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, Argonaute, DICER, RNAi

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.

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 (Y. Lee et al. 2004; Bartel 2018). 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 (Bernstein et al. 2001; Bartel 2018). 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; Tam et al. 2008; Malone and Hannon 2009). 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 (Shabalina and Koonin 2008; Gammon and Mello 2015). 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; Moran et al. 2017). 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 (Y.S. Lee et al. 2004; Matranga and Zamore 2007; Czech et al. 2008). 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; Malone and Hannon 2009; Gammon and Mello 2015). 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 (Weick and Miska 2014; Hirakata and Siomi 2016). 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 (Weick and Miska 2014; Hirakata and Siomi 2016). 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 et al. 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 (Grimson et al. 2008; Wheeler et al. 2009; Moran et al. 2013; Praher et al. 2017; Calcino et al. 2018; Fridrich et al. 2020). In some clades piRNA and endo-siRNA pathways have been secondarily lost (Wynant et al. 2017; Fontenla et al. 2021), while the miRNA pathway is likely to be ubiquitous in Metazoa (Fromm et al. 2022). 6

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; Fontenla et al. 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.

122 **Results**

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

131 All the annotated Argonaute proteins were aligned, and the Maximum-Likelihood (ML) tree was inferred. The
132 PIWI and AGO proteins of *Trypanosoma brucei* were obtained from Uniprot and used as outgroups. The
133 phylogenetic tree of the Argonaute superfamily supported the known main families (Fig. 1; supplementary
134 figure S1). The WAGO family, which is restricted to nematodes, included WAGOs and CSR-1 sequences from
135 *C. elegans*. The PIWI-like family was characterized by AGO3, PIWI and AUB of *D. melanogaster*, HILI and
136 HIWI of *H. sapiens* and PRG-1 of *C. elegans;* the AGO-like family comprised AGO1,2 of *D. melanogaster*,
137 AGO1-4 of *H. sapiens* and ALG1,2 of *C. elegans*. Every family was widely supported by UFBoot and the
138 SH-alrt test (Fig. 1).
139 The AGO-like family was divided in two groups. One monophyletic group (UFBoot = 99 and SH-alrt = 99.1)
140 included *H. sapiens* AGO1-4, *D. melanogaster* and *N. vectensis* AGO1 and *C. elegans* ALG1,2: all these
141 proteins interact with miRNAs. Almost all species featured a protein clustering within this clade, with very
142 few exceptions, and at least one organism from every phylum was recovered in this clade (Fig. 1;
143 supplementary table S1). We will refer to this clade as the "miAGO clade". The second group within the AGO-
144 like family was polyphyletic. *D. melanogaster* AGO2, *N. vectensis* AGO2 and *C. elegans* ERGO1 clustered
145 within this clade. All these proteins target endo-siRNAs, thus we will henceforth call this group the "siAGO
146 group". Few lophotrochozoan phyla are included in this group, since we annotated at least one siAGO protein
147 for all the Platyhelminthes and Entoprocta species, while three other phyla featured at least one species within
148 the group, namely Nemertea, Bryozoa and Syndermata. We did not retrieve any siAGO protein from the
149 remaining clades, namely Mollusca, Annelida, Brachiopoda, Phoronida (Fig. 1; supplementary table S1).
150 Similarly to the AGO-like family, protein clustering in the PIWI-like family can be divided in two groups: one
151 is monophyletic and one is polyphyletic. The monophyletic group (UFBoot = 98 and SH-alrt = 99.4) comprised
152 the *H. sapiens* HILI, the *N. vectensis* PIWI2 and the *D. melanogaster* AGO3; the polyphyletic group included
153 *H. sapiens* HIWI, the *N. vectensis* PIWI1, the *C. elegans* PRG1 and the *D. melanogaster* AUB and PIWI. Each
154 phylum was present in both groups with at least one species (supplementary table S1). PIWI-like proteins were
155 almost absent in Platyhelminthes, apart from *Schmitdea mediterranea* (Fig. 1).
156 Regarding the six non-lophotrochozoan species included in the analysis, the annotation of Argonautes proteins
157 matched the expectations: miAGO proteins were retrieved for all six species; siAGO proteins were absent in

158 *D. rerio, G. gallus* and *A. rubens* (Deuterostomia); all six species reported PIWI1 and PIWI2 proteins,
159 exception made for *S. ratti* (i.e. that lacked of both PIWI proteins) and, unexpectedly, *G. gallus* that lacked of
160 PIWI2 (Fig. 1; supplementary table S1).

