The evolution of RNA interference among Metazoa.

Alessandro Formaggioni[a,1], Gianmarco Cavalli[a,1], Mayuko Hamada[b], Tatsuya Sakamoto[b], Federico Plazzi[a,*], Marco Passamonti[a]

[a] Department of Biological, Geological and Environmental Sciences, University of Bologna, Italy

[b] Ushimado Marine Institute, Okayama University, Japan.

[1] These authors contributed equally to this work.

[*] Corresponding author: federico.plazzi@unibo.it

Keywords

#Metazoa #Lophotrochozoa #Spiralia #Argonaute #DICER #miRNA #piRNA #endo-siRNA #small RNA #RNA interference #Phylogenesis

Abstract

In Animals, three main RNA interference mechanisms have been described so far, which respectively maturate three types of small noncoding RNAs (sncRNAs): miRNAs, piRNAs and endo-siRNAs. The diversification of these mechanisms is deeply linked with the evolution of the Argonaute gene superfamily since each type of sncRNA is loaded by a specific Argonaute homolog protein. Moreover, other protein families play pivotal roles in the maturation of sncRNAs, like the DICER ribonuclease family, whose DICER1 and DICER2 paralogs maturate respectively miRNAs and endo-siRNAs. Among Metazoa, the distribution of these families has been only studied in major groups, and there are very few data for clades like Lophotrochozoa. Thus, we

here inferred the evolutionary history of the animal Argonaute and DICER families including 43 lophotrochozoan species. Phylogenetic analyses along with newly sequenced sncRNA libraries depicted a loss of the endo-siRNA pathway along the Lophotrochozoa evolution, with the absence of DICER2 in Nematoda and Polyzoa, and with the absence of DICER2 and the Argonaute homolog in the rest of Trochozoa phyla. On the contrary, early diverging phyla, Platyhelminthes and Syndermata, showed a complete endo-siRNA pathway. On the other hand, miRNAs were revealed the most conserved and ubiquitous mechanism of the metazoan RNA interference machinery, confirming their pivotal role in animal cell regulation.

Significance

The RNA interference (RNAi) pathway is one of the most important regulatory systems and it has been pointed out as a source of many evolutionary novelties. Lophotrochozoa comprehends a notable fraction of the animals' body plans, but an analysis of the main animal RNAi mechanisms has never been attempted. In this study we were able to depict novel patterns in the evolution of RNAi in Lophotrochozoan animals through the phylogenetic inference of Argonaute and DICER families and the analysis and sequencing of new small RNA libraries. These results strengthen the significance of the miRNA pathway among Metazoa and give support to some highly debated phylogenetic hypotheses within Lophotrochozoa.

Introduction

Argonaute proteins are cytoplasmic proteins that play a key role in most of the RNA interference (RNAi) pathways. They interact with a small non-coding RNA (sncRNA), forming an RNA-induced silencing complex (RISC). This ribonucleoprotein complex binds and silences target transcripts, using the complementary sncRNA as a probe (Iwakawa and Tomari 2022).Argonaute proteins can be found throughout most eukaryotic clades and share a common structure, featuring four domains: N-terminal (N), PIWI-Argonaute-Zwille (PAZ), Middle (MID) and P element-induced wimpy testis (PIWI) (Kuhn and Joshua-Tor 2013). The PIWI domain resembles the RNAse H domain's structure, but only some Argonautes have been reported to cleave the target

mRNA (Song et al. 2004). All the other Argonaute proteins repress the target trough proteins that interact with the RISC complex (Huntzinger and Izaurralde 2011; Wu et al. 2020).

Among the Argonaute superfamily, four different families have been characterized: *Trypanosoma*-AGO family, WAGO family, AGO-like family, and PIWI-like family (Swarts et al. 2014). *Trypanosoma*-AGO proteins have been identified only in the order Trypanosomatida (Garcia Silva et al. 2010). Conversely, WAGOs have been characterized in nematodes only, whereas AGO-like and PIWI-like proteins seem to be found in all animal phyla (Höck and Meister 2008; Swarts et al. 2014).

Three classes of metazoan interfering sncRNAs can be identified: micro-RNAs (miRNAs), small interfering RNAs (siRNAs), and piwi-interacting RNAs (piRNAs) (Iwakawa and Tomari 2022). Each class is matured by different pathways, is loaded by a different Argonaute protein and plays different cellular functions. Precursors of miRNAs (i.e., pri-miRNAs) are encoded by specific genes, transcribed by the RNA polymerase II (Bartel 2018; Y. Lee et al. 2004). Pri-miRNAs feature a hairpin secondary structure, with single-stranded ends at their 3' and 5' (Bartel 2018). These precursors undergo several maturation steps, starting immediately inside the nucleus, where they are targeted by the Microprocessor complex, which cleaves the overhanging nucleotides of pri-miRNAs, leaving a stem-loop structure with a 2bp offset, named pre-miRNA (Lee et al. 2003). Then, pre-miRNAs are exported in the cytoplasm, where they become substrate for DICER1, an endonuclease featuring one PAZ domain and two RNase III domains (Bernstein et al. 2001; Y. S. Lee et al. 2004). Through its catalytic activity, DICER1 removes the hairpin's loop, leaving a ~22bp ds-miRNA (Bartel 2018; Bernstein et al. 2001). Eventually, an AGO-like Argonaute binds to the ds-miRNA in the cytoplasm and disposes of one of the two strands, consequently resulting in a RISC complex. The miRNA-guided RISCs mostly target mRNAs by binding their 3'-UTR, interfering with their stability (Bartel 2018).

Unlike miRNAs, siRNAs may apparently be obtained from roughly every RNA capable of assuming a dsRNA structure (Shabalina and Koonin 2008). Therefore, siRNAs can originate from transcripts of transposons, repeated elements, or pseudogenes (Czech et al. 2008; Malone and Hannon 2009; Tam et al. 2008). Although some siRNAs have exogenous dsRNA precursors (e.g., viral RNAs), most of them have an endogenous origin, thus they are known as endo-siRNAs (Gammon and Mello 2015; Shabalina and Koonin 2008). Endo-siRNAs undergo a maturation pathway that is very similar to that of miRNAs, leading to the hypothesis that they evolved from a common ancestral RNAi system (Shabalina and Koonin 2008). Once the dsRNA precursors

reach the cytoplasm, they are processed by DICER2, a paralog of DICER1, but with a DExH helicase domain instead of a PAZ domain. DICER2 cleaves a ~21bp dsRNA that is loaded into an AGO-like Argonaute, called AGO2 in fruit flies (Czech et al. 2008; Y. S. Lee et al. 2004; Matranga and Zamore 2007). The resulting siRISC complex maintains one of the two strands, again using it as a probe (Matranga and Zamore 2007). The siRNA-mediated RNAi activity is not restricted to mRNAs, as siRNAs are known to partake in transposon silencing, DNA methylation, and antiviral defense (Galiana-Arnoux et al. 2006; Gammon and Mello 2015; Malone and Hannon 2009). A singular siRNA amplification pathway has been reported in nematodes, whose genomes encode RdRP enzymes, capable of using cleaved targets of the siRISC as a template to synthesize secondary siRNAs (Matranga and Zamore 2007).

Finally, piRNAs are 24-35 nt-long RNAs that originate from longer single strand precursors, consisting of either active transposons or transcripts of genomic piRNA clusters (Hirakata and Siomi 2016). Once these ssRNA precursors are exported through the nuclear pores, they undergo a rather complex maturation pathway, which takes place mostly in the perinuclear nuage (Hirakata and Siomi 2016; Weick and Miska 2014). Although piRNAs appear to be Metazoa-restricted (Grimson et al. 2008), biogenesis pathways vary significantly between clades (Weick and Miska 2014). Mature piRNAs interact with PIWI-like Argonautes, resulting in a piRISC that operates as a defense system against transposons, by means of its nuclease activity (Malone and Hannon 2009). Moreover, piRISCs have been linked to specific epigenetic modifications of the chromatin (i.e. H3K9me3; le Thomas et al., 2013). A peculiar piRNA amplification pathway called “ping-pong cycle” has been characterized in fruit flies, and later discovered in mice as well (Brennecke et al. 2007; Weick and Miska 2014). This amplification pattern is performed by two different PIWI-like Argonautes, namely one that binds sense-piRNAs (AGO3 in fruit flies) and one that loads antisense ones (Aubergine in fruit flies). As soon as one of these two proteins cleaves its target, it yields a secondary piRNA that can be loaded on the other PIWI-like protein, generating an amplification loop (Hirakata and Siomi 2016; Weick and Miska 2014). Another type of piRNA amplification is called “phasing”. PIWI loads a single strand RNA (pre-pre-piRNA) from the 5’ end, directing the endonucleolytic cleavage of the ribonuclease Zucchini at the 3’ end. This process is repeated along the pre-pre-piRNA leading to phased matured piRNAs(Mohn, Handler, and Brennecke 2015; Ozata et al. 2019).

The evolution and diversification of the Argonuate superfamily is deeply linked with the emergence of new RNAi pathways since each type of sncRNA is loaded by a different Argonaute homolog. miRNAs, piRNAs and endo-siRNAs were likely to be in the last metazoan common ancestor, since they have been described in Porifera, Cnidaria and most of the metazoan phyla (Calcino et al. 2018; Wheeler et al. 2009). In some clades piRNA and endo-siRNA pathways have been secondarily lost (Fontenla, Rinaldi, and Tort 2021; Wynant, Santos, and Vanden Broeck 2017), while the miRNA pathway is likely to be ubiquitous in Metazoa (Fromm et al. 2022).

RNAi mechanisms are by far less studied in Lophotrochozoa compared to Deuterostomia and Ecdysozoa. This superclade is comprised by several phyla, such as Platyhelminthes, Syndermata, Annelida, and Mollusca, showing an astonishing variability. For medical and nutritional reasons, most of the data are restricted to Mollusca and Platyhelminthes. Parasitic Platyhelminthes (i.e., Neodermata) lacks the PIWI pathway, but the miRNA and endo-siRNA pathways have been reported (Fontenla et al. 2017, 2021). The PIWI pathway has been confirmed in Mollusca, where it is visible a clear signature of ping-pong amplification (Jehn et al. 2018). Moreover, many miRNAs have been annotated in mollusks (Fromm et al. 2022). Outside those phyla, few data have been published. The endo-siRNA pathway seems absent in the Annelida *Capitella teleta* (Khanal et al. 2022) and the annotation of miRNA family is restricted to one Syndermata, one Brachiopoda and two Annelida species (Fromm et al. 2022).

