, the four adult libraries represent separately male and female gonads, as well as soma only anterior and posterior. In total, 187M small reads were acquired across the libraries, for which 84% were mapped to the *C. teleta* genome. Highly redundant sequences that reported more than 100 alignments per read were excluded from mapping (Supplementary Table S1). For initial assessment of small RNA species, all libraries were combined, followed by mapping of 15–32 nt reads and application of a small RNA loci finding strategy based on read sites with coverage  $\geq 200$  reads and merging of features within 500 bp. From this, >20,000 locus-expressing small RNAs were identified (Supplementary File S1). The size distribution of mapping reads to all loci showed two peaks, one at 22–23 nt and another at 29–31 nt (Figure 1B). Based on observations from other invertebrates, these two peaks likely represent Ago-loaded Dicer products (miRNAs/siRNAs) and piRNAs, respectively. The next size distribution per locus was calculated, which allowed segregation into nine clusters (Figure 1B). A super majority (~90%) of loci represented by six of the nine clusters showed mapping of 29–31 nt reads, suggesting that 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% were 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 heterogeneous sizes (clusters: 3, 7, 8, and 9) or mapping of small,  $\leq 20$  nt sizes (cluster 1) comprised ~6% of all alignments. Next, we applied seqlogo analysis to further characterize small RNAs represented in each group (Supplementary Figure S2). Cluster 1 was predominantly represented by the sequence of the miRNA *bantam*—the one miRNA found in this group. The other miRNA clusters had little

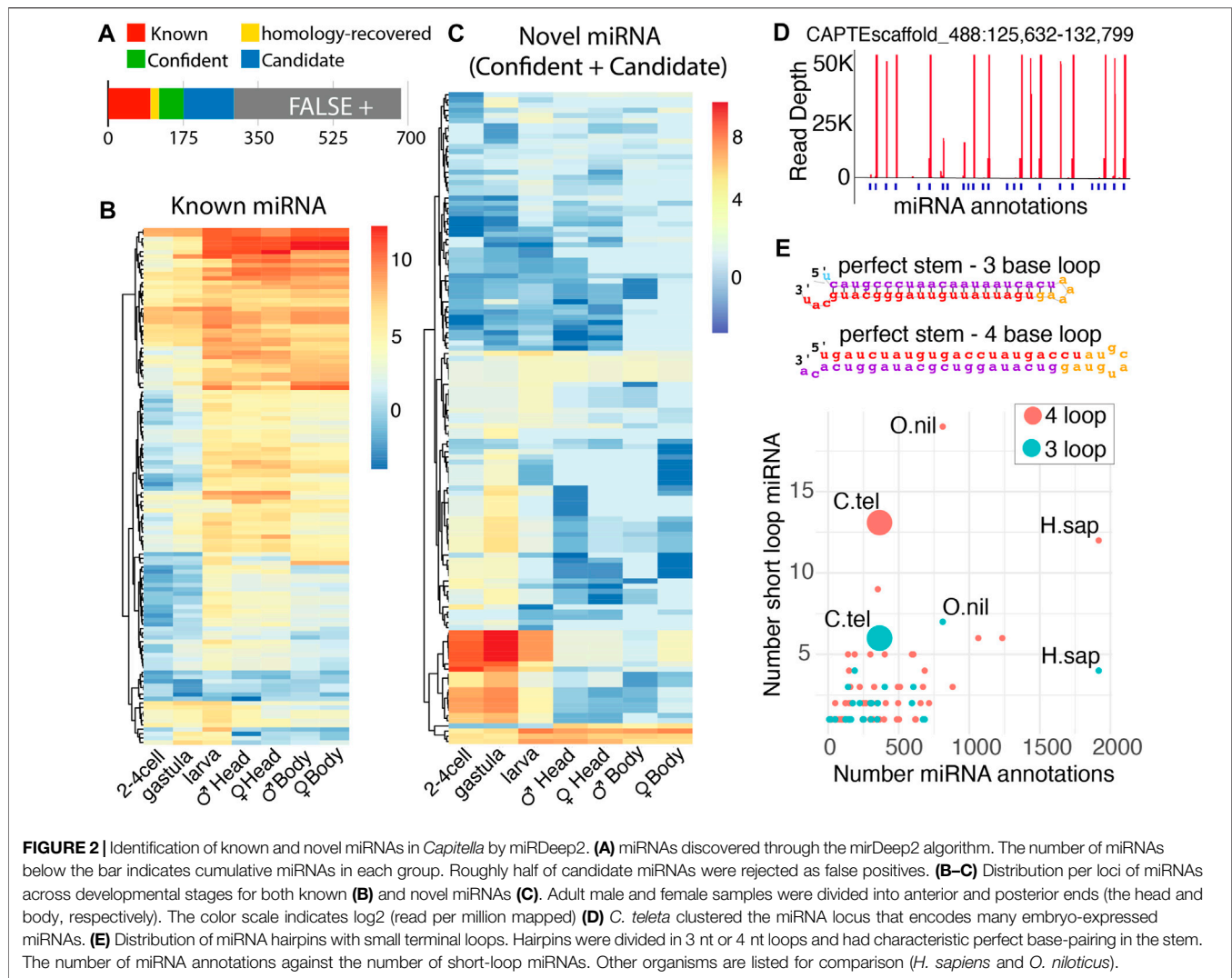
bias, which 1T (or 1U) bias was seen for several putative piRNA groupings, such as cluster #4. Together, these results show the expected miRNA but also abundant piRNA populations in *C. teleta* datasets.