161 The phylogenetic analysis highlighted the presence of miAGO proteins in each phylum, but only some of them
162 feature siAGO proteins. In Arthropoda, the preliminary structures of siRNAs and miRNAs are processed by
163 two distinct DICER paralogs: DICER2 and DICER1, respectively (Shabalina and Koonin 2008). DICER2 has
164 been found in other phyla, like Cnidaria and Platyhelminthes (Mukherjee et al. 2013a), while clades lacking
165 siAGO proteins lack DICER2 as well. We annotated DICER proteins and inferred the phylogeny to understand
166 whether Lophotrochozoa follow the same pattern.
167 We annotated DICER proteins querying the lophotrochozoan sequences against an annotated metazoan DICER
168 set (Mukherjee et al. 2013a) and looking for the ribonucleases 3 domain. The phylogenetic analysis included
169 the metazoan DICER set of Mukherjee and colleagues (2013) as references, including the *Zea mays* DICER
170 proteins as outgroups (Fig. 2; supplementary figure S2). The resulting ML tree clustered the DICER proteins
171 in two distinct groups. Recall the position of the reference sequences, DICER1 and DICER2 sequences are
172 accordingly split in the two groups, apart from *Litopenaeus vannamei* DICER2, which is basal to all the other
173 proteins. The monophyly of the DICER1 group is supported by the SH-alrt test (86.4) and the UFBoot (96).
174 On the contrary, the low support values of the DICER2 node (UFBoot= 58 and SH-alrt = 64.3) undermine the
175 hypothesis of a unique common origin of DICER2 proteins. Nonetheless, the presence of DICER2 was
176 restricted to few lophotrochozoan phyla: all Platyhelminthes and three Syndermata species showed a DICER2
177 protein. On the other hand, DICER1 was annotated for every phylum: thus, Mollusca, Annelida, Phoronida,
178 Brachiopoda, Nemertea, Bryozoa and Entoprocta reported DICER1 proteins, but no DICER2 proteins (Fig. 2;
179 supplementary table S1). In line with the expectations, DICER2 was reported for *A. muricata, A. gambiae,*
180 while it was lacking in *D. rerio, G. gallus, A. rubens* and *S. ratti.* On the contrary, DICER1 was annotated in
181 all of them.
182 Overall, the DICER family phylogenetic analysis confirmed that the absence of siAGO proteins coincides with
183 the absence of DICER2, and *vice versa*, exception made for Nemertea, Bryozoa and Entoprocta, where we
184 annotated siAGO proteins for most species; however, none of these species featured DICER2. DICER1 and
185 miAGO proteins showed lower evolutionary rates than their paralogous counterparts: the root-to-tip distances
186 of miAGO branches were significantly lower than the root-to-tip distances of siAGO, PIWI1 and PIWI2
187 branches, and the root-to-tip distances of DICER1 branches were significantly lower than their DICER2
188 counterparts (supplementary figure S3). We also estimated the ratio of non-synonymous to synonymous
189 substitution rates (ω) along the AGO family tree and the DICER family tree: we confirmed that, during the
190 evolution of DICER and AGO family, purifying selection on DICER1 (LRT=34.34, p-value = 4.62×10^{-9};
191 supplementary figure S4B) and miAGO (LRT=204,64, p-value = 0; supplementary figure S4A) has intensified
192 compared to the rest of the family tree (Wertheim et al. 2015).

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

204 According to the phylogenomic analysis, some lophotrochozoan phyla lack pivotal proteins related to the endo-
205 siRNA pathway. DICER2 generally processes double stranded RNAs producing two 21 bases siRNA duplexes
206 that overlaps of 19 bases. Similarly, piRNAs produced by the ping-pong cycle go in pairs that overlaps of 10
207 bases (Antoniewski 2014; Khanal et al. 2022). Thus, siRNAs and piRNAs have a unique signature that can be
208 spot in sncRNA libraries by looking for overlapping pairs of reads. We retrieved eight sncRNA libraries from
209 lophotrochozoan and non-lophotrochozoan species (i.e., *Danio rerio, Apostichopus japonicus, Acropora*
210 *muricata, Anopheles gambiae, Drosophila melanogaster, Schmitdea mediterranea, Schistosoma japonicum,*
211 *Crassostrea gigas*) and we sequenced the sncRNA pool of *Notospermus geniculatus*, since no nemertean
212 sncRNA library was available. Using the overlapping_reads.py script (Antoniewski 2014) we calculated the
213 number of read pairs that overlapped for the same number of bases, from 4 to 20 bases. Then, we calculated
214 the Z-Score among the number of read pairs for each overlap group. A Z-score equal to 1 means that the
215 number of reads pairs in that overlap group is one standard deviation higher from the mean size of all the
216 overlap groups of a given species. Taking into consideration only 21 bp reads (i.e., the expected length of
217 endo-siRNAs), Arthropoda (*D. melanogaster* and *A. gambiae*), Cnidaria (*A. muricata*) and Platyhelminthes (*S.*
218 *mediterranea* and *S. japonicum*) representatives reported a Z-score higher than 1 in the 19-overlap group,
219 corresponding to the siRNA signature (Fig. 3). On the other hand, Deuterostomia (*D. rerio* and *Apostichopus*
220 *japonicus*), Mollusca (*C. gigas*) and Nemertea (*N. geniculatus*) representatives showed a Z-score lower than
221 1. Some species, namely *N. geniculatus*, *A. japonicus*, *C. gigas*, and *S. mediterranea*, also reported a sharp
222 increase of the Z-score for the 10-overlap group for 21-bp long reads only. A 10-bases overlap would
223 correspond to the piRNA signature, although the length of piRNAs in Arthropoda ranges from 25 to 30 pb.
224 Overall, considering the phylogenomic analysis, all the species that exhibited a complete siRNA pathway
225 (using the presence of DICER2 and siAGO as a proxy) showed a high Z-score in the 19-overlap group (Fig.
226 3).

227 Finally, we investigated whether the range of action of the three sncRNA types is uniform among the three
228 main Metazoa superclades, choosing tree representative for each one (i.e., *C. gigas* for Lophotrochozoa, *A.*

229 *gambiae* for Ecdysozoa and *D. rerio* for Deuterostomia). We annotated miRNAs, piRNAs and endo-siRNAs
230 evaluating their expression at different sncRNA lengths (supplementary figure S5A). In terms of reads per
231 million (RPM), in all three species miRNAs are the most expressed and their length ranges from 20 – 25 nc.
232 On the contrary, piRNAs were annotated in the in the 24 – 30 nc length range. In most of the cases, endo-
233 siRNAs were not able to discern the endo-siRNA signal from noise, but *A. gambiae* is the species that showed
234 highest expression levels (supplementary figure S4A). We also compared ovary and somatic tissue sncRNAs
235 libraries. As expected, in all three species piRNAs is the class more expressed in the ovaries. (supplementary
236 figure S5B). Overall, in the three species we did not reported notable differences in terms of sncRNA length
237 or differential expression in somatic/ovarian tissue among piRNA and miRNA types.
238