During last years, the increase of -omics data have made possible to compare and study the evolution of protein families along Lophotrochozoa. In this study, we exploited various -omics resources from nine lophotrochozoan phyla to annotate and characterize the diversification of the Argonaute and DICER proteins. We also analyzed sncRNA libraries to annotate the three sncRNA types and confirm the presence or absence of a particular sncRNA type in some phyla. The pattern reported by our results has never been described in any eukaryotic clade. Along the Lophotrochozoa evolution the endo-siRNA pathway has been progressively lost, starting with DICER2 in all Trochozoa (i.e., Nemertea, Phoronida, Brachiopoda, Annelida and Mollusca) and Polyzoa (Entoprocta + Bryozoa), followed by the loss of the fruit fly AGO2-like proteins. This pattern is confirmed by the distribution of DICER2 and AGO2-like proteins in the analyzed organisms. On the contrary, the piRNA and miRNA pathways appeared to be conserved in almost all Lophotrochozoa.

Results

We annotated Argonaute proteins of 43 lophotrochozoan species analyzing 19 proteomes, 16 genomes and 8 transcriptomes. Argonaute sequences of *Homo sapiens* (Chordata), *Drosophila melanogaster* (Arthropoda), *Caenorhabditis elegans* (Nematoda) and *Nematostella vectensis* (Cnidaria) were retrieved from SwissProt and included in the phylogenetic analysis as references. Moreover, we include metazoan species whose RNAi pathways have been already studied (see Introduction), testing whether our Argonaute and DICER annotation matches results from literature. We chose proteomes of *Danio rerio* and *Gallus gallus* (Chordata), *Asterias rubens* (Echinodermata), *Anopheles gambiae* (Arthropoda), *Strongyloides ratti* (Nematoda) and the *Acropora muricata* (Cnidaria) genome assembly.

All the annotated Argonaute proteins were aligned, and the Maximum-Likelihood (ML) tree was inferred. The PIWI and AGO proteins of *Trypanosoma brucei* were obtained from Uniprot and used as outgroups. The phylogenetic tree of the Argonaute superfamily supported the known main families (Fig. 1; Fig. S1). The WAGO family, which is restricted to nematodes, included WAGOs and CSR-1 sequences from *C. elegans*. The PIWI-like family was characterized by AGO3, PIWI and AUB of *D. melanogaster*, HILI and HIWI of *H. sapiens* and PRG-1 of *C. elegans;* the AGO-like family comprised AGO1,2 of *D. melanogaster*, AGO1-4 of *H. sapiens* and ALG1,2 of *C. elegans.* Every family was widely supported by UFBoot and the SH-alrt test (Fig. 1).

The AGO-like family was divided in two groups. One monophyletic group (UFBoot = 99 and SH-alrt = 99.1) included *H. sapiens* AGO1-4, *D. melanogaster* and *N. vectensis* AGO1 and *C. elegans* ALG1,2: all these proteins interact with miRNAs. Almost all species featured a protein clustering within this clade, with very few exceptions, and at least one organism from every phylum was recovered in this clade (Fig. 1; Tab. S1). We will refer to this clade as the “miAGO clade”. The second group within the AGO-like family was polyphyletic. *D. melanogaster* AGO2, *N. vectensis* AGO2 and *C. elegans* ERGO1 clustered within this clade. All these proteins target endo-siRNAs, thus we will henceforth call this group the “siAGO group”. Few lophotrochozoan phyla are included in this group, since we annotated at least one siAGO protein for all the Platyhelminthes and Entoprocta species, while three other phyla featured at least one species within the group, namely Nemertea, Bryozoa and Syndermata. We did not retrieve any siAGO protein from the remaining clades, namely Mollusca, Annelida, Brachiopoda, Phoronida (Fig. 1; Tab. S1).

Similarly to the AGO-like family, protein clustering in the PIWI-like family can be divided in two groups: one is monophyletic and one is polyphyletic. The monophyletic group (UFBoot = 98 and SH-alrt = 99.4) comprised the *H. sapiens* HILI, the *N. vectensis* PIWI2 and the *D. melanogaster* AGO3; the polyphyletic group included *H. sapiens* HIWI, the *N. vectensis* PIWI1, the *C. elegans* PRG1 and the *D. melanogaster* AUB and PIWI. Each phylum was present in both groups with at least one species (Tab. S1). PIWI-like proteins were almost absent in Platyhelminthes, apart from *Schmitdea mediterranea* (Fig. 1).

Regarding the six non-lophotrochozoan species included in the analysis, the annotation of Argonautes proteins matched the expectations: miAGO proteins were retrieved for all six species; siAGO proteins were absent in *D. rerio, G. gallus* and *A. rubens* (Deuterostomia); all six species reported PIWI1 and PIWI2 proteins, exception made for *S. ratti* (i.e. that lacked of both PIWI proteins) and, unexpectedly, *G. gallus* that lacked of PIWI2 (Fig. 1; Tab S1).

The phylogenetic analysis highlighted the presence of miAGO proteins in each phylum, but only some of them feature siAGO proteins. In Arthropoda, the preliminary structures of siRNAs and miRNAs are processed by two distinct DICER paralogs: DICER2 and DICER1, respectively (Shabalina and Koonin 2008). DICER2 has been found in other phyla, like Cnidaria and Platyhelminthes (Mukherjee, Campos, and Kolaczkowski 2013a), while clades lacking siAGO proteins lack DICER2 as well. We annotated DICER proteins and inferred the phylogeny to understand whether Lophotrochozoa follow the same pattern.

We annotated DICER proteins querying the lophotrochozoan sequences against an annotated metazoan DICER set (Mukherjee et al. 2013a) and looking for the ribonucleases 3 domain. The phylogenetic analysis included the metazoan DICER set of Mukherjee and colleagues (2013) as references, including the *Zea mays* DICER proteins as outgroups (Fig. 2; Fig. S2). The resulting ML tree clustered the DICER proteins in two distinct groups. Recall the position of the reference sequences, DICER1 and DICER2 sequences are accordingly split in the two groups, apart from *Litopenaeus vannamei* DICER2, which is basal to all the other proteins. The monophyly of the DICER1 group is supported by the SH-alrt test (86.4) and the UFBoot (96). On the contrary, the low support values of the DICER2 node (UFBoot= 58 and SH-alrt = 64.3) undermine the hypothesis of a unique common origin of DICER2 proteins. Nonetheless, the presence of DICER2 was restricted to few lophotrochozoan phyla: all Platyhelminthes and three Syndermata species showed a DICER2 protein. On the other hand, DICER1 was annotated for every phylum: thus, Mollusca, Annelida, Phoronida, Brachiopoda,

Nemertea, Bryozoa and Entoprocta reported DICER1 proteins, but no DICER2 proteins (Fig. 2; Tab. S1). In line with the expectations, DICER2 was reported for *A. muricata, A. gambiae,* while it was lacking in *D. rerio, G. gallus, A. rubens* and *S. ratti.* On the contrary, DICER1 was annotated in all of them.

Overall, the DICER family phylogenetic analysis confirmed that the absence of siAGO proteins coincides with the absence of DICER2, and *vice versa*, exception made for Nemertea, Bryozoa and Entoprocta, where we annotated siAGO proteins for most species; however, none of these species featured DICER2. DICER1 and miAGO proteins showed lower evolutionary rates than their paralogous counterparts: the root-to-tip distances of miAGO branches were significantly lower than the root-to-tip distances of siAGO, PIWI1 and PIWI2 branches, and the root-to-tip distances of DICER1 branches were significantly lower than their DICER2 counterparts (Fig. S3). We also estimated the ratio of non-synonymous to synonymous substitution rates (ω) along the AGO family tree and the DICER family tree: we confirmed that, during the evolution of DICER and AGO family, purifying selection on DICER1 (LRT=34.34, p-value = 4.62×10^{-9}; Fig. S4B) and miAGO (LRT=204,64, p-value = 0; Fig. S4A) has intensified compared to the rest of the family tree (Wertheim et al. 2015).

In both phylogenetic analyses the presence of a protein in some phyla was not confirmed by all the species. To understand whether it might be related to the quality of the data, we evaluated the completeness of proteomes, genomes, and transcriptomes: Argonaute and DICER proteins were missing more commonly in transcriptomes than in proteomes or genome assemblies, regardless of completeness (Tab. S1). All proteomes showed a similar presence/absence pattern and high completeness values, while the completeness of the genomes varied significantly between species. In general, in more complete genome assemblies we were also able to annotate more proteins (Tab. S1). Overall, phyla that are generally more represented in protein databases (i.e., Mollusca, Annelida, Platyhelminthes) showed a more constant presence/absence pattern between species than under-represented phyla (i.e., Nemertea, Bryozoa, Syndermata). It is worth noting that the annotation of Argonaute and DICER was necessarily dependent on databases of orthologous genes (see Materials and methods).

According to the phylogenomic analysis, some lophotrochozoan phyla lack pivotal proteins related to the endo-siRNA pathway. DICER2 generally processes double stranded RNAs producing two 21 bases siRNA duplexes that overlaps of 19 bases. Similarly, piRNAs produced by the ping-pong cycle go in pairs that overlaps of 10 bases (Antoniewski 2014; Khanal et al. 2022). Thus, siRNAs and piRNAs have a unique signature that can be

spot in sncRNA libraries by looking for overlapping pairs of reads. We retrieved eight sncRNA libraries from lophotrochozoan and non-lophotrochozoan species (i.e., *Danio rerio, Apostichopus japonicus, Acropora muricata, Anopheles gambiae, Drosophila melanogaster, Schmitdea mediterranea, Schistosoma japonicum, Crassostrea gigas*) and we sequenced the sncRNA pool of *Notospermus geniculatus*, since no nemertean sncRNA library was available. Using the overlapping_reads.py script (Antoniewski 2014) we calculated the number of read pairs that overlapped for the same number of bases, from 4 to 20 bases. Then, we calculated the Z-Score among the number of read pairs for each overlap group. A Z-score equal to 1 means that the number of reads pairs in that overlap group is one standard deviation higher from the mean size of all the overlap groups of a given species. Taking into consideration only 21 bp reads (i.e., the expected length of endo-siRNAs), Arthropoda (*D. melanogaster* and *A. gambiae*), Cnidaria (*A. muricata*) and Platyhelminthes (*S. mediterranea* and *S. japonicum*) representatives reported a Z-score higher than 1 in the 19-overlap group, corresponding to the siRNA signature (Fig. 3). On the other hand, Deuterostomia (*D. rerio* and *Apostichopus japonicus*), Mollusca (*C. gigas*) and Nemertea (*N. geniculatus*) representatives showed a Z-score lower than 1. Some species, namely *N. geniculatus*, *A. japonicus*, *C. gigas*, and *S. mediterranea*, also reported a sharp increase of the Z-score for the 10-overlap group for 21-bp long reads only. A 10-bases overlap would correspond to the piRNA signature, although the length of piRNAs in Arthropoda ranges from 25 to 30 pb. Overall, considering the phylogenomic analysis, all the species that exhibited a complete siRNA pathway (using the presence of DICER2 and siAGO as a proxy) showed a high Z-score in the 19-overlap group (Fig. 3).