## C. teleta miRNAs

To further examine *C. teleta* small RNA classes, we first sought to annotate miRNAs using the miRDeep2 algorithm (Supplementary File S2–S8) (Friedländer et al., 2008). In total, ~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* (Bhambri et al., 2018; Kozomara et al., 2018; Fromm et al., 2021). Of the 102 annotations in MirGeneDB, 99 were found (Supplementary File S3). 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 (Supplementary File S4). Nevertheless, this shows that 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* (Supplementary File S5). In some cases, these homology-rescued annotations appear to result from near identical duplication of *C. teleta* miRNAs or to have arisen earlier in the annelid lineage (Supplementary File S6).

Through this analysis, we were able to annotate novel miRNAs, which were individually vetted based on the expression, presence of Drosha and Dicer cleavage signatures, and precision of 5' end processing (Figure 2A) (Supplementary File S5–S7) (Mohammed et al., 2018). Out of 569 predicted novel miRNAs, 58 showed all features and were categorized as confident (Supplementary File S7), and 117 were denoted as candidates due to suboptimal features that were not indicative of RNase III processing (Supplementary File S8). The last group of 390 was labeled as false positives and was 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 the 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 a confident status and uncover additional miRNAs. Based on efforts in *Drosophila*, saturating the sequencing depth for recovery of all miRNAs in the 333 Mb genome of *C. teleta* would be achieved with ~600 million reads (Mohammed et al., 2018). 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 stages (Figure 2B). This has been



**FIGURE 2 |** Identification of known and novel miRNAs in *Capitella* by miRDeep2. **(A)** miRNAs discovered through the miRDeep2 algorithm. The number of miRNAs below the bar indicates cumulative miRNAs in each group. Roughly half of candidate miRNAs were rejected as false positives. **(B–C)** Distribution per loci of miRNAs across developmental stages for both known **(B)** and novel miRNAs **(C)**. Adult male and female samples were divided into anterior and posterior ends (the head and body, respectively). The color scale indicates log2 (read per million mapped) **(D)** *C. teleta* clustered the miRNA locus that encodes many embryo-expressed miRNAs. **(E)** Distribution of miRNA hairpins with small terminal loops. Hairpins were divided in 3 nt or 4 nt loops and had characteristic perfect base-pairing in the stem. The number of miRNA annotations against the number of short-loop miRNAs. Other organisms are listed for comparison (*H. sapiens* and *O. niloticus*).

observed in other species and reflects the greater cell type diversity that arises during differentiation (Flynt et al., 2007). The 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 the brain and pharynx. Minor differences are seen between genders for both posterior and anterior segments. The expression of miRNAs is strikingly similar between the 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 the post-zygotic stage through larval stages. They were not observed in the gravid female body sample, suggesting they are either produced from embryonic transcription or maternally deposited as intact precursors. The discovery of these miRNAs is likely due to sampling of embryonic stages, which was not performed for other annelids or mollusks. Considering their prolonged expressions, these miRNAs may have a role in regulating 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 (**Supplementary Figure S3A**) (Agarwal et al., 2015). Ingestion does not begin until swimming larvae 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 a nearly identical expression during early embryogenesis (**Figure 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 (Mohammed et al., 2014; Liu et al., 2020). 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 (**Supplementary Figure S3B**). 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 (**Supplementary Figure S4**).

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 (**Figure 2E**). In the *C. teleta* genome, we observed 13 species with a four-base loop and six species with a three-base loop. Both were reminiscent of miR-451, which is processed in a Dicer-independent manner (Cifuentes et al., 2010; Yang et al., 2010). However, for these miRNAs, this does not appear to be the case. They exhibited strong evidence of the 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 the miRbase (Kozomara et al., 2018). 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 (Pinhal et al., 2018). 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 reported for *C. teleta*, we recovered even more than the perhaps over-aggressively annotated human genome.

### **C. teleta piRNAs**

Even more abundant than miRNAs in datasets were the apparent piRNAs in *C. teleta*, both in terms of the number of putative piRNA loci and abundance of piRNA molecules (**Figure 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 and trigger the piRNA-mediated cleavage that initiates phasi-piRNAs (Mohn et al., 2015a). This read mapping profile was overlaid with all loci with 15–31 nt aligning reads that were over 1,000 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 the size and expression of each (**Figure 3A**). Loci run the gamut from 1,000 to 10,000 bp with a relatively low expression for any given locus. This was rather unexpected. In other invertebrates, massive loci are present that generate extensive amounts of piRNAs. An example of this is the ~180 kb “piRNA cluster” flamenco locus, which in *Drosophila* serves as a repository of unlicensed transcripts (Goriaux et al., 2014). No such locus was found in *C. teleta*. The most extensive piRNA loci found in our efforts were less than 20 kb (**Figure 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 (Mondal et al., 2018).