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 (Höck and Meister 2008; Garcia Silva et al. 2010; 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). It is still uncertain how the four families emerged during eukaryotes evolution. For instance, several eukaryotic clades have a miRNA-like pathway, but it is not clear whether these pathways are analogous or homologous to the metazoan miRNA pathway (Moran et al. 2017). It is likely that RNAi systems diverged from an ancestral siRNA system and, considering the distribution of Eukaryota clades in the four families (Swarts et al. 2014), the divergence took place 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, where only one of the two groups is monophyletic. This pattern is confirmed in other Argonaute phylogenetic analyses (Swarts et al. 2014; Praher et al. 2017; 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 (supplementary figure 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. 2013a). 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; supplementary table 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

274 the three reference species; they are 20-25 nt long, they are not more expressed in the ovaries compared to
275 somatic tissue and overall they are by far the most expressed sncRNA class.
276 Conversely, the endo-siRNA pathway is by far less studied than the miRNA and piRNA ones, which are also
277 more ubiquitous in Metazoa. The absence of siAGO proteins have been reported in Vertebrata (Wynant et al.
278 2017) and in the annelid *Capitella teleta* (Khanal et al. 2022), where the absence of a siRNA signature in
279 sncRNA libraries have been also confirmed. Our analysis on Argonaute and DICER proteins revealed a unique
280 pattern in the evolution of this pathway. DICER2 and siAGO proteins were retrieved in Platyhelminthes and
281 Syndermata. The two proteins were completely absent in Mollusca, Annelida, Phoronida and Brachiopoda. A
282 unique condition was recovered in Nemertea, Bryozoa and Entoprocta; they are equipped with siAGO proteins
283 (albeit not every species), but DICER2 proteins were not annotated in any of those. Thus, we speculate that
284 Nematoda, Bryozoa and Entoprocta are an intermediate state during the loss of the endo-siRNA pathway in
285 Lophotrochozoa. The phylogenetic relationships between lophotrochozoan phyla are not completely resolved
286 yet. However, a phylotranscriptomic analysis has shown that the best ML tree, inferred by reducing the branch-
287 length heterogeneity, places Nemertea in sister relationship with Mollusca + Phoronida + Annelida +
288 Brachiopoda (Kocot et al. 2016), the group of phyla totally lacking the endo-siRNA pathway. The same tree
289 places Bryozoa and Entoprocta as the sister group of Trochozoa (Nemertea + Mollusca + Phoronida + Annelida
290 + Brachiopoda) and the Polyzoa hypothesis (i.e., monophyly of Bryozoa+Entoprocta) has been recently
291 supported (Khalturin et al. 2022). Therefore, the characterization of endo-siRNA pathway in Nemertea is of
292 paramount importance to achieve a complete figure of the evolution of this pathway. Overall, although the
293 relationships between lophotrochozoan phyla are still uncertain, in Nemertea and Polyzoa the presence of
294 siAGO and absence of DICER2 proteins may have a phylogenetic meaning. We hypothesize that the loss of
295 DICER2 took place at the base of Polyzoa+Trochozoa; subsequently, nemerteans and Polyzoa retained siAGO
296 proteins that were left in the last common ancestor of remaining Trochozoa (Fig. 4). This hypothesis would
297 confirm the phylogenetic position of Polyzoa as sisters of Trochozoa, which is inferred but not highly
298 supported in the latest phylogenetic analyses (Kocot et al. 2016).
299 Indeed, species lacking DICER2 showed less sncRNA pairs overlapping for 19 bp, which is the siRNA
300 signature, compared to species with a complete endo-siRNA pathway. The absence of a siRNA signature in
301 the nemertean *N. geniculatus* is in line with our scenario since the typical signature is due to the DICER2
302 cleavage on double-stranded RNAs. To our knowledge, an uncomplete endo-siRNA pathway has never been
303 described, at least among Metazoa. Thus, it is difficult to say whether in Nemertea and Polyzoa the siAGO
304 proteins have been coopted to load a different sncRNA type or they have lost their function in the RNAi
305 pathway after the loss of DICER2. Although the miRNA pathway is highly conserved among Metazoa, many
306 other non-canonical pathways have been discovered, which bypass the Microprocessor or the DICER1 step,
307 but are still loaded by miAGOs (Havens et al. 2012). In the same way, nemerteans and polyzoans siAGO
308 proteins might be involved in pathways that do not require the action of DICER2. Alternatively, DICER2
309 might have been coopted for completely different functions.

310 Concluding, the loss of siAGO proteins and, more in general, of the endo-siRNA pathway possibly in all
311 Trochozoa have been strongly supported by the joint phylogenetic analysis and analysis of sncRNA libraries.
312 The situation is comparable with that obtained from Deuterostomia, where the absence of the canonical (i.e.,
313 involving both DICER and siAGOs) endo-siRNA pathway have already been reported using the lack of
314 annotated siAGO proteins as a proxy (Wynant et al. 2017). Our analysis support this hypothesis, since neither
315 siAGO nor DICER2 homologs were retrieved in deuterostomes (Fig. 1,2). The distribution of the endo-siRNA
316 pathway in Cnidaria (Calcino et al. 2018; Fridrich et al. 2020), Arthropoda and basal Lophotrochozoa (i.e.,
317 Platyhelminthes and Syndermata; Fig. 4) suggests that most trochozoans and deuterostomes lost it secondarily.
318 That said, in mammals it has been demonstrated that endogenous dsRNAs can still be processed and included
319 in the RNAi regulation by proteins of the miRNA pathway (Babiarz et al. 2008; Watanabe et al. 2008; Laura
320 et al. 2020). Thus, a different endo-siRNA pathway seems to have evolved from other pathways. In the same
321 way, it is possible that in most Trochozoa endo-siRNAs might not be completely absent, rather processed by
322 different proteins: however, due to the lack of DICER2, they would not display the typical siRNA signature
323 (Park et al. 2021). As for mammals, immunoprecipitation or knockout experiments in Trochozoa might
324 elucidate whether Argonaute proteins can load other sncRNAs types beside piRNAs and miRNAs.
325 Through the present work we were able to depict the evolution of the main RNAi mechanisms in Metazoa.
326 Unique patterns emerged, such as the uncomplete endo-siRNA pathway in Nemertea and Polyzoa. More efforts
327 should be placed in studying RNAi in less considered phyla, in order to better characterize understudied
328 patterns, casting light on the diversity of RNAi across Metazoa. Comparing different RNAi pathways allowed
329 us to highlight the evolutionary constraints on the miRNA pathway among Metazoa. The conservation and
330 ubiquity of this pathway is likely due to its importance in many regulative systems, and further research should
331 focus on detecting the key features that made the miRNA pathway the main RNAi system.