Finally, we investigated whether the range of action of the three sncRNA types is uniform among the three main Metazoa superclades, choosing tree representative for each one (i.e., *C. gigas* for Lophotrochozoa, *A. gambiae* for Ecdysozoa and *D. rerio* for Deuterostomia). We annotated miRNAs, piRNAs and endo-siRNAs evaluating their expression at different sncRNA lengths (Fig. S5A). In terms of reads per million (RPM), in all three species miRNAs are the most expressed and their length ranges from 20 – 25 nc. On the contrary, piRNAs were annotated in the in the 24 – 30 nc length range. In most of the cases, endo-siRNAs were not able to discern the endo-siRNA signal from noise, but *A. gambiae* is the species that showed highest expression levels (Fig. S4A). We also compared ovary and somatic tissue sncRNAs libraries. As expected, in all three species piRNAs is the class more expressed in the ovaries. (Fig. S5B). Overall, in the three species we did not

reported notable differences in terms of sncRNA length or differential expression in somatic/ovarian tissue among piRNA and miRNA types.

Discussion

RNAi pathways play a central role in many molecular aspects, from mRNA regulation to defense mechanisms, and Argonaute proteins are the RNAi key players in all eukaryotes. The expansion and differentiation of this superfamily have reflected the emergence of different RNAi mechanisms, each represented by a sncRNA class loaded by a specific Argonaute. All the phylogenetic analyses agree to divide the eukaryotic Argonaute superfamily in four main families, namely the *Trypanosoma*-AGO family, the WAGO family, the AGO family, and the PIWI family (Garcia Silva et al. 2010; Höck and Meister 2008; Swarts et al. 2014). Excluding the *Trypanosoma*-AGO family, all the other families are represented within Metazoa; moreover, the PIWI and the WAGO family are restricted to animals or even nematodes, respectively (Swarts et al. 2014). According to the distribution of Eukaryota in the four families (Swarts et al. 2014), it is likely that the Argonaute families differentiated at least 1.5 billions of years ago (Strassert et al. 2021)

The inferred Argonaute phylogenetic tree confirmed and highly supported the three metazoan Argonaute families (Fig. 1). Moreover, we could identify two distinct groups in the AGO and PIWI families that might suggest a duplication at the base of the protein family. However, this hypothesis would hold true if both groups were monophyletic, but our and other Argonaute phylogenetic analyses recursively confirmed the polyphyly of either groups in PIWI and AGO families (Swarts et al. 2014; Wynant et al. 2017). Recalling the deep divergence of these proteins, the signal might be saturated. Accordingly, most of the nodes at the base of the family are not strongly supported (Fig. S1). The same pattern has been observed in the DICER phylogeny, with DICER2 proteins grouped in a polyphyletic group while DICER1 proteins clustered in a well-supported monophyletic clade (Fig. 2; Mukherjee et al. 2013). Concordantly, DICER2 and DICER1 are related respectively to siAGO and miAGO proteins. Overall, within each clade the phylogenetic reconstruction is substantially in agreement with the state-of-art animal phylogeny, recalling that the signal has been inferred from single markers.

Almost all lophotrochozoan showed two distant related PIWI proteins (i.e., one AUB-like and one AGO3-like; Fig. 1; Tab. S1), with the only exception of Neodermata (Platyhelminthes), which lacks the whole piRNA pathway (Fontenla et al. 2021). It is likely that the ping-pong cycle, which has been already described in some mollusks (Jehn et al. 2018), has been maintained in most of lophotrochozoan. The piRNA expression of *C.*

gigas is in line with that of *D. rerio* and *A. gambiae* and the differential expression analysis confirms that piRNAs are more expressed in the gonads in all the three clades.

The miRNA pathway appears the most ubiquitous RNAi pathway among Metazoa, almost all species reported a miAGO and a DICER1 protein. For these proteins we even detected lower root-to-tip distances and a decrease of ω along their branches, which confirms a higher selective pressure. Therefore, the high conservation of proteins involved in the miRNA pathway reflects the well-known conservation of miRNAs among Metazoa (Tarver et al. 2013) . Even in our case, we confirmed that the annotated miRNAs showed similar features in the three reference species; they are 20-25 nt long, they are not more expressed in the ovaries compared to somatic tissue and overall they are by far the most expressed sncRNA class.

Conversely, the endo-siRNA pathway is by far less studied than the miRNA and piRNA ones, which are also more ubiquitous in Metazoa. The absence of siAGO proteins have been reported in Vertebrata (Wynant et al. 2017) and in the annelid *Capitella teleta* (Khanal et al. 2022), where the absence of a siRNA signature in sncRNA libraries have been also confirmed. Our analysis on Argonaute and DICER proteins revealed a unique pattern in the evolution of this pathway. DICER2 and siAGO proteins were retrieved in Platyhelminthes and Syndermata. The two proteins were completely absent in Mollusca, Annelida, Phoronida and Brachiopoda. A unique condition was recovered in Nemertea, Bryozoa and Entoprocta; they are equipped with siAGO proteins (albeit not every species), but DICER2 proteins were not annotated in any of those. Thus, we speculate that Nematoda, Bryozoa and Entoprocta are an intermediate state during the loss of the endo-siRNA pathway in Lophotrochozoa. The phylogenetic relationships between lophotrochozoan phyla are not completely resolved yet. However, a phylotranscriptomic analysis has shown that the best ML tree, inferred by reducing the branch-length heterogeneity, places Nemertea in sister relationship with Mollusca + Phoronida + Annelida + Brachiopoda (Kocot et al. 2016), the group of phyla totally lacking the endo-siRNA pathway. The same tree places Bryozoa and Entoprocta as the sister group of Trochozoa (Nemertea + Mollusca + Phoronida + Annelida + Brachiopoda) and the Polyzoa hypothesis (i.e., monophyly of Bryozoa+Entoprocta) has been recently supported (Khalturin et al. 2022). Therefore, the characterization of endo-siRNA pathway in Nemertea is of paramount importance to achieve a complete figure of the evolution of this pathway. Overall, although the relationships between lophotrochozoan phyla are still uncertain, in Nemertea and Polyzoa the presence of siAGO and absence of DICER2 proteins may have a phylogenetic meaning. We hypothesize that the loss of

DICER2 took place at the base of Polyzoa+Trochozoa; subsequently, nemerteans retained siAGO proteins that were left in the last common ancestor of remaining Trochozoa (Fig. 4).

Indeed, species lacking DICER1 showed less sncRNA pairs overlapping for 19 bp, which is the siRNA signature, compared to species with a complete endo-siRNA pathway. The absence of a siRNA signature in the nemertean *N. geniculatus* is in line with our scenario since the typical signature is due to the DICER2 cleavage on double-stranded RNAs. To our knowledge, an uncomplete endo-siRNA pathway has never been described, at least among Metazoa. Thus, it is difficult to say whether in Nemertea and Polyzoa the siAGO proteins have been coopted to load a different sncRNA type or they have lost their function in the RNAi pathway after the loss of DICER2. Although the miRNA pathway is highly conserved among Metazoa, many other non-canonical pathways have been discovered, which bypass the Microprocessor or the DICER1 step, but are still loaded by miAGOs (Havens et al. 2012). In the same way, nemerteans and polyzoans siAGO proteins might be involved in pathways that do not require the action of DICER2. Alternatively, DICER2 might have been coopted for completely different functions.

Concluding, the loss of siAGO proteins and, more in general, of the endo-siRNA pathway possibly in all Trochozoa have been strongly supported by the joint phylogenetic analysis and analysis of sncRNA libraries. The situation is comparable with that obtained from Deuterostomia, where the absence of the canonical (i.e., involving both DICER and siAGOs) endo-siRNA pathway have already been reported using the lack of annotated siAGO proteins as a proxy (Wynant et al. 2017). Our analysis support this hypothesis, since neither siAGO nor DICER2 homologs were retrieved in deuterostomes (Fig. 1,2). The distribution of the endo-siRNA pathway in Cnidaria (Calcino et al. 2018), Arthropoda and basal Lophotrochozoa (i.e., Platyhelminthes and Syndermata) suggests that most trochozoans and deuterostomes lost it secondarily. That said, in mammals it has been demonstrated that endogenous dsRNAs can still be processed and included in the RNAi regulation by proteins of the miRNA pathway (Babiarz et al. 2008; Laura, Srividya, and Vidigal 2020; Watanabe et al. 2008). Thus, a different endo-siRNA pathway seems to have evolved from other pathways. In the same way, it is possible that in most Trochozoa endo-siRNAs might not be completely absent, rather processed by different proteins: however, due to the lack of DICER2, they would not display the typical siRNA signature (Park, Kang, and Shin 2021). As for mammals, immunoprecipitation or knockout experiments in Trochozoa might elucidate whether Argonaute proteins can load other sncRNAs types beside piRNAs and miRNAs.

Through the present work we were able to depict the evolution of the main RNAi mechanisms in Metazoa. Unique patterns emerged, such as the uncomplete endo-siRNA pathway in Nemertea and Polyzoa. More efforts should be placed in studying RNAi in less considered phyla, in order to better characterize understudied patterns, casting light on the diversity of RNAi across Metazoa. Comparing different RNAi pathways allowed us to highlight the evolutionary constraints on the miRNA pathway among Metazoa. The conservation and ubiquity of this pathway is likely due to its importance in many regulative systems, and further research should focus on detecting the key features that made the miRNA pathway the main RNAi system.