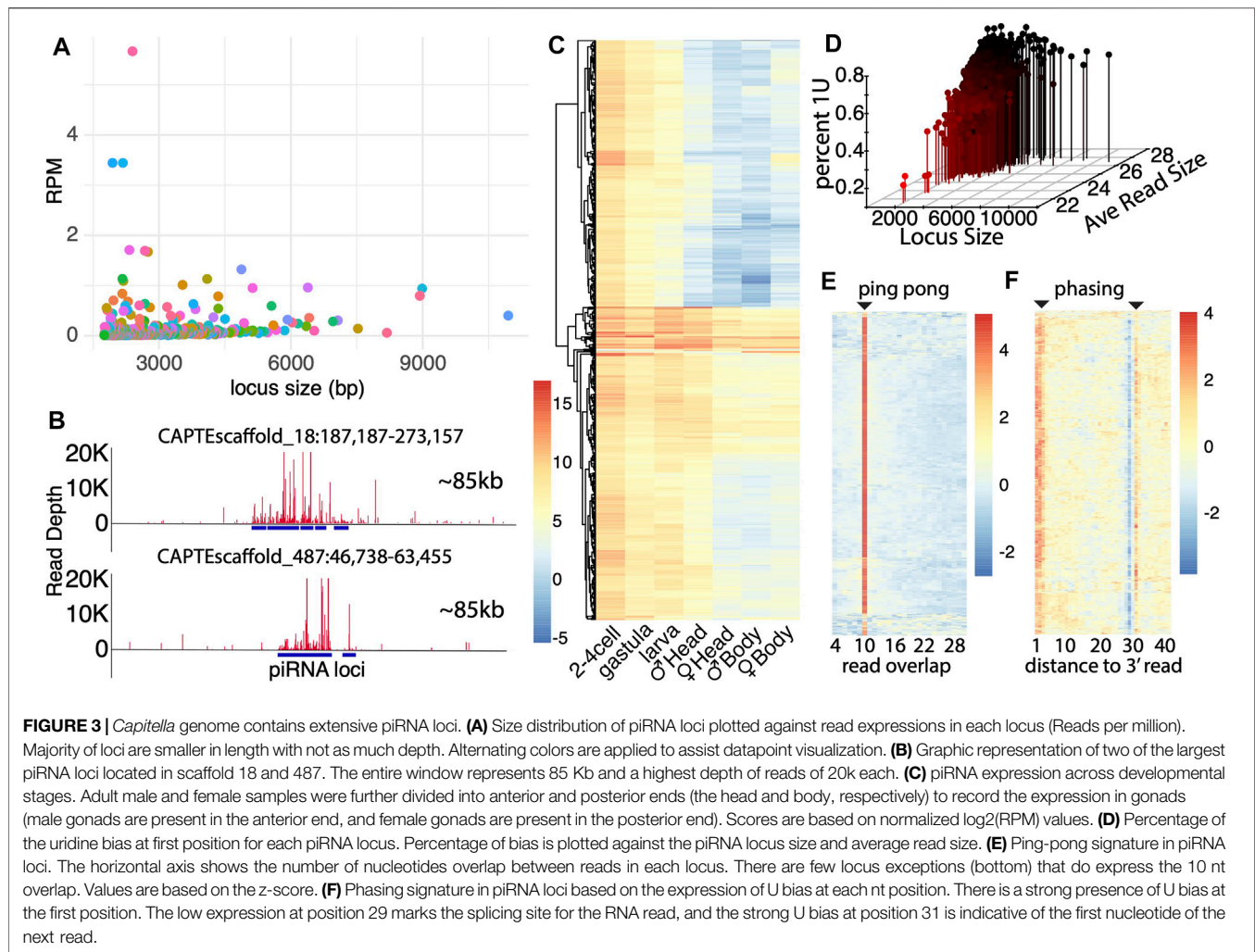
Another unexpected aspect of the piRNA expression in *C. teleta* is the greater expression of piRNA in early developmental stages vs. later stages (**Figure 3C**). Comparing expressions across development found the highest level of expression in embryo stages, which was not seen in gravid females. Thus,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

miRNAs, piRNAs are either the product of early transcription or post-fertilization processing. Here, more so the cluster miRNA expression is higher than 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 (Fabry et al., 2021). Nevertheless, some of the most highly expressed piRNA loci are expressed throughout development in all conditions. We also observed that gonad-containing tissues (the male head and female body) have higher expressions of piRNAs, consistent with the presence of PIWI1-positive cells in germ cells (Giani et al., 2011). However, even the non-gonad-containing tissues have abundant piRNAs, which might be associated with regeneration processes based on the PIWI1 expression in the posterior growth zone, for example.

To further characterize biogenesis of putative piRNAs, we compared parameters that were 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 found that longer loci and reads with higher average read lengths 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 piRNA subjects to turnover by the piRNA-mediated cleavage. In addition to the first base bias, we also examined ping-pong and phasing biogenesis (**Figures 3E,F**). Using a similar computational approach to find 10 nt overlap loci, we sought to visualize ping-pong biogenesis in all loci. Unsurprisingly, in these loci, we observed that most have a very clear signature of 10 nt overlaps (**Figure 3E**). However, a handful of loci are present that do not have a clear signature. Next, we also assessed the presence of piRNA phasing (**Figure 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*, we found that arthropods with ~98% of loci had shared characteristics such as clear phasing and ping-pong signatures (**Figure 4A**). For these loci, there is a seemingly random accumulation of reads in the region. The remaining 2%, however, showed a highly distinct expression pattern (**Figures 4A,B**) (**Supplementary File S9**). 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 (**Supplementary Figure S5**).

To further characterize these loci, we curated features of both arthropod and non-arthropod varieties (**Figure 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, biased towards one strand, or only expressed from a single strand. In contrast, the majority of atypical loci are uni-strand. A major concern with annotation of piRNAs is multi-mapping of piRNA-derived reads. We found 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 are not found due to mapping artifacts. Finally, we found that both types of loci have similar sizes.