Materials and Methods

Annotation of Argonaute and DICER proteins

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 (Stanke et al. 2008; Hoff et al. 2019; Brůna et al. 2021). 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). 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 et al. 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 (Camacho et al. 2009; Buchfink et al. 2015) 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 49 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 et al. 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

367 among the best five hits. No ambiguous results (i.e., proteins showing both DICER and DROSHA within the
368 best five hits) were obtained.

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

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

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

386

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

388 Six specimens (three males and three females) of the nemertean Notospermus geniculatus were sampled in
389 June 2018 near Ushimado (Okayama prefecture, Japan). Animals were left in seawater and 7% $MgCl_2 \cdot H_2O$
390 (1:1 ratio) for 15'; gonads were then dissected in 7% $MgCl_2 \cdot H_2O$ on ice and stored in RNAlater (Thermo
391 Fisher Scientific Inc., Waltham, USA), following manufacturer's instructions. Total RNA was extracted using
392 a standard chloforom:TRI Reagent® (Merck KGaA, Darmstadt, Germany) protocol, following manufacturer's
393 instructions. The TruSeq Small RNA library kit (Illumina, San Diego, USA) was used to prepare six sRNA
394 libraries that were sequenced on an Illumina HiSeq2500 platform. Both library preparation and high-
395 throughput sequencing were carried out at the Macrogen Inc. facility (Seoul, South Korea).

396 For this study we sequenced the sncRNA pool from six samples of *N. geniculatus.* These libraries were
397 analyzed along with publicly available sncRNA libraries of other five species. The libraries were selected and
398 downloaded from the Sequence Reads Archive (SRA) provided by NCBI (Table S2). When more samples
399 from the same project were available, libraries were pulled together, obtaining a single fastq file for each
400 species. Adapters and low-quality bases were removed from reads with Cutadapt v3.9.7 (Martin 2011), setting

401 the options -e 0.2 -O 5 --quality-cutoff 6 --discard-untrimmed. Trimmed reads were mapped on the reference
402 genome with Bowtie (Langmead et al. 2009), allowing up to 100 multiple alignments (-m 100). We estimated
403 the distribution of the overlaps between reads with the python script *overlapping_reads.py*
404 (https://github.com/ARTbio/tools-artbio/blob/master/tools/small_rna_signatures/overlapping_reads.py;
405 Antoniewski 2014).

406 For *Crassotrea gigas*, *Danio rerio* and *Anopheles gambiae* we annotated miRNAs, siRNAs and piRNAs. The
407 tool miRDeep2 (Friedländer et al. 2012) was used to predict miRNAs providing the already annotated miRNA
408 set of that species from MiRGeneDB (Fromm et al. 2022) or miRBase (Kozomara et al. 2019) and the
409 annotated miRNAs of up to five closely related species. The novel predicted miRNAs were reconsidered
410 following the rules established by Fromm and colleagues (2015) and discarding novel miRNAs with a STAR
411 sequence coverage lower than 5 reads. Putative siRNA and piRNA pairs were predicted according to reads
412 overlaps (i.e. siRNA pairs overlap of the read length – 2, piRNA pairs overlap of 10 nucleotides) calculated
413 with *overlapping_reads.py.* Pairs of siRNAs and piRNAs were discarded when: one of the paired sRNAs had
414 a coverage lower than 5; the logarithmic ratio of the pair was above 1.5; the pair mapped in a miRNA region.
415 The differential expression of sncRNAs between somatic tissues and ovaries was tested with edgeR (Robinson
416 et al. 2010) using a generalized linear model and a quasi-likelihood F-test (Lund et al. 2012). We chose a strict
417 P-Value threshold of 0.001 to consider a sRNA to be significantly differentially expressed.

418

419 **Data availability**

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

421 **Authors contributions**

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

425 **Acknowledgements**

426 This work was supported by the Canziani Bequest and by the Italian Ministry of University and Research PRIN
427 2020 (2020BE2BC3), both funded to M.P.

428 **Figures captions**

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

434 **Figure 2**. ML tree of the lophotrochozoan DICER proteins. For two marked nodes it is reported respectively
435 the UFBoot and the SH-alrt value. Clades formed by paralogs of the same species were collapsed and
436 represented with a triangle. References species retrieved from the analysis of Mukherjee and colleagues (2013)
437 are marked with an asterisk. Species are colored according to the phylum. The color legend on the bottom left
438 reconstructs the main phylogenetic relationships according to the ML-tree of Kocot and colleagues (2016).

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

444 **Figure 4** The loss of DICER2 and siAGO along the Metazoa evolution. The lophotrochozoan phylogenetic
445 tree is reconstructed according to the latest Lophotrochozoa phylogenetic analyses (Kocot et al. 2016;
446 Khalturin et al. 2022). The position of Polyzoa as sister group of Trochozoa is weakly supported in the literature
447 (red node), but it would be confirmed by the presence of siAGO and absence of DICER2. Colored boxes
448 represent the presence of the related protein in that clade, while crossed boxes represent the absence of the
449 related protein. Each protein is associated with a color, as reported in the legend.

450 **References**

451 Antoniewski C. 2014. Computing siRNA and piRNA Overlap Signatures. In: Animal Endo-SiRNAs. Vol. 1173.
452 New York, NY: Humana Press. p. 135–146.