Materials and Methods

Supporting methods for annotation of Argonaute and DICER proteins and phylogenetic inference

Argonaute and DICER proteins were annotated for a wide range of omics-data. We first analyzed all the lophotrochozoan proteomes which were annotated following the NCBI Eukaryotic Genome Annotation Pipeline (Thibaud-Nissen et al. 2016). To increase the sampling in underrepresented clades, we selected 17 assemblies (Table S1) and we predicted gene models with the BRAKER2 automated pipeline (Brůna et al. 2021; Hoff et al. 2019; Stanke et al. 2008). To improve the DICER and Argonaute model prediction, the Metazoa OrthoDB database provided by BRAKER2 was enriched of Argonaute and DICER sequences annotated from the lophotrochozoan NCBI proteomes. We collapsed all the isoforms, keeping the longer ones, using the perl script agat_sp_keep_longest_isoform.pl (Dainat n.d.). Lastly, we searched Argonaute and DICER proteins in transcriptomes when phyla showed few assemblies (Table S1). We trimmed the reads with Trimmomatic-0.39(Bolger, Lohse, and Usadel 2014) with the following settings: ILLUMINACLIP: TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:75 ; we assembled the transcriptome with Trinity v2.1.1 with default settings (Grabherr et al. 2011); we filtered out contaminants locally aligning the transcripts against the non-redundant protein database (Sayers et al. 2022) with DIAMOND blastp (Buchfink, Xie, and Huson 2015; Camacho et al. 2009) and discarding all the transcripts with a non-metazoan best hit. Coding regions were predicted with Transdecoder v5.5.0 (https://github.com/TransDecoder/TransDecoder), scanning all ORFs for homology using DIAMOND blastp and HMMER (Mistry et al. 2013). Overall, we obtained the coding sequences of 42 lophotrochozoan species and six other metazoan species. Coding sequences were translated into amino acid sequences; then Argonaute and DICER proteins were annotated as follow. We looked for Argonaute proteins by annotating the conserved PIWI and PAZ domains in all 48 species. Domain alignments were retrieved from Pfam (PIWI accession: pfam02171; PAZ accession: pfam02170) (Mistry et al. 2021). Using HMMER, we built a profile for each multiple sequence alignment and we searched the profiles against each (--e-value 10e-6). Only proteins with both domains annotated were considered Argonaute proteins and kept for the phylogenetic analysis.

To annotate DICER proteins we aligned the amino acid sequences against the bilaterian annotated set of Mukherjee and colleagues (Mukherjee, Campos, and Kolaczkowski 2013b) using blastp. In a second round of

filtering, we retrieved the ribonucleases 3 domain alignment from Pfam (accession: pf14622.9), which is the only domain shared between all the metazoan DICERs (Mukherjee et al. 2013b); we searched with HMMER (--e-value 10xe-6) and selected only sequences containing the ribonucleases 3 domain. However, orthologs of the endoribonuclease DROSHA were possibly included among the annotated DICER at this stage. Therefore, we downloaded the DROSHA orthologs from OrthoDB (reference: 9211at3208) (Zdobnov et al. 2021) and we built a custom dataset with both DICER from Mukherjee and colleagues (2013) and DROSHA sequences. We locally aligned the set of putatively annotated DICER proteins against this dataset. We retained proteins whose best five hits were all with DICER orthologs, whereas we discarded proteins with only DROSHA orthologs among the best five hits. No ambiguous results (i.e., proteins showing both DICER and DROSHA within the best five hits) were obtained.

The Argonaute and DICER phylogenetic trees were inferred from datasets being comprised of all the annotated sequences from proteomes, genomes and transcriptomes, with the addition of reference sequences chosen from SwissProt (Fig. 1) or the bilaterian annotated set for the DICER dataset. Datasets were aligned with MAFFT v7.508 (Katoh and Standley 2013), using the options --maxiterate 1000 --localpair. Uninformative columns were masked from the alignments using Gblocks (Castresana 2000), setting -b2= (3 × number of sequences)/5 -b3=10 -b4=5 -b5=a. The ML trees were inferred with IQ-Tree(Nguyen et al. 2015) with the predefined protein mixture model LG+C20+R4. To assess the robustness of the clades we calculated the ultrafast bootstrap approximation (UFBoot) with 1,000 bootstrap replicates(Hoang et al. 2018) and the SH-like approximate likelihood ratio test with 1,000 replicates (Guindon et al. 2010).

We tested whether DICER1 and miAGO proteins have experienced an intensified selection with HyPhy RELAX (Wertheim et al. 2015). We tagged all the branches belonging to DICER1 and miAGO clades as foreground. All the other branches belonging to DICER or AGO family clades were tagged as background. For some proteins we were not able to retrieve the respective coding sequence, namely the *Saccostrea glomerata* miAGO and DICER1, *Anopheles gambiae* DICER2, *Trobolium castaneum* and *Brugia malayi* DICER1. Those proteins were removed during the selection analysis.

The completeness of proteomes, assembly and transcriptomes was evaluated with BUSCO v. 5.4.3 (Simão et al. 2015) using the Metazoa dataset.

***Notospermus geniculatus* sncRNA libraries sequencing and analysis of sncRNA libraries**

Six specimens (three males and three females) of the nemertean Notospermus geniculatus were sampled in June 2018 near Ushimado (Okayama prefecture, Japan). Animals were left in seawater and 7% MgCl2·H2O (1:1 ratio) for 15'; gonads were then dissected in 7% $MgCl_2 \cdot H_2O$ on ice and stored in RNAlater (Thermo Fisher Scientific Inc., Waltham, USA), following manufacturer's instructions. Total RNA was extracted using a standard chloforom:TRI Reagent® (Merck KGaA, Darmstadt, Germany) protocol, following manufacturer's instructions. The TruSeq Small RNA library kit (Illumina, San Diego, USA) was used to prepare six sRNA libraries that were sequenced on an Illumina HiSeq2500 platform. Both library preparation and high-throughput sequencing were carried out at the Macrogen Inc. facility (Seoul, South Korea).

For this study we sequenced the sncRNA pool from six samples of *N. geniculatus.* These libraries were analyzed along with publicly available sncRNA libraries of other five species. The libraries were selected and downloaded from the Sequence Reads Archive (SRA) provided by NCBI (Table S2). When more samples from the same project were available, libraries were pulled together, obtaining a single fastq file for each species. Adapters and low-quality bases were removed from reads with Cutadapt v3.9.7(Martin 2011), setting the options -e 0.2 -O 5 --quality-cutoff 6 --discard-untrimmed. Trimmed reads were mapped on the reference genome with Bowtie(Langmead et al. 2009), allowing up to 100 multiple alignments (-m 100). We estimated the distribution of the overlaps between reads with the python script *overlapping_reads.py* (https://github.com/ARTbio/tools-artbio/blob/master/tools/small_rna_signatures/overlapping_reads.py; Antoniewski 2014).

For *Crassotrea gigas*, *Danio rerio* and *Anopheles gambiae* we annotated miRNAs, siRNAs and piRNAs. The tool miRDeep2 (Friedländer et al. 2012) was used to predict miRNAs providing the already annotated miRNA set of that species from MiRGeneDB (Fromm et al. 2022) or miRBase(Kozomara, Birgaoanu, and Griffiths-Jones 2019) and the annotated miRNAs of up to five closely related species. The novel predicted miRNAs were reconsidered following the rules established by Fromm and colleagues (2015) and discarding novel miRNAs with a STAR sequence coverage lower than 5 reads. Putative siRNA and piRNA pairs were predicted according to reads overlaps (i.e. siRNA pairs overlap of the read length – 2, piRNA pairs overlap of 10

nucleotides) calculated with *overlapping_reads.py*. Pairs of siRNAs and piRNAs were discarded when: one of the paired sRNAs had a coverage lower than 5; the logarithmic ratio of the pair was above 1.5; the pair mapped in a miRNA region. The differential expression of sncRNAs between somatic tissues and ovaries was tested with edgeR (Robinson, McCarthy, and Smyth 2010) using a generalized linear model and a quasi-likelihood F-test (Lund et al. 2012). We chose a strict P-Value threshold of 0.001 to consider a sRNA to be significantly differentially expressed.

Data availability

Raw reads were uploaded to Sequence Read Archive (SRA) under the BioProject PRJNA942081.

Authors contributions

A.F., F.P. and M.P. designed the research; M.H. and T.S. sampled and prepared the specimens; A.F. and M.C. analyzed the data and wrote the first draft of the article; F.P. and M.P. supervised the project. All authors accepted and contributed to the final version of the article.

Acknowledgements

This work was supported by the Canziani bequest funded to M.P.

Figures captions

Figure 1. ML tree of the lophotrochozoan Argonaute proteins. For the six marked nodes the label shows UFBoot/SH-alrt value. Clades formed by paralogs of the same species were collapsed and represented with a triangle. For reference proteins the Uniprot accession code is reported in brackets. Species are colored according to the phylum. The color legend on the bottom left reconstructs the main phylogenetic relationships according to the ML-tree of Kocot and colleagues (2016).

Figure 2. ML tree of the lophotrochozoan DICER proteins. For two marked nodes it is reported respectively the UFBoot and the SH-alrt value. Clades formed by paralogs of the same species were collapsed and represented with a triangle. References species retrieved from the analysis of Mukherjee and colleagues (2013)

are marked with an asterisk. Species are colored according to the phylum. The color legend on the bottom left reconstructs the main phylogenetic relationships according to the ML-tree of Kocot and colleagues (2016).

Figure 3. Evaluating the siRNA signature in sncRNA libraries. The plot reports the Z-score between the number of pairs of all possible overlaps. A Z-score greater than 1 means that pairs overlapping of that length are at least a standard deviation more numerous than the mean of all the overlaps. The two colors distinguish species equipped with both endo-siRNA proteins (DICER2 and siAGO) from species that lack at least one of them.

Figure 4 The loss of DICER2 and siAGO along the Metazoa evolution. The lophotrochozoan phylogenetic tree is reconstructed according to the ML-tree of Kocot and colleagues (2016). Colored boxes represent the presence of the related protein in that clade, while crossed boxes represent the absence of the related protein. Each protein is associated with a color, as reported in the legend.

References

Antoniewski, Christophe. 2014. "Computing SiRNA and PiRNA Overlap Signatures." Pp. 135–46 in *Animal Endo-SiRNAs*. Vol. 1173. New York, NY: Humana Press.

Babiarz, Joshua E., J. Graham Ruby, Yangming Wang, David P. Bartel, and Robert Blelloch. 2008. "Mouse ES Cells Express Endogenous ShRNAs, SiRNAs, and Other Microprocessor-Independent, Dicer-Dependent Small RNAs." *Genes & Development* 22(20):2773–85. doi: 10.1101/gad.1705308.

Bartel, David P. 2018. "Metazoan MicroRNAs." *Cell* 173(1):20–51. doi: 10.1016/j.cell.2018.03.006.

Bernstein, Emily, Amy A. Caudy, Scott M. Hammond, and Gregory J. Hannon. 2001. "Role for a Bidentate Ribonuclease in the Initiation Step of RNA Interference." *Nature* 409(6818):363–66. doi: 10.1038/35053110.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30(15):2114–20. doi: 10.1093/bioinformatics/btu170.

Brennecke, Julius, Alexei A. Aravin, Alexander Stark, Monica Dus, Manolis Kellis, Ravi Sachidanandam, and Gregory J. Hannon. 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.

Brůna, Tomáš, Katharina J. Hoff, Alexandre Lomsadze, Mario Stanke, and Mark Borodovsky. 2021. "BRAKER2: Automatic Eukaryotic Genome Annotation with GeneMark-EP+ and AUGUSTUS Supported by a Protein Database." *NAR Genomics and Bioinformatics* 3(1). doi: 10.1093/nargab/lqaa108.