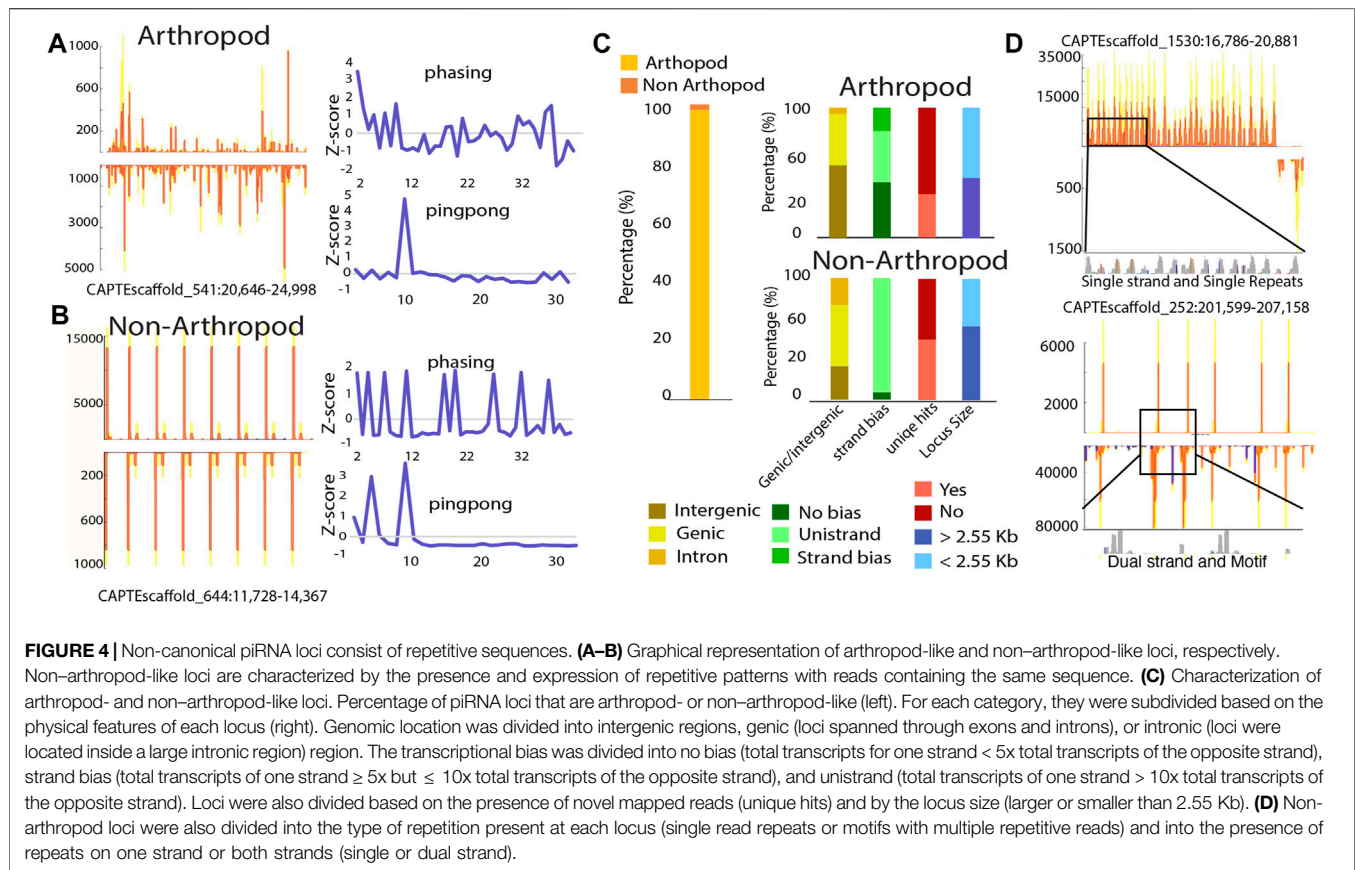
Inspection of sequence identity at non-arthropod loci revealed that for some, they consist of short repeat elements, while others have more complex sequences (Figure 4D). Even for highly repetitive loci, polymorphisms are present within different repeat elements that when perfect mapping is enforced, the expression profile is retained; all 16 loci retained their profiles. 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 (Huang et al., 2019).

## 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 mediterranea*, a related bilateral animal where RNAi induced by long dsRNA is confidently validated (Rouhana et al., 2013). Small RNA studies in this planarian report all three classes of small RNAs (miRNAs, siRNAs, and piRNAs). There are substantial differences in *S. mediterranea* RNAi machinery, where there are 2 Dicers and 3 Agos compared to *C. teleta* 1 Dicer and 1 Ago (Figure 1A; Supplementary Figure S1). 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 in Figure 1B for *S. mediterranea*, using a public small RNA dataset (Friedländer et al., 2009) (Figures 1B, 5A). In the planarian analysis, most loci have mapping in the