453 Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R. 2008. Mouse ES cells express endogenous shRNAs,
454 siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev* 22:2773–
455 2785.

456 Bartel DP. 2018. Metazoan MicroRNAs. *Cell* 173:20–51.

457 Bernstein E, Caudy AA, Hammond SM, Hannon GJ. 2001. Role for a bidentate ribonuclease in the initiation
458 step of RNA interference. *Nature* 409:363–366.

459 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.
460 *Bioinformatics* 30:2114–2120.

461 Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete Small RNA-
462 Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* 128:1089–1103.

463 Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome
464 annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom*
465 *Bioinform* 3.

466 Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*
467 12:59–60.

468 Calcino AD, Fernandez-Valverde SL, Taft RJ, Degnan BM. 2018. Diverse RNA interference strategies in early-
469 branching metazoans. *BMC Evol Biol* 18:160.

470 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+:
471 architecture and applications. *BMC Bioinformatics* 10:421.

472 Castresana J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic
473 Analysis. *Mol Biol Evol* 17:540–552.

474 Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA,
475 Sachidanandam R, et al. 2008. An endogenous small interfering RNA pathway in Drosophila. *Nature*
476 453:798–802.

477 Dainat J. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format.

478 Fontenla S, Rinaldi G, Smircich P, Tort JF. 2017. Conservation and diversification of small RNA pathways
479 within flatworms. *BMC Evol Biol* 17:215.

480 Fontenla S, Rinaldi G, Tort JF. 2021. Lost and Found: Piwi and Argonaute Pathways in Flatworms. *Front Cell*
481 *Infect Microbiol* 11.

482 Fridrich A, Modepalli V, Lewandowska M, Aharoni R, Moran Y. 2020. Unravelling the developmental and
483 functional significance of an ancient Argonaute duplication. *Nat Commun* 11:6187.

484 Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. 2012. miRDeep2 accurately identifies known
485 and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res* 40:37–52.

486 Fromm B, Høye E, Domanska D, Zhong X, Aparicio-Puerta E, Ovchinnikov V, Umu SU, Chabot PJ, Kang W,
487 Aslanzadeh M, et al. 2022. MirGeneDB 2.1: toward a complete sampling of all major animal phyla.
488 *Nucleic Acids Res* 50:D204–D210.

489 Galiana-Arnoux D, Dostert C, Schneemann A, Hoffmann JA, Imler J-L. 2006. Essential function in vivo for
490 Dicer-2 in host defense against RNA viruses in drosophila. *Nat Immunol* 7:590–597.

491 Gammon DB, Mello CC. 2015. RNA interference-mediated antiviral defense in insects. *Curr Opin Insect Sci*
492 8:111–120.

493 Garcia Silva MR, Tosar JP, Frugier M, Pantano S, Bonilla B, Esteban L, Serra E, Rovira C, Robello C, Cayota A.
494 2010. Cloning, characterization and subcellular localization of a Trypanosoma cruzi argonaute protein
495 defining a new subfamily distinctive of trypanosomatids. *Gene* 466:26–35.

496 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng
497 Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome.
498 *Nat Biotechnol* 29:644–652.

499 Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP.
500 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature*
501 455:1193–1197.

502 Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New Algorithms and Methods
503 to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol*
504 59:307–321.

505 Havens MA, Reich AA, Duelli DM, Hastings ML. 2012. Biogenesis of mammalian microRNAs by a non-
506 canonical processing pathway. *Nucleic Acids Res* 40:4626–4640.

507 Hirakata S, Siomi MC. 2016. piRNA biogenesis in the germline: From transcription of piRNA genomic sources
508 to piRNA maturation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1859:82–92.

509 Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast
510 Bootstrap Approximation. *Mol Biol Evol* 35:518–522.

511 Höck J, Meister G. 2008. The Argonaute protein family. *Genome Biol* 9:210.

512 Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-Genome Annotation with BRAKER. In: p. 65–
513 95.

514 Huntzinger E, Izaurralde E. 2011. Gene silencing by microRNAs: contributions of translational repression
515 and mRNA decay. *Nat Rev Genet* 12:99–110.

516 Iwakawa H, Tomari Y. 2022. Life of RISC: Formation, action, and degradation of RNA-induced silencing
517 complex. *Mol Cell* 82:30–43.

518 Jehn J, Gebert D, Pipilescu F, Stern S, Kiefer JST, Hewel C, Rosenkranz D. 2018. PIWI genes and piRNAs are
519 ubiquitously expressed in mollusks and show patterns of lineage-specific adaptation. *Commun Biol*
520 1:137.

521 Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in
522 Performance and Usability. *Mol Biol Evol* 30:772–780.

523 Khalturin K, Shunatova N, Shchenkov S, Sasakura Y, Kawamitsu M, Satoh N. 2022. Polyzoa is back: The
524 effect of complete gene sets on the placement of Ectoprocta and Entoprocta. *Sci Adv* 8.

525 Khanal S, Zancanela BS, Peter JO, Flynt AS. 2022. The Small RNA Universe of Capitella teleta. *Front Mol*
526 *Biosci* 9.

527 Kocot KM, Struck TH, Merkel J, Waits DS, Todt C, Brannock PM, Weese DA, Cannon JT, Moroz LL, Lieb B, et
528 al. 2016. Phylogenomics of Lophotrochozoa with Consideration of Systematic Error. *Syst Biol*:syw079.

529 Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to function. *Nucleic*
530 *Acids Res* 47:D155–D162.

531 Kuhn C-D, Joshua-Tor L. 2013. Eukaryotic Argonautes come into focus. *Trends Biochem Sci* 38:263–271.

532 Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA
533 sequences to the human genome. *Genome Biol* 10:R25.

534 Laura S, Srividya C, Vidigal JA. 2020. AGO unchained Canonical and non-canonical roles of Argonaute
535 proteins in mammals. *Frontiers in Bioscience* 25:4793.