Buchfink, Benjamin, Chao Xie, and Daniel H. Huson. 2015. "Fast and Sensitive Protein Alignment Using DIAMOND." *Nature Methods* 12(1):59–60. doi: 10.1038/nmeth.3176.

Calcino, Andrew D., Selene L. Fernandez-Valverde, Ryan J. Taft, and Bernard M. Degnan. 2018. "Diverse RNA Interference Strategies in Early-Branching Metazoans." *BMC Evolutionary Biology* 18(1):160. doi: 10.1186/s12862-018-1274-2.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. 2009. “BLAST+: Architecture and Applications.” *BMC Bioinformatics* 10(1):421. doi: 10.1186/1471-2105-10-421.

Castresana, J. 2000. “Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis.” *Molecular Biology and Evolution* 17(4):540–52. doi: 10.1093/oxfordjournals.molbev.a026334.

Czech, Benjamin, Colin D. Malone, Rui Zhou, Alexander Stark, Catherine Schlingeheyde, Monica Dus, Norbert Perrimon, Manolis Kellis, James A. Wohlschlegel, Ravi Sachidanandam, Gregory J. Hannon, and Julius Brennecke. 2008. “An Endogenous Small Interfering RNA Pathway in Drosophila.” *Nature* 453(7196):798–802. doi: 10.1038/nature07007.

Dainat, J. n.d. “AGAT: Another Gff Analysis Toolkit to Handle Annotations in Any GTF/GFF Format.”

Fontenla, Santiago, Gabriel Rinaldi, Pablo Smircich, and Jose F. Tort. 2017. “Conservation and Diversification of Small RNA Pathways within Flatworms.” *BMC Evolutionary Biology* 17(1):215. doi: 10.1186/s12862-017-1061-5.

Fontenla, Santiago, Gabriel Rinaldi, and Jose F. Tort. 2021. “Lost and Found: Piwi and Argonaute Pathways in Flatworms.” *Frontiers in Cellular and Infection Microbiology* 11. doi: 10.3389/fcimb.2021.653695.

Friedländer, Marc R., Sebastian D. Mackowiak, Na Li, Wei Chen, and Nikolaus Rajewsky. 2012. “MiRDeep2 Accurately Identifies Known and Hundreds of Novel MicroRNA Genes in Seven Animal Clades.” *Nucleic Acids Research* 40(1):37–52. doi: 10.1093/nar/gkr688.

Fromm, Bastian, Eirik Høye, Diana Domanska, Xiangfu Zhong, Ernesto Aparicio-Puerta, Vladimir Ovchinnikov, Sinan U. Umu, Peter J. Chabot, Wenjing Kang, Morteza Aslanzadeh, Marcel Tarbier, Emilio Mármol-Sánchez, Gianvito Urgese, Morten Johansen, Eivind Hovig, Michael Hackenberg, Marc R. Friedländer, and Kevin J. Peterson. 2022. “MirGeneDB 2.1: Toward a Complete Sampling of All Major Animal Phyla.” *Nucleic Acids Research* 50(D1):D204–10. doi: 10.1093/nar/gkab1101.

Galiana-Arnoux, Delphine, Catherine Dostert, Anette Schneemann, Jules A. Hoffmann, and Jean-Luc Imler. 2006. "Essential Function in Vivo for Dicer-2 in Host Defense against RNA Viruses in Drosophila." *Nature Immunology* 7(6):590–97. doi: 10.1038/ni1335.

Gammon, Don B., and Craig C. Mello. 2015. "RNA Interference-Mediated Antiviral Defense in Insects." *Current Opinion in Insect Science* 8:111–20. doi: 10.1016/j.cois.2015.01.006.

Garcia Silva, Maria R., Juan P. Tosar, Magali Frugier, Sergio Pantano, Braulio Bonilla, Luis Esteban, Esteban Serra, Carlos Rovira, Carlos Robello, and Alfonso Cayota. 2010. "Cloning, Characterization and Subcellular Localization of a Trypanosoma Cruzi Argonaute Protein Defining a New Subfamily Distinctive of Trypanosomatids." *Gene* 466(1–2):26–35. doi: 10.1016/j.gene.2010.06.012.

Grabherr, Manfred G., Brian J. Haas, Moran Yassour, Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir Hacohen, Andreas Gnirke, Nicholas Rhind, Federica di Palma, Bruce W. Birren, Chad Nusbaum, Kerstin Lindblad-Toh, Nir Friedman, and Aviv Regev. 2011. "Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome." *Nature Biotechnology* 29(7):644–52. doi: 10.1038/nbt.1883.

Grimson, Andrew, Mansi Srivastava, Bryony Fahey, Ben J. Woodcroft, H. Rosaria Chiang, Nicole King, Bernard M. Degnan, Daniel S. Rokhsar, and David P. Bartel. 2008. "Early Origins and Evolution of MicroRNAs and Piwi-Interacting RNAs in Animals." *Nature* 455(7217):1193–97. doi: 10.1038/nature07415.

Guindon, Stéphane, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and Olivier Gascuel. 2010. "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0." *Systematic Biology* 59(3):307–21. doi: 10.1093/sysbio/syq010.

Havens, Mallory A., Ashley A. Reich, Dominik M. Duelli, and Michelle L. Hastings. 2012. "Biogenesis of Mammalian MicroRNAs by a Non-Canonical Processing Pathway." *Nucleic Acids Research* 40(10):4626–40. doi: 10.1093/nar/gks026.

Hirakata, Shigeki, and Mikiko C. Siomi. 2016. “PiRNA Biogenesis in the Germline: From Transcription of PiRNA Genomic Sources to PiRNA Maturation.” *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1859(1):82–92. doi: 10.1016/j.bbagrm.2015.09.002.

Hoang, Diep Thi, Olga Chernomor, Arndt von Haeseler, Bui Quang Minh, and Le Sy Vinh. 2018. “UFBoot2: Improving the Ultrafast Bootstrap Approximation.” *Molecular Biology and Evolution* 35(2):518–22. doi: 10.1093/molbev/msx281.

Höck, Julia, and Gunter Meister. 2008. “The Argonaute Protein Family.” *Genome Biology* 9(2):210. doi: 10.1186/gb-2008-9-2-210.

Hoff, Katharina J., Alexandre Lomsadze, Mark Borodovsky, and Mario Stanke. 2019. “Whole-Genome Annotation with BRAKER.” Pp. 65–95 in.

Huntzinger, Eric, and Elisa Izaurralde. 2011. “Gene Silencing by MicroRNAs: Contributions of Translational Repression and MRNA Decay.” *Nature Reviews Genetics* 12(2):99–110. doi: 10.1038/nrg2936.

Iwakawa, Hiro-oki, and Yukihide Tomari. 2022. “Life of RISC: Formation, Action, and Degradation of RNA-Induced Silencing Complex.” *Molecular Cell* 82(1):30–43. doi: 10.1016/j.molcel.2021.11.026.

Jehn, Julia, Daniel Gebert, Frank Pipilescu, Sarah Stern, Julian Simon Thilo Kiefer, Charlotte Hewel, and David Rosenkranz. 2018. “PIWI Genes and PiRNAs Are Ubiquitously Expressed in Mollusks and Show Patterns of Lineage-Specific Adaptation.” *Communications Biology* 1(1):137. doi: 10.1038/s42003-018-0141-4.

Katoh, K., and D. M. Standley. 2013. “MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.” *Molecular Biology and Evolution* 30(4):772–80. doi: 10.1093/molbev/mst010.

Khalturin, Konstantin, Natalia Shunatova, Sergei Shchenkov, Yasunori Sasakura, Mayumi Kawamitsu, and Noriyuki Satoh. 2022. “Polyzoa Is Back: The Effect of Complete Gene Sets on the Placement of Ectoprocta and Entoprocta.” *Science Advances* 8(26). doi: 10.1126/sciadv.abo4400.

Khanal, Sweta, Beatriz Schueng Zancanela, Jacob Oche Peter, and Alex Sutton Flynt. 2022. “The Small RNA Universe of Capitella Teleta.” *Frontiers in Molecular Biosciences* 9. doi: 10.3389/fmolb.2022.802814.

Kocot, Kevin M., Torsten H. Struck, Julia Merkel, Damien S. Waits, Christiane Todt, Pamela M. Brannock, David A. Weese, Johanna T. Cannon, Leonid L. Moroz, Bernhard Lieb, and Kenneth M. Halanych. 2016. “Phylogenomics of Lophotrochozoa with Consideration of Systematic Error.” *Systematic Biology* syw079. doi: 10.1093/sysbio/syw079.

Kozomara, Ana, Maria Birgaoanu, and Sam Griffiths-Jones. 2019. “MiRBase: From MicroRNA Sequences to Function.” *Nucleic Acids Research* 47(D1):D155–62. doi: 10.1093/nar/gky1141.

Kuhn, Claus-D., and Leemor Joshua-Tor. 2013. “Eukaryotic Argonautes Come into Focus.” *Trends in Biochemical Sciences* 38(5):263–71. doi: 10.1016/j.tibs.2013.02.008.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. 2009. “Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome.” *Genome Biology* 10(3):R25. doi: 10.1186/gb-2009-10-3-r25.

Laura, Sala, Chandrasekhar Srividya, and Joana A. Vidigal. 2020. “AGO Unchained Canonical and Non-Canonical Roles of Argonaute Proteins in Mammals.” *Frontiers in Bioscience* 25(1):4793. doi: 10.2741/4793.

Lee, Yoontae, Chiyoung Ahn, Jinju Han, Hyounjeong Choi, Jaekwang Kim, Jeongbin Yim, Junho Lee, Patrick Provost, Olof Rådmark, Sunyoung Kim, and V. Narry Kim. 2003. “The Nuclear RNase III Drosha Initiates MicroRNA Processing.” *Nature* 425(6956):415–19. doi: 10.1038/nature01957.

Lee, Yoontae, Minju Kim, Jinju Han, Kyu-Hyun Yeom, Sanghyuk Lee, Sung Hee Baek, and V. Narry Kim. 2004. “MicroRNA Genes Are Transcribed by RNA Polymerase II.” *The EMBO Journal* 23(20):4051–60. doi: 10.1038/sj.emboj.7600385.

Lee, Young Sik, Kenji Nakahara, John W. Pham, Kevin Kim, Zhengying He, Erik J. Sontheimer, and Richard W. Carthew. 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.

Lund, Steven P., Dan Nettleton, Davis J. McCarthy, and Gordon K. Smyth. 2012. "Detecting Differential Expression in RNA-Sequence Data Using Quasi-Likelihood with Shrunken Dispersion Estimates." *Statistical Applications in Genetics and Molecular Biology* 11(5). doi: 10.1515/1544-6115.1826.