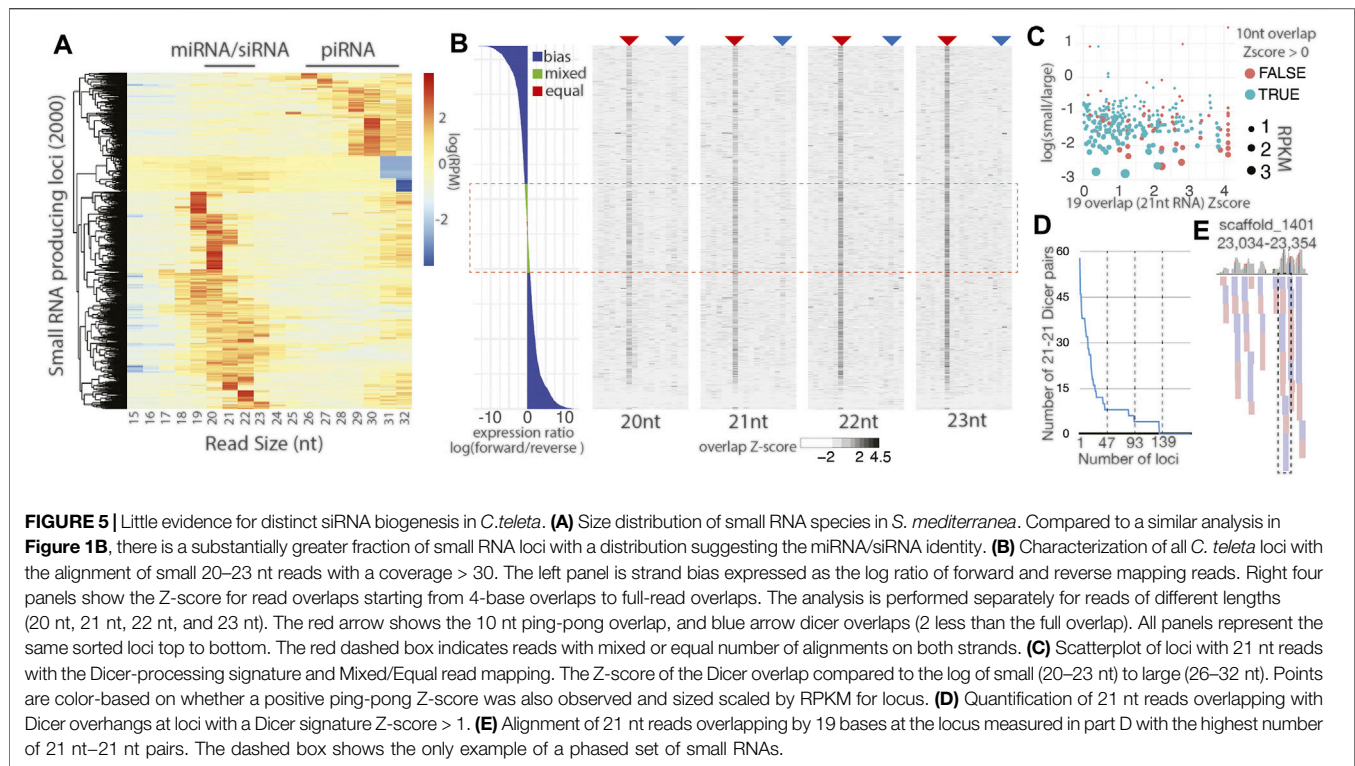


19–23 nt size range and are likely siRNAs or miRNAs. Only a quarter of the small RNA loci were 29–30 nt piRNAs, a striking difference to *C. teleta*.

Next, we sought small RNA loci with Dicer-processing signatures based on the presence of reads with 2 nt overhangs that range from 20 to 23 nt long (Mondal et al., 2020). These alignments were intersected with highly expressed loci (Figures 1B, 5A). Comparison of the 500 loci with the highest small RNA expression by size and Dicer overhang mapping revealed a substantially different distribution of loci when compared to *S. mediterranea* (Supplementary Figure S6A). *C. teleta* loci were substantially shorter than some examples in *S. mediterranea* that were longer than 40 kb. In *C. teleta*, the 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 (Supplementary Figures S6B, S6C). In the locus from *C. teleta*, a couple isolated mappings were seen whereas the *S. mediterranea* locus showed 21 nt abundant reads that overlapped with 2 nt 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 100s of such endogenous siRNA-loci that correspond to *cis*-NAT transcripts and transposable elements, as well as cryptic dsRNA-producing loci (Okamura and Lai, 2008). Due to the incomplete nature of *C.*

*teleta* gene annotations, only 122 instances of overlapping gene annotations are reported, and among those, only 22 are overlaps greater than 40 bases (Supplementary Figure 6D). Alignments to the overlapping gene regions are limited with very few showing alignments of small RNAs in the size range of Dicer products (20–23 nt). Even if these reads are present, they do not coincide with opposite strand alignments that would be expected from the Dicer cleavage (Supplementary Figure 6E). To circumvent this limitation, we identified all regions of the *C. teleta* genome with a mapping of 20–23 nt reads and greater than 30-read coverage that did not correspond to miRNAs (Figure 5B). For each of the 5,191 loci, the ratio of forward to reverse reads was calculated and binned into three categories: “bias” where there was 2-fold greater alignments on one strand, “mixed” where the ratio fell between 1.2 and 1.9 times greater on one strand, and “Equal” with the ratio being 1–1.1 (Zhang and Ruvkun, 2012). The alignment of 20–23 nt reads, at these loci, were simultaneously assessed for read overlaps such as 10 nt overlaps seen for ping-pong processing and 2 nt less than full overlaps associated with the Dicer cleavage. Loci were sorted by the log strand bias ratio and visualized as a bar plot for bias and heatmaps for read overlap Z-scores (Figure 5B). Combined mixed and equal loci were roughly a quarter (24%) of the loci. In the overlap heatmaps, a strong ping-pong signature was observed, suggesting that these 20–23 nt reads are predominantly piRNA-type possibly truncated by trimming. Indeed, ~60% of the loci overlap with cluster 4 from Figure 1B. Simultaneously, a very minor signal was observable at the Dicer overlap position,



particularly for 21 nt reads (**Figure 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 the 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 21 nt reads (19 nt overlaps) (**Figure 5C**). For this subset, we sought to understand the identity and expression of small RNAs by comparing RPKM values, the presence of the ping-pong signature, and the relative abundance of short (20–23 nt) reads, relative to long (26–32 nt) reads. Nearly all loci had greater alignment of longer reads than shorter 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, the expression per locus tapered significantly away from those with a greater portion of piRNA-sized read alignments. Next, we quantified the number of 21 nt read pairs that overlap by 19 nt. Calculation of Z-scores in **Figure 5B** reports overlap biases of the indicated read size paired with a read of any other size; thus, we sought to quantify 21 nt reads paired with other 21 nts (**Figure 5D**). From this, we found that over 75% of loci in **Figure 5C** have less than 10 unique 21 nt 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 21–21 nt pairs overlapping by 19 nt. Nearly all pair alignments were isolated, suggesting distributive processing (**Figure 5E**; **Supplementary Figure S7**). Indeed, of the 47 loci examined, only six had phased reads that would occur from processive cleavage. Together, there is