536 Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Rådmark O, Kim S, et al. 2003. The nuclear RNase
537 III Drosha initiates microRNA processing. *Nature* 425:415–419.

538 Lee Y, Kim M, Han J, Yeom K-H, Lee S, Baek SH, Kim VN. 2004. MicroRNA genes are transcribed by RNA
539 polymerase II. *EMBO J* 23:4051–4060.

540 Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, Carthew RW. 2004. Distinct Roles for Drosophila
541 Dicer-1 and Dicer-2 in the siRNA/miRNA Silencing Pathways. *Cell* 117:69–81.

542 Lund SP, Nettleton D, McCarthy DJ, Smyth GK. 2012. Detecting Differential Expression in RNA-sequence
543 Data Using Quasi-likelihood with Shrunken Dispersion Estimates. *Stat Appl Genet Mol Biol* 11.

544 Malone CD, Hannon GJ. 2009. Small RNAs as Guardians of the Genome. *Cell* 136:656–668.

545 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*
546 17:10.

547 Matranga C, Zamore PD. 2007. Small silencing RNAs. *Current Biology* 17:R789–R793.

548 Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S,
549 Richardson LJ, et al. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Res* 49:D412–
550 D419.

551 Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. 2013. Challenges in homology search: HMMER3 and
552 convergent evolution of coiled-coil regions. *Nucleic Acids Res* 41:e121–e121.

553 Mohn F, Handler D, Brennecke J. 2015. piRNA-guided slicing specifies transcripts for Zucchini-dependent,
554 phased piRNA biogenesis. *Science (1979)* 348:812–817.

555 Moran Y, Agron M, Praher D, Technau U. 2017. The evolutionary origin of plant and animal microRNAs. *Nat*
556 *Ecol Evol* 1:0027.

557 Moran Y, Praher D, Fredman D, Technau U. 2013. The Evolution of MicroRNA Pathway Protein Components
558 in Cnidaria. *Mol Biol Evol* 30:2541–2552.

559 Mukherjee K, Campos H, Kolaczkowski B. 2013a. Evolution of Animal and Plant Dicers: Early Parallel
560 Duplications and Recurrent Adaptation of Antiviral RNA Binding in Plants. *Mol Biol Evol* 30:627–641.

561 Mukherjee K, Campos H, Kolaczkowski B. 2013b. Evolution of Animal and Plant Dicers: Early Parallel
562 Duplications and Recurrent Adaptation of Antiviral RNA Binding in Plants. *Mol Biol Evol* 30:627–641.

563 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm
564 for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* 32:268–274.

565 Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. 2019. PIWI-interacting RNAs: small RNAs with big
566 functions. *Nat Rev Genet* 20:89–108.

567 Park S, Kang I, Shin C. 2021. MicroRNA clustering on the biogenesis of suboptimal microRNAs. *Appl Biol*
568 *Chem* 64:51.

569 Praher D, Zimmermann B, Genikhovich G, Columbus-Shenkar Y, Modepalli V, Aharoni R, Moran Y, Technau
570 U. 2017. Characterization of the piRNA pathway during development of the sea anemone
571 *Nematostella vectensis*. *RNA Biol* 14:1727–1741.

572 Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression
573 analysis of digital gene expression data. *Bioinformatics* 26:139–140.

574 Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, et al. 2022.
575 Database resources of the national center for biotechnology information. *Nucleic Acids Res* 50:D20–
576 D26.

577 Shabalina S, Koonin E. 2008. Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol* 23:578–
578 587.

579 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015. BUSCO: assessing genome
580 assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.

581 Song J-J, Smith SK, Hannon GJ, Joshua-Tor L. 2004. Crystal Structure of Argonaute and Its Implications for
582 RISC Slicer Activity. *Science (1979)* 305:1434–1437.

583 Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA
584 alignments to improve de novo gene finding. *Bioinformatics* 24:637–644.

585 Strassert JFH, Irisarri I, Williams TA, Burki F. 2021. A molecular timescale for eukaryote evolution with
586 implications for the origin of red algal-derived plastids. *Nat Commun* 12:1879.

587 Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin E V, Patel DJ, van der Oost J. 2014. The
588 evolutionary journey of Argonaute proteins. *Nat Struct Mol Biol* 21:743–753.

589 Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R,
590 Schultz RM, et al. 2008. Pseudogene-derived small interfering RNAs regulate gene expression in
591 mouse oocytes. *Nature* 453:534–538.

592 Tarver JE, Sperling EA, Nailor A, Heimberg AM, Robinson JM, King BL, Pisani D, Donoghue PCJ, Peterson KJ.
593 2013. miRNAs: Small Genes with Big Potential in Metazoan Phylogenetics. *Mol Biol Evol* 30:2369–
594 2382.

595 Thibaud-Nissen F, DiCuccio M, Hlavina W, Kimchi A, Kitts PA, Murphy TD, Pruitt KD, Souvorov A. 2016.
596 P8008 The NCBI Eukaryotic Genome Annotation Pipeline. *J Anim Sci* 94:184–184.

597 Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Tóth KF. 2013.
598 Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state.
599 *Genes Dev* 27:390–399.

600 Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T,
601 Nakano T, et al. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in
602 mouse oocytes. *Nature* 453:539–543.

603 Weick E-M, Miska EA. 2014. piRNAs: from biogenesis to function. *Development* 141:3458–3471.

604 Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2015. RELAX: Detecting Relaxed
605 Selection in a Phylogenetic Framework. *Mol Biol Evol* 32:820–832.