Malone, Colin D., and Gregory J. Hannon. 2009. "Small RNAs as Guardians of the Genome." *Cell* 136(4):656–68. doi: 10.1016/j.cell.2009.01.045.

Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." *EMBnet.Journal* 17(1):10. doi: 10.14806/ej.17.1.200.

Matranga, Christian, and Phillip D. Zamore. 2007. "Small Silencing RNAs." *Current Biology* 17(18):R789–93. doi: 10.1016/j.cub.2007.07.014.

Mistry, Jaina, Sara Chuguransky, Lowri Williams, Matloob Qureshi, Gustavo A. Salazar, Erik L. L. Sonnhammer, Silvio C. E. Tosatto, Lisanna Paladin, Shriya Raj, Lorna J. Richardson, Robert D. Finn, and Alex Bateman. 2021. "Pfam: The Protein Families Database in 2021." *Nucleic Acids Research* 49(D1):D412–19. doi: 10.1093/nar/gkaa913.

Mistry, Jaina, Robert D. Finn, Sean R. Eddy, Alex Bateman, and Marco Punta. 2013. "Challenges in Homology Search: HMMER3 and Convergent Evolution of Coiled-Coil Regions." *Nucleic Acids Research* 41(12):e121–e121. doi: 10.1093/nar/gkt263.

Mohn, F., D. Handler, and J. Brennecke. 2015. "PiRNA-Guided Slicing Specifies Transcripts for Zucchini-Dependent, Phased PiRNA Biogenesis." *Science* 348(6236):812–17. doi: 10.1126/science.aaa1039.

Mukherjee, Krishanu, Henry Campos, and Bryan Kolaczkowski. 2013a. "Evolution of Animal and Plant Dicers: Early Parallel Duplications and Recurrent Adaptation of Antiviral RNA Binding in Plants." *Molecular Biology and Evolution* 30(3):627–41. doi: 10.1093/molbev/mss263.

Mukherjee, Krishanu, Henry Campos, and Bryan Kolaczkowski. 2013b. "Evolution of Animal and Plant Dicers: Early Parallel Duplications and Recurrent Adaptation of Antiviral RNA Binding in Plants." *Molecular Biology and Evolution* 30(3):627–41. doi: 10.1093/molbev/mss263.

Nguyen, Lam-Tung, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2015. “IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.” *Molecular Biology and Evolution* 32(1):268–74. doi: 10.1093/molbev/msu300.

Ozata, Deniz M., Ildar Gainetdinov, Ansgar Zoch, Dónal O’Carroll, and Phillip D. Zamore. 2019. “PIWI-Interacting RNAs: Small RNAs with Big Functions.” *Nature Reviews Genetics* 20(2):89–108. doi: 10.1038/s41576-018-0073-3.

Park, Sehee, Igojo Kang, and Chanseok Shin. 2021. “MicroRNA Clustering on the Biogenesis of Suboptimal MicroRNAs.” *Applied Biological Chemistry* 64(1):51. doi: 10.1186/s13765-021-00624-3.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. “EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data.” *Bioinformatics* 26(1):139–40. doi: 10.1093/bioinformatics/btp616.

Sayers, Eric W., Evan E. Bolton, J. Rodney Brister, Kathi Canese, Jessica Chan, Donald C. Comeau, Ryan Connor, Kathryn Funk, Chris Kelly, Sunghwan Kim, Tom Madej, Aron Marchler-Bauer, Christopher Lanczycki, Stacy Lathrop, Zhiyong Lu, Francoise Thibaud-Nissen, Terence Murphy, Lon Phan, Yuri Skripchenko, Tony Tse, Jiyao Wang, Rebecca Williams, Barton W. Trawick, Kim D. Pruitt, and Stephen T. Sherry. 2022. “Database Resources of the National Center for Biotechnology Information.” *Nucleic Acids Research* 50(D1):D20–26. doi: 10.1093/nar/gkab1112.

Shabalina, S., and E. Koonin. 2008. “Origins and Evolution of Eukaryotic RNA Interference.” *Trends in Ecology & Evolution* 23(10):578–87. doi: 10.1016/j.tree.2008.06.005.

Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2015. “BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.” *Bioinformatics* 31(19):3210–12. doi: 10.1093/bioinformatics/btv351.

Song, Ji-Joon, Stephanie K. Smith, Gregory J. Hannon, and Leemor Joshua-Tor. 2004. “Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity.” *Science* 305(5689):1434–37. doi: 10.1126/science.1102514.

Stanke, Mario, Mark Diekhans, Robert Baertsch, and David Haussler. 2008. “Using Native and Syntenically Mapped CDNA Alignments to Improve de Novo Gene Finding.” *Bioinformatics* 24(5):637–44. doi: 10.1093/bioinformatics/btn013.

Strassert, Jürgen F. H., Iker Irisarri, Tom A. Williams, and Fabien Burki. 2021. “A Molecular Timescale for Eukaryote Evolution with Implications for the Origin of Red Algal-Derived Plastids.” *Nature Communications* 12(1):1879. doi: 10.1038/s41467-021-22044-z.

Swarts, Daan C., Kira Makarova, Yanli Wang, Kotaro Nakanishi, René F. Ketting, Eugene V Koonin, Dinshaw J. Patel, and John van der Oost. 2014. “The Evolutionary Journey of Argonaute Proteins.” *Nature Structural & Molecular Biology* 21(9):743–53. doi: 10.1038/nsmb.2879.

Tam, Oliver H., Alexei A. Aravin, Paula Stein, Angelique Girard, Elizabeth P. Murchison, Sihem Cheloufi, Emily Hodges, Martin Anger, Ravi Sachidanandam, Richard M. Schultz, and Gregory J. Hannon. 2008. “Pseudogene-Derived Small Interfering RNAs Regulate Gene Expression in Mouse Oocytes.” *Nature* 453(7194):534–38. doi: 10.1038/nature06904.

Tarver, James E., Erik A. Sperling, Audrey Nailor, Alysha M. Heimberg, Jeffrey M. Robinson, Benjamin L. King, Davide Pisani, Philip C. J. Donoghue, and Kevin J. Peterson. 2013. “MiRNAs: Small Genes with Big Potential in Metazoan Phylogenetics.” *Molecular Biology and Evolution* 30(11):2369–82. doi: 10.1093/molbev/mst133.

Thibaud-Nissen, F., M. DiCuccio, W. Hlavina, A. Kimchi, P. A. Kitts, T. D. Murphy, K. D. Pruitt, and A. Souvorov. 2016. “P8008 The NCBI Eukaryotic Genome Annotation Pipeline.” *Journal of Animal Science* 94:184–184. doi: 10.2527/jas2016.94supplement4184x.

Watanabe, Toshiaki, Yasushi Totoki, Atsushi Toyoda, Masahiro Kaneda, Satomi Kuramochi-Miyagawa, Yayoi Obata, Hatsune Chiba, Yuji Kohara, Tomohiro Kono, Toru Nakano, M. Azim Surani, Yoshiyuki Sakaki, and Hiroyuki Sasaki. 2008. “Endogenous SiRNAs from Naturally Formed DsRNAs Regulate Transcripts in Mouse Oocytes.” *Nature* 453(7194):539–43. doi: 10.1038/nature06908.

Weick, Eva-Maria, and Eric A. Miska. 2014. “PiRNAs: From Biogenesis to Function.” *Development* 141(18):3458–71. doi: 10.1242/dev.094037.

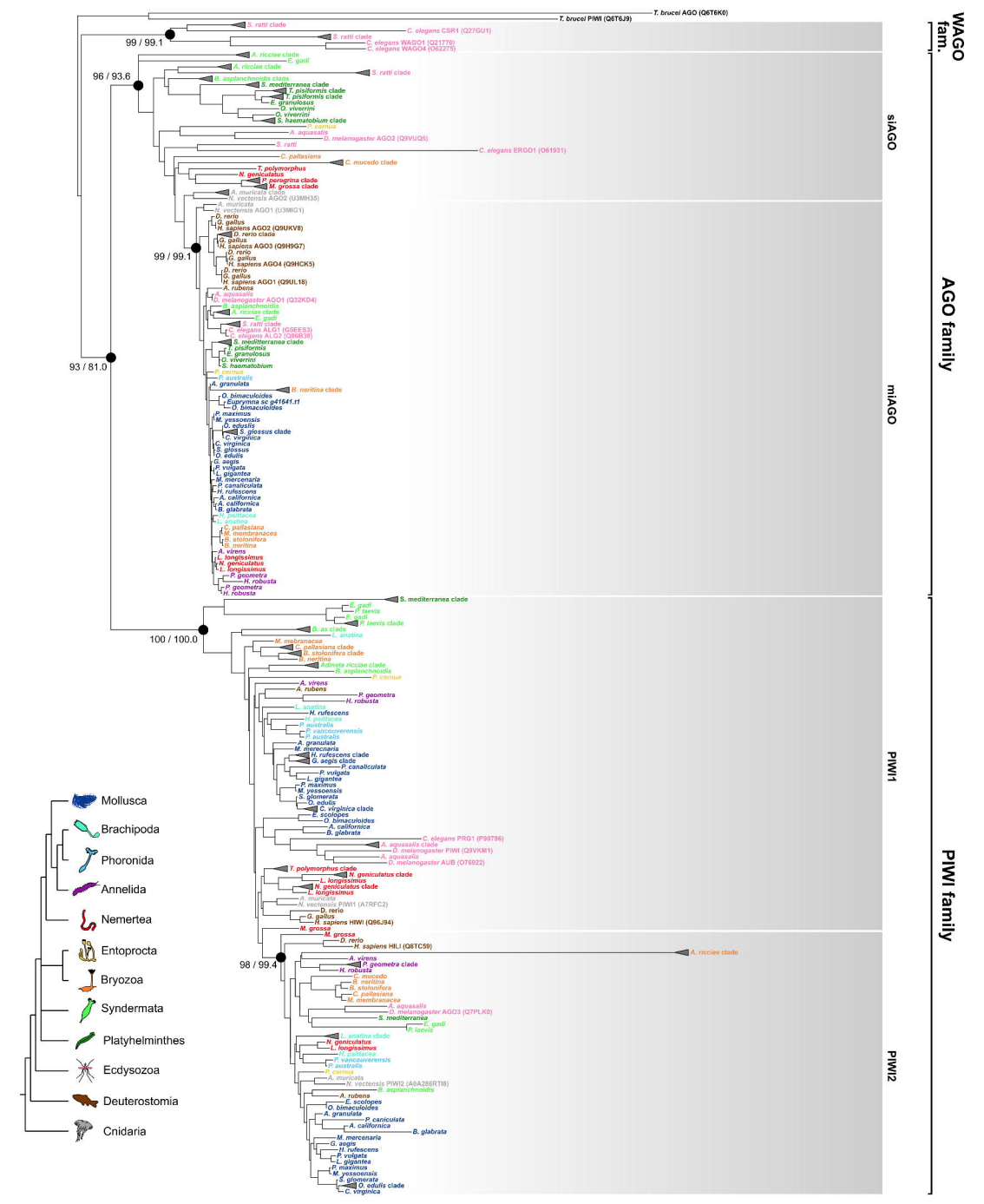
Wertheim, Joel O., Ben Murrell, Martin D. Smith, Sergei L. Kosakovsky Pond, and Konrad Scheffler. 2015. "RELAX: Detecting Relaxed Selection in a Phylogenetic Framework." *Molecular Biology and Evolution* 32(3):820–32. doi: 10.1093/molbev/msu400.