little evidence for siRNAs processed from long dsRNAs. Instead, our results suggest the *C. teleta* Dicer may be involved with the 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 (**Supplementary Figure S8**). A key domain involved in the processive Dicer activity is the N-terminal helicase domain that hydrolyzes ATP as a part of substrate engagement (Sinha et al., 2018). 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 a processive behavior. *C. teleta* Ago possesses 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 the absence of siAgo proteins and was borne out by lack of compelling signatures of Dicer processing. Similar configurations of RNAi pathway components are found in other lophotrochozoans, such as gastropods, cephalopods, and brachiopods, and leeches (Jehn et al., 2018). The exception might be bivalves where there are two Ago proteins; however, they seem to be the 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 antiviral defense (Gammon and Mello, 2015). 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 (Waldron et al., 2018). 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 antiviral defense has evolved independently.

In addition to the loss of siRNAs, we observed 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 the Dicer enzymatic activity. Interestingly, these hairpins would be excellent candidates for an 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 process 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 observed some unexpected features of *C. teleta* piRNAs. First, we did not observe a massive piRNA cluster which is observed in a variety of animals. Instead, we found loci that reached a maximum of 10kb, none of which exhibited 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 (Gebert et al., 2021). Thus, in *C. teleta*, even though there is no large piRNA clusters, 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 did not observe that the piRNA expression correlated with gonads. In this worm, we found a greater expression in the early embryos. Thus, germline defense may occur in post-fertilization vs. during gametogenesis, and the collection of inherited piRNAs peaks at a different point in this animal's life cycle.

In this species, we also observed 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 these types of loci are found in related species (Huang et al., 2019). 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* (Mohn et al., 2015b). Alternately, these piRNAs may be like *C. elegans* piRNAs produced by discrete transcriptional units.

Similarly, 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 be unsurprising that dsRNA molecules have become a pathogen pattern and could lead to an antiviral response instead of a 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 observed may provide additional configurations for exploiting the miRNA pathway. Finally, it may also be possible to exploit piRNA mechanisms by introducing synthetic RNAs that possess complementarity to known piRNAs (Mondal et al., 2020). This species has abundant piRNAs which would allow a variety of cell types to be targeted. In addition, there are two mechanisms of action (phasing and ping-pong amplification) that could be modeled for gene silencing technology.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/sra/PRJNA777269>.

## AUTHOR CONTRIBUTIONS

SK isolated RNAs, managed sequencing, performed analysis, and wrote the manuscript. BS performed analysis and equally contributed to writing. JP provided computational methods. AF managed the project and oversaw analysis and writing.

## FUNDING

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 the grant number P20GM103476. Computational resources were provided by NSF: ACI 1626217.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.802814/full#supplementary-material>