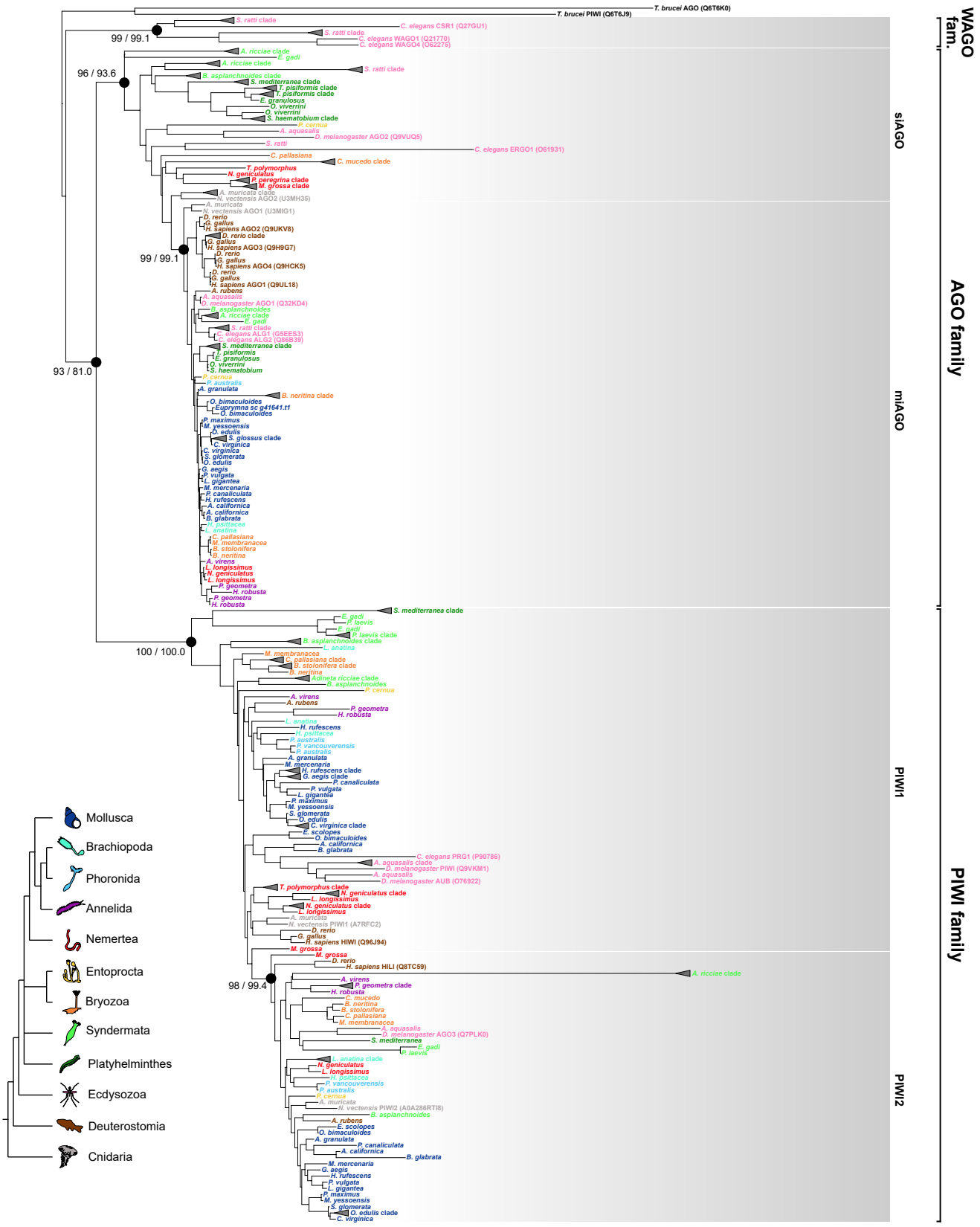
606 Wheeler BM, Heimberg AM, Moy VN, Sperling EA, Holstein TW, Heber S, Peterson KJ. 2009. The deep
607 evolution of metazoan microRNAs. *Evol Dev* 11:50–68.

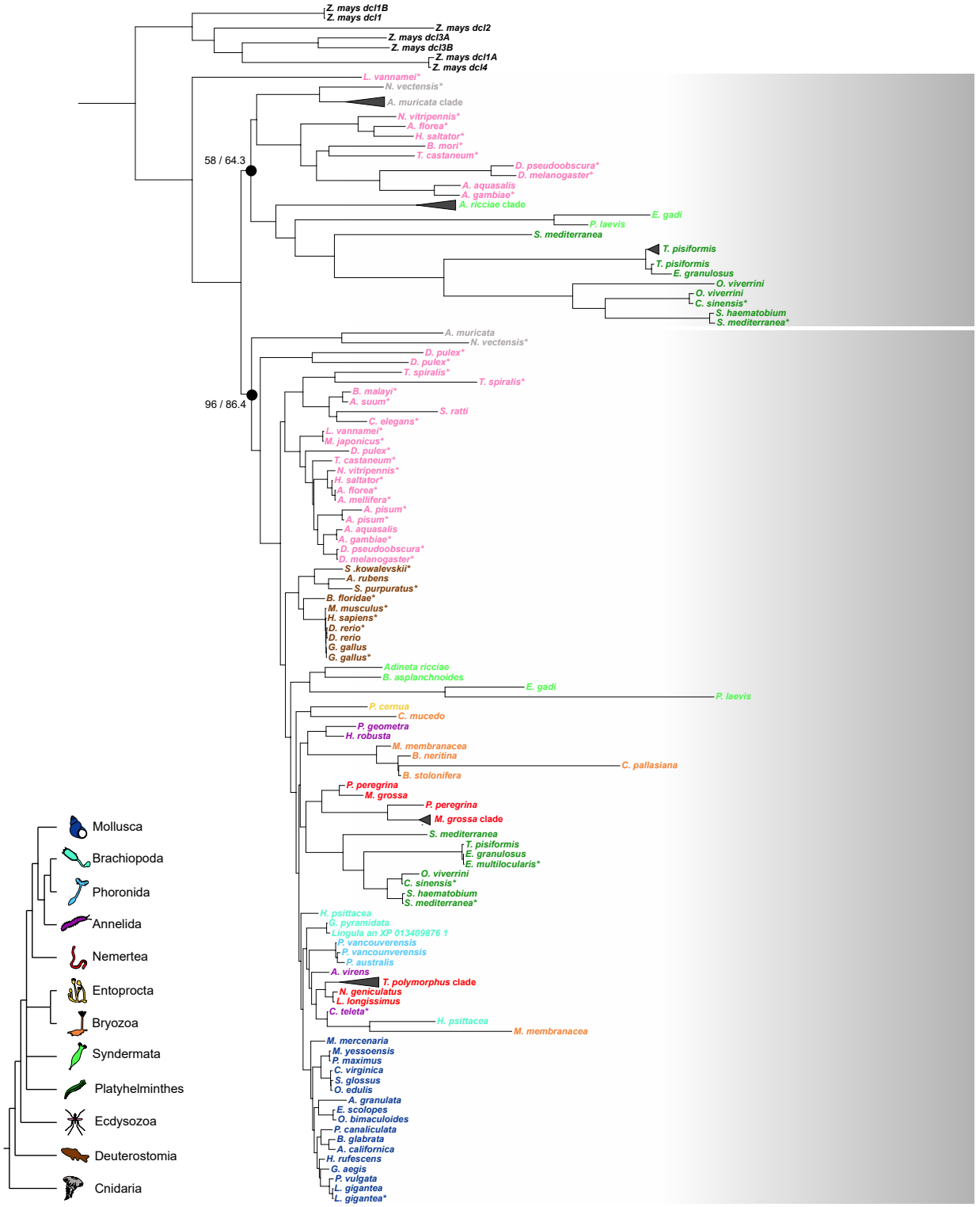
608 Wu J, Yang J, Cho WC, Zheng Y. 2020. Argonaute proteins: Structural features, functions and emerging
609 roles. *J Adv Res* 24:317–324.