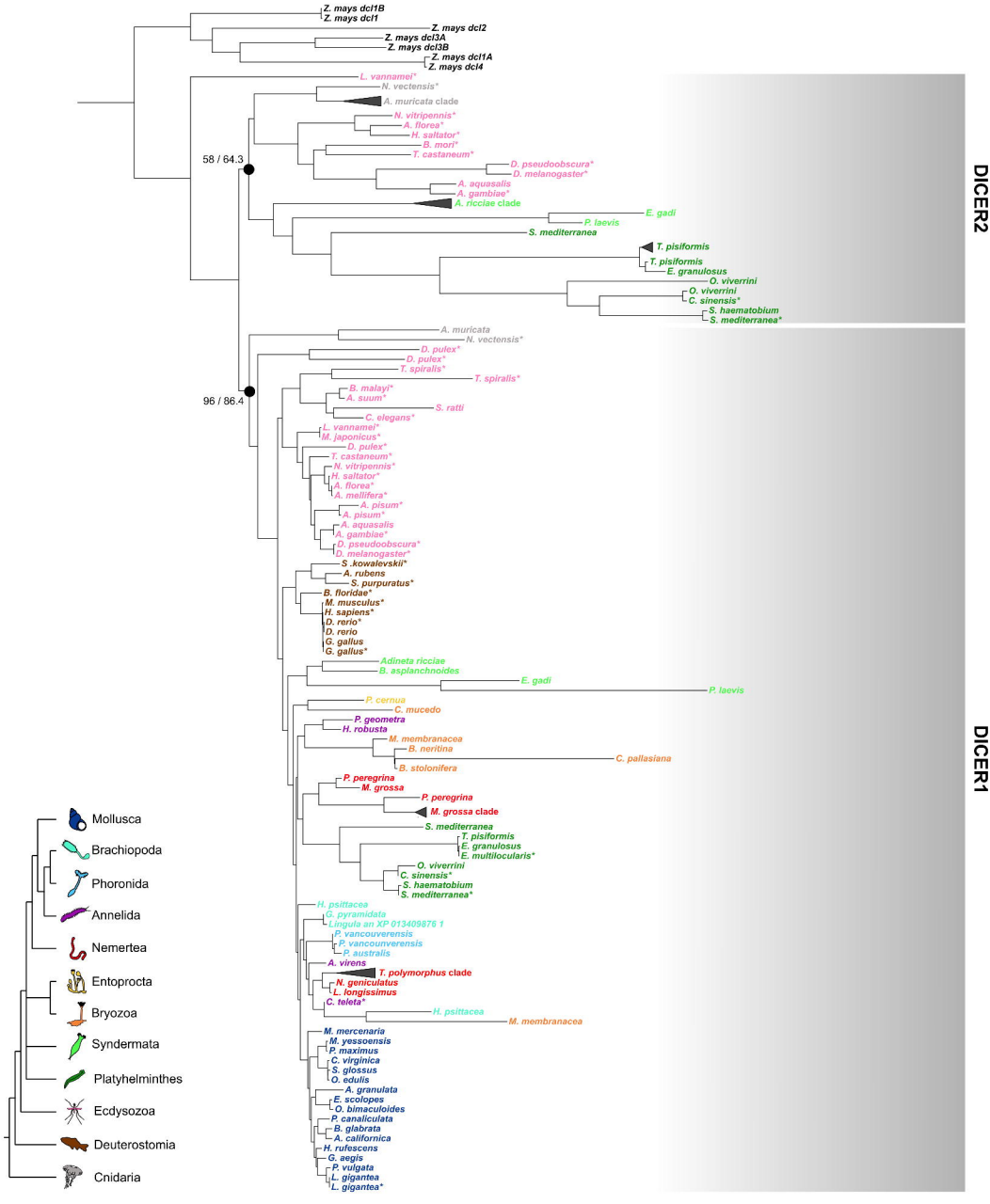
Wheeler, Benjamin M., Alysha M. Heimberg, Vanessa N. Moy, Erik A. Sperling, Thomas W. Holstein, Steffen Heber, and Kevin J. Peterson. 2009. "The Deep Evolution of Metazoan MicroRNAs." *Evolution & Development* 11(1):50–68. doi: 10.1111/j.1525-142X.2008.00302.x.

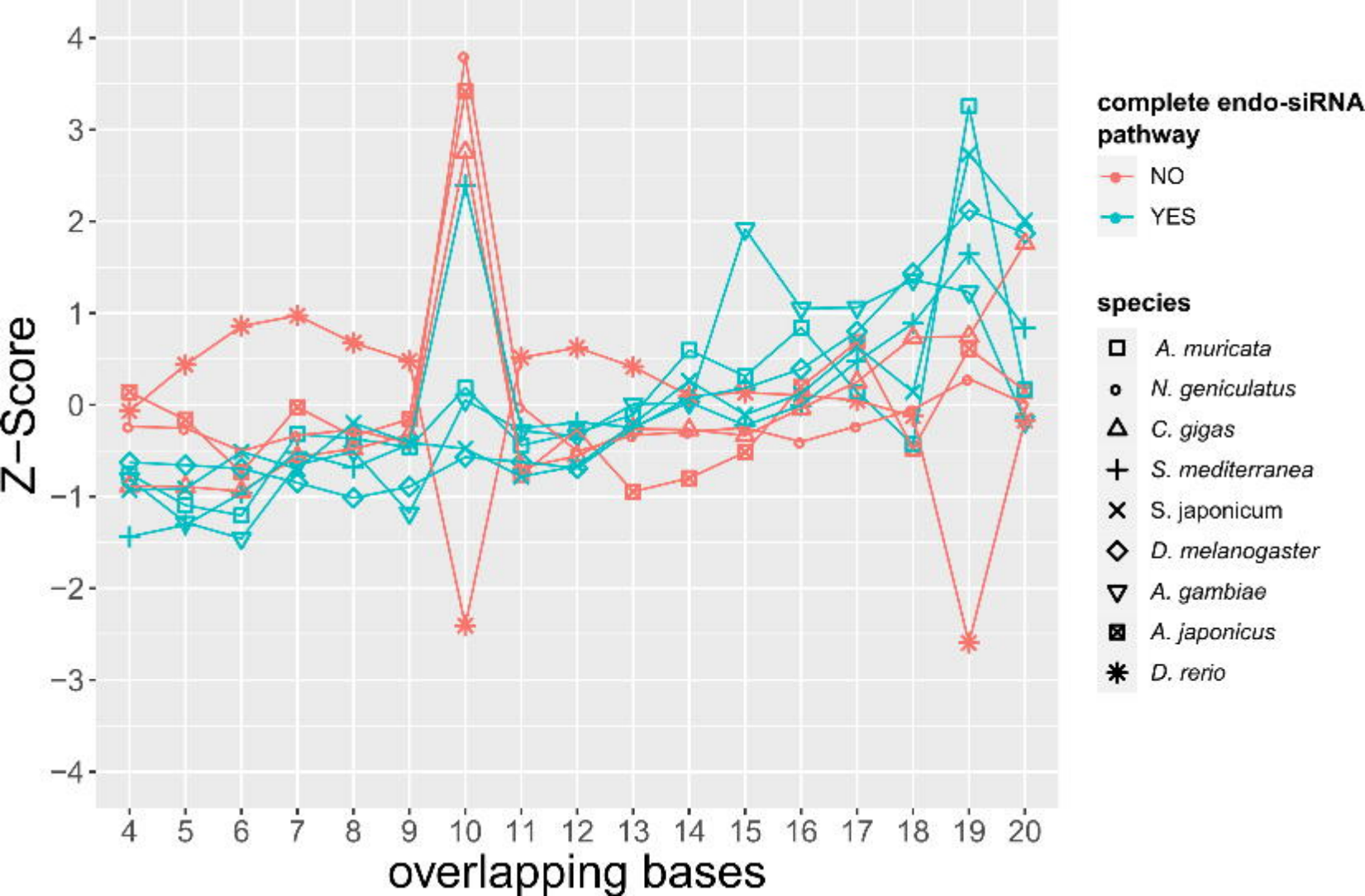
Wu, Jin'en, Jing Yang, William C. Cho, and Yadong Zheng. 2020. "Argonaute Proteins: Structural Features, Functions and Emerging Roles." *Journal of Advanced Research* 24:317–24. doi: 10.1016/j.jare.2020.04.017.