## REFERENCES

- Agarwal, V., Bell, G. W., Nam, J.-W., and Bartel, D. P. (2015). Predicting Effective microRNA Target Sites in Mammalian mRNAs. *eLife* 4, e05005. doi:10.7554/eLife.05005
- Antoniewski, C. (2014). "Computing siRNA and piRNA Overlap Signatures," in *Animal Endo-SiRNAs: Methods and Protocols*. Editor A. Werner (New York, NY: Springer New York), 135–146. doi:10.1007/978-1-4939-0931-5\_12
- Bartel, D. P. (2018). Metazoan MicroRNAs. *Cell* 173 (1), 20–51. doi:10.1016/j.cell.2018.03.006
- Berezikov, E., and Plasterk, R. H. A. (2005). Camels and Zebrafish, Viruses and Cancer: A microRNA Update. *Hum. Mol. Genet.* 14 (Suppl. 1\_2), R183–R190. doi:10.1093/hmg/ddi271
- Bhambri, A., Dhaunta, N., Patel, S. S., Hardikar, M., Bhatt, A., Srikakulam, N., et al. (2018). Large Scale Changes in the Transcriptome of *Eisenia F* during Regeneration. *PLOS ONE* 13 (9), e0204234. doi:10.1371/journal.pone.0204234
- Billi, A. C., Freeberg, M. A., Day, A. M., Chun, S. Y., Khivansara, V., and Kim, J. K. (2013). A Conserved Upstream Motif Orchestrates Autonomous, Germline-Enriched Expression of *Caenorhabditis E* piRNAs. *Plos Genet.* 9 (3), e1003392. doi:10.1371/journal.pgen.1003392
- Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., et al. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. *Cell* 128 (6), 1089–1103. doi:10.1016/j.cell.2007.01.043
- Cao, Z., Rosenkranz, D., Wu, S., Liu, H., Pang, Q., Zhang, X., et al. (2020). Different Classes of Small RNAs Are Essential for Head Regeneration in the Planarian *Dugesia Japonica*. *BMC Genomics* 21 (1), 876. doi:10.1186/s12864-020-07234-1
- Cifuentes, D., Xue, H., Taylor, D. W., Patnode, H., Mishima, Y., Cheloufi, S., et al. (2010). A Novel miRNA Processing Pathway Independent of Dicer Requires Argonaute2 Catalytic Activity. *Science* 328 (5986), 1694–1698. doi:10.1126/science.1190809
- Doench, J. G., Petersen, C. P., and Sharp, P. A. (2003). siRNAs Can Function as miRNAs. *Genes Dev.* 17 (4), 438–442. doi:10.1101/gad.1064703
- Fabry, M. H., Falconio, F. A., Joud, F., Lythgoe, E. K., Czech, B., and Hannon, G. J. (2021). Maternally Inherited piRNAs Direct Transient Heterochromatin Formation at Active Transposons during Early *Drosophila* Embryogenesis. *eLife* 10, e68573. doi:10.7554/eLife.68573
- Flynt, A. S. (2021). Insecticidal RNA Interference, Thinking Beyond Long dsRNA. *Pest Manag. Sci.* 77 (5), 2179–2187. doi:10.1002/ps.6147
- Flynt, A. S., Li, N., Thatcher, E. J., Solnica-Krezel, L., and Patton, J. G. (2007). Zebrafish miR-214 Modulates Hedgehog Signaling to Specify Muscle Cell Fate. *Nat. Genet.* 39 (2), 259–263. doi:10.1038/ng1953
- Friedländer, M. R., Adamidi, C., Han, T., Lebedeva, S., Isenbarger, T. A., Hirst, M., et al. (2009). High-Resolution Profiling and Discovery of Planarian Small RNAs. *Proc. Natl. Acad. Sci. United States America* 106 (28), 11546–11551. doi:10.1073/pnas.0905222106
- Friedländer, M. R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., et al. (2008). Discovering microRNAs from Deep Sequencing Data Using miRDeep. *Nat. Biotechnol.* 26 (4), 407–415. doi:10.1038/nbt1394
- Fromm, B., Høye, E., Domanska, D., Zhong, X., Aparicio-Puerta, E., Ovchinnikov, V., et al. (2021). MirGeneDB 2.1: Toward a Complete Sampling of All Major Animal Phyla. *Nucleic Acids Res.* 50, D204–D210. doi:10.1093/nar/gkab1101
- Gammon, D. B., and Mello, C. C. (2015). RNA Interference-Mediated Antiviral Defense in Insects. *Curr. Opin. Insect Sci.* 8, 111–120. doi:10.1016/j.cois.2015.01.006
- Gebert, D., Neubert, L. K., Lloyd, C., Gui, J., Lehmann, R., and Teixeira, F. K. (2021). Large *Drosophila* Germline piRNA Clusters Are Evolutionarily Labile and Dispensable for Transposon Regulation. *Mol. Cell* 81 (19), 3965–3978. doi:10.1016/j.molcel.2021.07.011
- Giani, V. C., Yamaguchi, E., Boyle, M. J., and Seaver, E. C. (2011). Somatic and Germline Expression of Piwi during Development and Regeneration in the Marine Polychaete Annelid *Capitella Teleta*. *EvoDevo* 2 (1), 10. doi:10.1186/2041-9139-2-10
- Goriaux, C., Théron, E., Brasset, E., and Vauray, C. (2014). History of the Discovery of a Master Locus Producing piRNAs: the Flamenco/COM Locus in *Drosophila M.* *Front. Genet.* 5, 257. doi:10.3389/fgene.2014.00257
- Han, B. W., Wang, W., Zamore, P. D., and Weng, Z. (2015). 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), 593–595. doi:10.1093/bioinformatics/btu647
- Huang, S., Ichikawa, Y., Igarashi, Y., Yoshitake, K., Kinoshita, S., Omori, F., et al. (2019). Piwi-interacting RNA (piRNA) Expression Patterns in Pearl Oyster (*Pinctada Fucata*) Somatic Tissues. *Sci. Rep.* 9 (1), 247. doi:10.1038/s41598-018-36726-0
- Jehn, J., Gebert, D., Pipilescu, F., Stern, S., Kiefer, J. S. T., Hewel, C., et al. (2018). PIWI Genes and piRNAs Are Ubiquitously Expressed in Mollusks and Show Patterns of Lineage-Specific Adaptation. *Commun. Biol.* 1 (1), 137. doi:10.1038/s42003-018-0141-4
- Kang, W., Eldfjell, Y., Fromm, B., Estivill, X., Biryukova, I., and Friedländer, M. R. (2018). miRTrace Reveals the Organismal Origins of microRNA Sequencing Data. *Genome Biol.* 19 (1), 213. doi:10.1186/s13059-018-1588-9
- Kim, V. N., Han, J., and Siomi, M. C. (2009). Biogenesis of Small RNAs in Animals. *Nat. Rev. Mol. Cell Biol.* 10 (2), 126–139. doi:10.1038/nrm2632
- Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2018). miRBase: from microRNA Sequences to Function. *Nucleic Acids Res.* 47 (D1), D155–D162. doi:10.1093/nar/gky1141
- Lee, Y. S., Nakahara, K., Pham, J. W., Kim, K., He, Z., Sontheimer, E. J., et al. (2004). Distinct Roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA Silencing Pathways. *Cell* 117 (1), 69–81. doi:10.1016/s0092-8674(04)00261-2
- Lewis, S. H., Quarles, K. A., Yang, Y., Tanguy, M., Frézal, L., Smith, S. A., et al. (2018). Pan-arthropod Analysis Reveals Somatic piRNAs as an Ancestral Defence against Transposable Elements. *Nat. Ecol. Evol.* 2 (1), 174–181. doi:10.1038/s41559-017-0403-4
- Liu, Y., Zhu, Z., Ho, I. H. T., Shi, Y., Li, J., Wang, X., et al. (2020). Genetic Deletion of miR-430 Disrupts Maternal-Zygotic Transition and Embryonic Body Plan. *Front. Genet.* 11, 853. doi:10.3389/fgene.2020.00853
- Mohammed, J., Bortolamiol-Becet, D., Flynt, A. S., Gronau, I., Siepel, A., and Lai, E. C. (2014). Adaptive Evolution of Testis-Specific, Recently Evolved, Clustered miRNAs in *Drosophila*. *RNA (New York, N.Y.)* 20 (8), 1195–1209. doi:10.1261/rna.044644.114
- Mohammed, J., Flynt, A. S., Panzarino, A. M., Mondal, M. M. H., DeCruz, M., Siepel, A., et al. (2018). Deep Experimental Profiling of microRNA Diversity, Deployment, and Evolution across the *Drosophila* Genus. *Genome Res.* 28 (1), 52–65. doi:10.1101/gr.226068.117
- Mohn, F., Handler, D., and Brennecke, J. (2015). piRNA-Guided Slicing Specifies Transcripts for Zucchini-dependent, Phased piRNA Biogenesis. *Science* 348 (6236), 812–817. doi:10.1126/science.aaa1039
- Mohn, F., Handler, D., and Brennecke, J. (2015). Noncoding RNA. piRNA-Guided Slicing Specifies Transcripts for Zucchini-Dependent, Phased piRNA Biogenesis. *Science* 348 (6236), 812–817. doi:10.1126/science.aaa1039
- Mondal, M., Brown, J. K., and Flynt, A. (2020). Exploiting Somatic piRNAs in *Bemisia Tabaci* Enables Novel Gene Silencing through RNA Feeding. *Life Sci. Alliance* 3 (e), e202000731. doi:10.26508/lsa.202000731
- Mondal, M., Klimov, P., and Flynt, A. S. (2018). Rewired RNAi-Mediated Genome Surveillance in House Dust Mites. *Plos Genet.* 14 (1), e1007183. doi:10.1371/journal.pgen.1007183
- Müller, M., Fazi, F., and Ciaudo, C. (2020). Argonaute Proteins: From Structure to Function in Development and Pathological Cell Fate Determination. *Front. Cell Dev. Biol.* 7, 360. doi:10.3389/fcell.2019.00360
- Okamura, K., and Lai, E. C. (2008). Endogenous Small Interfering RNAs in Animals. *Nat. Rev. Mol. Cell Biol.* 9 (9), 673–678. doi:10.1038/nrm2479
- Palmer, W. H., Hadfield, J. D., and Obbard, D. J. (2018). RNA-Interference Pathways Display High Rates of Adaptive Protein Evolution in Multiple Invertebrates. *Genetics* 208 (4), 1585–1599. doi:10.1534/genetics.117.300567
- Pinhal, D., Bovolenta, L. A., Moxon, S., Oliveira, A. C., Nachtigall, P. G., Acencio, M. L., et al. (2018). Genome-wide microRNA Screening in Nile tilapia Reveals Pervasive isomiRs' Transcription, Sex-Biased Arm Switching and Increasing Complexity of Expression throughout Development. *Sci. Rep.* 8 (1), 8248. doi:10.1038/s41598-018-26607-x
- Robine, N., Lau, N. C., Balla, S., Jin, Z., Okamura, K., Kuramochi-Miyagawa, S., et al. (2009). A Broadly Conserved Pathway Generates 3'UTR-Directed Primary piRNAs. *Curr. Biol.* 19 (24), 2066–2076. doi:10.1016/j.cub.2009.11.064
- Ross, R. J., Weiner, M. M., and Lin, H. (2014). PIWI Proteins and PIWI-Interacting RNAs in the Soma. *Nature* 505 (7483), 353–359. doi:10.1038/nature12987



- Rouhana, L., Weiss, J. A., Forsthoefel, D. J., Lee, H., King, R. S., Inoue, T., et al. (2013). RNA Interference by Feeding In Vitro-Synthesized Double-Stranded RNA to Planarians: Methodology and Dynamics. *Dev. Dyn.* 242 (6), 718–730. doi:10.1002/dvdy.23950
- Sinha, N. K., Iwasa, J., Shen, P. S., and Bass, B. L. (2018). Dicer Uses Distinct Modules for Recognizing dsRNA Termini. *Science* 359 (6373), 329–334. doi:10.1126/science.aag0921
- Waldron, F. M., Stone, G. N., and Obbard, D. J. (2018). Metagenomic Sequencing Suggests a Diversity of RNA Interference-Like Responses to Viruses across Multicellular Eukaryotes. *Plos Genet.* 14 (7), e1007533. doi:10.1371/journal.pgen.1007533
- Weick, E.-M., and Miska, E. A. (2014). piRNAs: From Biogenesis to Function. *Development* 141 (18), 3458–3471. doi:10.1242/dev.094037
- Wienholds, E., and Plasterk, R. H. A. (2005). MicroRNA Function in Animal Development. *FEBS Lett.* 579 (26), 5911–5922. doi:10.1016/j.febslet.2005.07.070
- Wynant, N., Santos, D., and Vanden Broeck, J. (2017). The Evolution of Animal Argonautes: Evidence for the Absence of Antiviral AGO Argonautes in Vertebrates. *Sci. Rep.* 7 (1), 9230. doi:10.1038/s41598-017-08043-5
- Yamanaka, S., Siomi, M. C., and Siomi, H. (2014). piRNA Clusters and Open Chromatin Structure. *Mobile DNA* 5 (1), 22. doi:10.1186/1759-8753-5-22
- Yang, J.-S., Maurin, T., Robine, N., Rasmussen, K. D., Jeffrey, K. L., Chandwani, R., et al. (2010). Conserved Vertebrate Mir-451 Provides a Platform for Dicer-independent, Ago2-Mediated microRNA Biogenesis. *Proc. Natl. Acad. Sci.* 107 (34), 15163–15168. doi:10.1073/pnas.1006432107
- Zamore, P. D. (2010). Somatic piRNA Biogenesis. *Embo J.* 29 (19), 3219–3221. doi:10.1038/emboj.2010.232
- Zhang, C., and Ruvkun, G. (2012). New Insights into siRNA Amplification and RNAi. *RNA Biol.* 9 (8), 1045–1049. doi:10.4161/rna.21246

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Khanal, Zancanela, Peter and Flynt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.