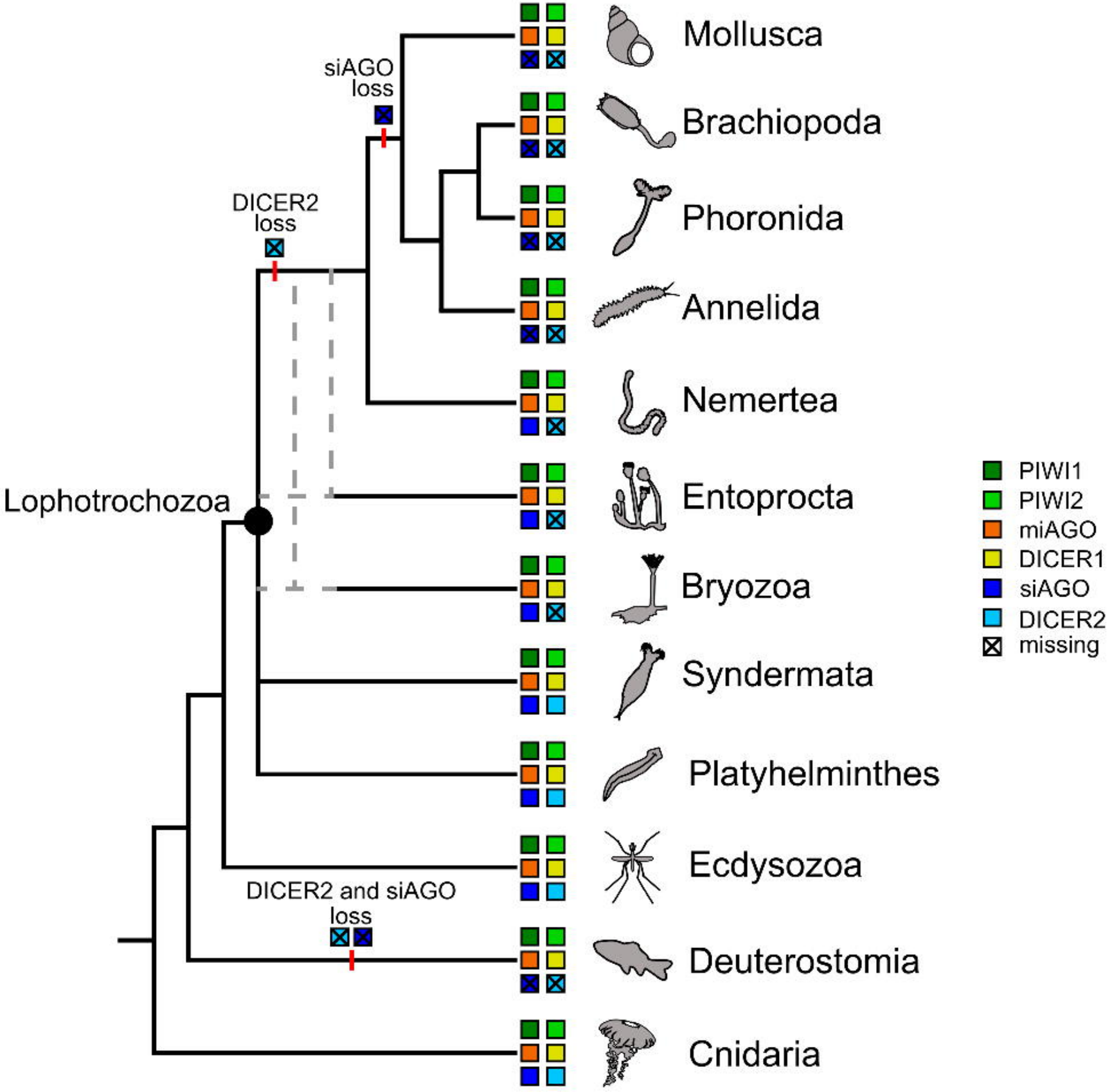
Wynant, Niels, Dulce Santos, and Jozef Vanden Broeck. 2017. "The Evolution of Animal Argonautes: Evidence for the Absence of Antiviral AGO Argonautes in Vertebrates." *Scientific Reports* 7(1):9230. doi: 10.1038/s41598-017-08043-5.

Zdobnov, Evgeny M., Dmitry Kuznetsov, Fredrik Tegenfeldt, Mosè Manni, Matthew Berkeley, and Evgenia V Kriventseva. 2021. "OrthoDB in 2020: Evolutionary and Functional Annotations of Orthologs." *Nucleic Acids Research* 49(D1):D389–93. doi: 10.1093/nar/gkaa1009.


T. brucei AGO (Q6T6K0)
T. brucei PIWI (Q6T6J9)
S. ratti clade
C. elegans CSR1 (Q27GU1)
S. ratti clade
C. elegans WAGO1 (Q21770)
C. elegans WAGO4 (O62275)
99 / 99.1
A. ricciae clade
E. gadi
A. ricciae clade
S. ratti clade
96 / 93.6
B. asplanchnoidis clade
S. mediterranea clade
T. pisiformis clade
T. pisiformis clade
E. granulosus
O. viverrini
O. viverrini
S. haematobium clade
P. cornua
A. aquasalis
D. melanogaster AGO2 (Q9VUQ5)
S. ratti
C. elegans ERGO1 (O61931)
C. pallasiana
C. mucedo clade
T. polymorphus
N. geniculatus
P. peregrina clade
M. grossa clade
A. muricata clade
N. vectensis AGO2 (U3MH35)
A. muricata
N. vectensis AGO1 (U3MIG1)
D. rerio
G. gallus
H. sapiens AGO2 (Q9UKV8)
D. rerio clade
G. gallus
H. sapiens AGO3 (Q9H9G7)
D. rerio
G. gallus
H. sapiens AGO4 (Q9HCK5)
D. rerio
G. gallus
H. sapiens AGO1 (Q9UL18)
A. rubens
A. aquasalis
D. melanogaster AGO1 (Q32KD4)
B. asplanchnoidis
A. ricciae clade
E. gadi
S. ratti clade
C. elegans ALG1 (G5EES3)
C. elegans ALG2 (Q86B38)
S. mediterranea clade
T. pisiformis
E. granulosus
O. viverrini
S. haematobium
P. cornua
P. australis
B. neritina clade
99 / 99.1
93 / 81.0
O. bimaculoides
Euprymna sc g41641.t1
O. bimaculoides
P. maximus
M. yessoensis
D. edulis
S. glossus clade
C. virginica
C. virginica
C. glossus
D. edulis
G. aegis
P. vulgata
L. gigantea
M. mercenaria
P. canaliculata
H. rufescens
A. californica
A. californica
B. glabrata
L. gigantea
C. pallasiana
M. membranacea
B. stolonifera
B. neritina
A. virens
L. longissimus
N. geniculatus
L. longissimus
P. gnometra
H. robusta
P. geometra
H. robusta
S. mediterranea clade
E. gadi
P. laevis
E. gadi
P. laevis clade
B. as clade
L. anatina
100 / 100.0
M. mabranacea
C. pallasiana clade
B. stolonifera clade
B. neritina
Aiptasia ricciae clade
B. asplanchnoidis
P. cornua
A. virens
A. rubens
P. geometra
H. robusta
L. longissimus
N. rufescens
H. pontacea
P. australis
P. vancouverensis
P. australis
A. granulata
M. mercenaria
H. rufescens clade
G. aegis clade
P. canaliculata
P. vulgata
L. gigantea
P. maximus
M. yessoensis
S. glomerata
D. edulis
C. virginica clade
E. scolopes
O. bimaculoides
A. californica
B. glabrata
C. elegans PRG1 (P90786)
A. aquasalis clade
D. melanogaster PIWI (Q9VKM1)
A. aquasalis
D. melanogaster AUB (O76922)
T. polymorphus clade
N. geniculatus clade
L. longissimus
N. geniculatus clade
L. longissimus
A. muricata
N. vectensis PIWI1 (A7RFC2)
D. rerio
G. gallus
H. sapiens HIWI (Q96J94)
M. grossa
M. grossa
D. rerio
H. sapiens HILI (Q8TC59)
98 / 99.4
A. ricciae clade
A. virens
P. geometra clade
H. robusta
C. mucedo
B. neritina
B. stolonifera
C. pallasiana
M. membranacea
A. aquasalis
D. melanogaster AGO3 (Q7PLK0)
S. mediterranea
E. gadi
P. laevis
L. longissimus clade
N. geniculatus
L. longissimus
C. pallasiana
P. vancouverensis
P. australis
P. cornua
A. muricata
N. vectensis PIWI2 (A0A286RTI8)
B. asplanchnoidis
A. rubens
E. scolopes
O. bimaculoides
A. granulata
P. caniculata
A. californica
B. glabrata
M. mercenaria
G. aegis
H. rufescens
P. vulgata
L. gigantea
P. maximus
M. yessoensis
S. glomerata
O. edulis clade
C. virginica
Mollusca
Brachiopoda
Phoronida
Annelida
Nemertea
Entoprocta
Bryozoa
Syndermata
Platyhelminthes
Ecdysozoa
Deuterostomia
Cnidaria
WAGO fam.
AGO family
siAGO
miAGO
PIWI family
PIWI1
PIWI2


Z. mays dcl1B
Z. mays dcl1
Z. mays dcl2
Z. mays dcl3A
Z. mays dcl3B
Z. mays dcl1A
Z. mays dcl4
L. vannamei*
N. vectensis*
A. muricata clade
N. vitripennis*
A. florea*
H. saltator*
B. mori*
T. castaneum*
58 / 64.3
D. pseudoobscura*
D. melanogaster*
A. aquasalis
A. gambiae*
A. ricciae clade
E. gadi
P. laevis
S. mediterranea
T. pisiformis
T. pisiformis
E. granulosus
O. viverrini
O. viverrini
C. sinensis*
S. haematobium
S. mediterranea*
DICER2
A. muricata
N. vectensis*
D. pulex*
D. pulex*
T. spiralis*
T. spiralis*
96 / 86.4
B. malayi*
A. suum*
S. ratti
C. elegans*
L. vannamei*
M. japonicus*
D. pulex*
T. castaneum*
N. vitripennis*
H. saltator*
A. florea*
A. mellifera*
A. pisum*
A. pisum*
A. aquasalis
A. gambiae*
D. pseudoobscura*
D. melanogaster*
S. kowalevskii*
A. rubens
S. purpuratus*
B. floridae*
M. musculus*
H. sapiens*
D. rerio*
D. rerio
G. gallus
G. gallus*
Adineta ricciae
B. asplanchnoides
E. gadi
P. laevis
P. cernua
C. mucedo
P. geometra
H. robusta
M. membranacea
B. neritina
C. pallasiana
B. stolonifera
P. peregrina
M. grossa
P. peregrina
M. grossa clade
S. mediterranea
T. pisiformis
E. granulosus
E. multilocularis*
O. viverrini
C. sinensis*
S. haematobium
S. mediterranea*
H. psittacea
G. pyramidata
Lingula an XP 013409876 1
P. vancouverensis
P. vancouverensis
P. australis
A. virens
T. polymorphus clade
N. geniculatus
L. longissimus
C. teleta*
H. psittacea
M. membranacea
M. mercenaria
M. yessoensis
P. maximus
C. virginica
S. glossus
O. edulis
A. granulata
E. scolopes
O. bimaculoides
P. canaliculata
B. glabrata
A. californica
H. rufescens
G. aegis
P. vulgata
L. gigantea
L. gigantea*
DICER1
Mollusca
Brachiopoda
Phoronida
Annelida
Nemertea
Entoprocta
Bryozoa
Syndermata
Platyhelminthes
Ecdysozoa
Deuterostomia
Cnidaria


Z−Score
4
3
2
1
0
−1
−2
−3
−4
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
overlapping bases
complete endo-siRNA pathway
NO
YES
species
A. muricata
N. geniculatus
C. gigas
S. mediterranea
S. japonicum
D. melanogaster
A. gambiae
A. japonicus
D. rerio


Mollusca
Brachiopoda
Phoronida
Annelida
Nemertea
Entoprocta
Bryozoa
Syndermata
Platyhelminthes
Ecdysozoa
Deuterostomia
Cnidaria
siAGO loss
DICER2 loss
Lophotrochozoa
DICER2 and siAGO loss
PIWI1
PIWI2
miAGO
DICER1
siAGO
DICER2
missing