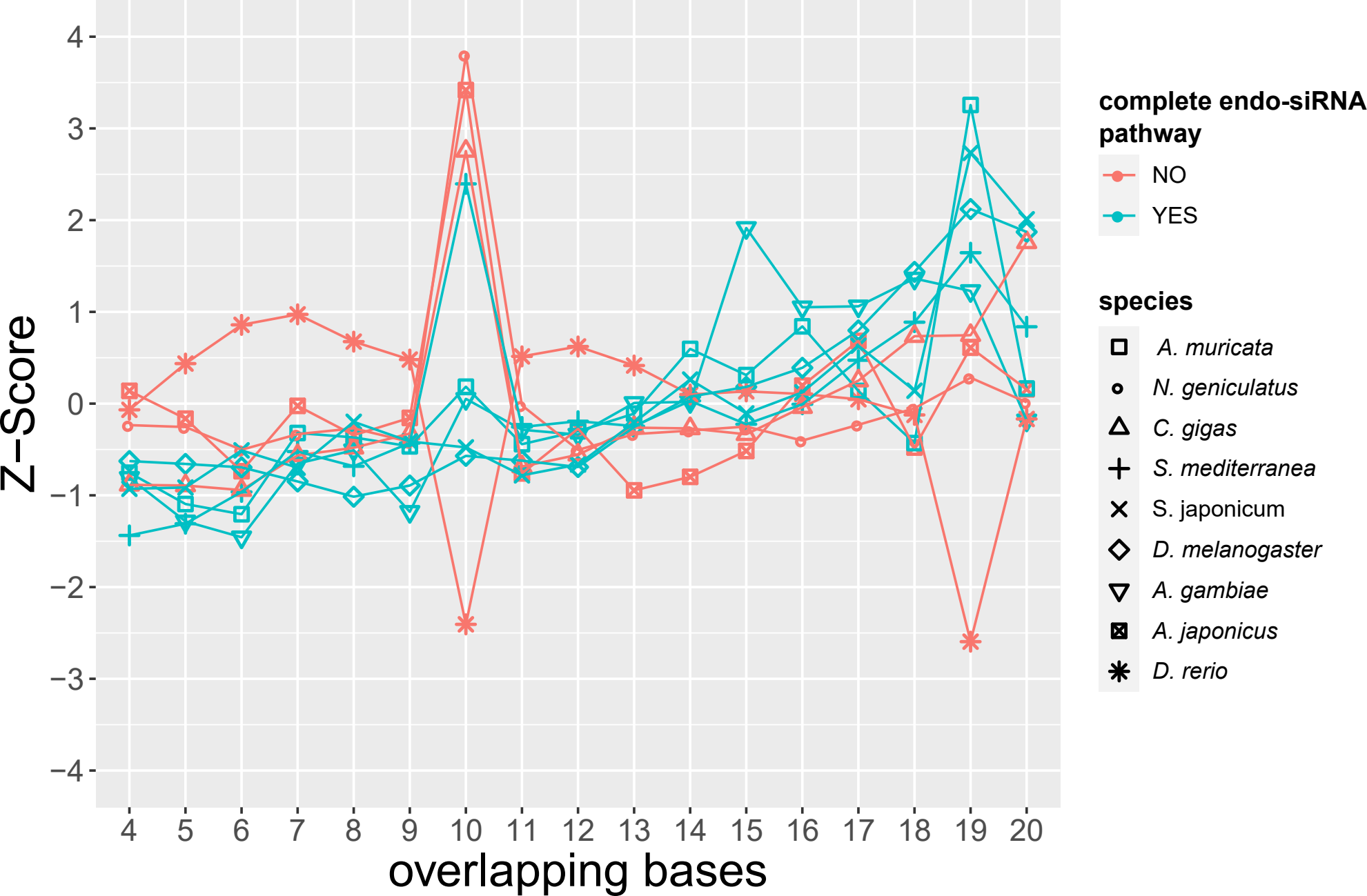
610 Wynant N, Santos D, Vanden Broeck J. 2017. The evolution of animal Argonautes: evidence for the absence
611 of antiviral AGO Argonautes in vertebrates. *Sci Rep* 7:9230.

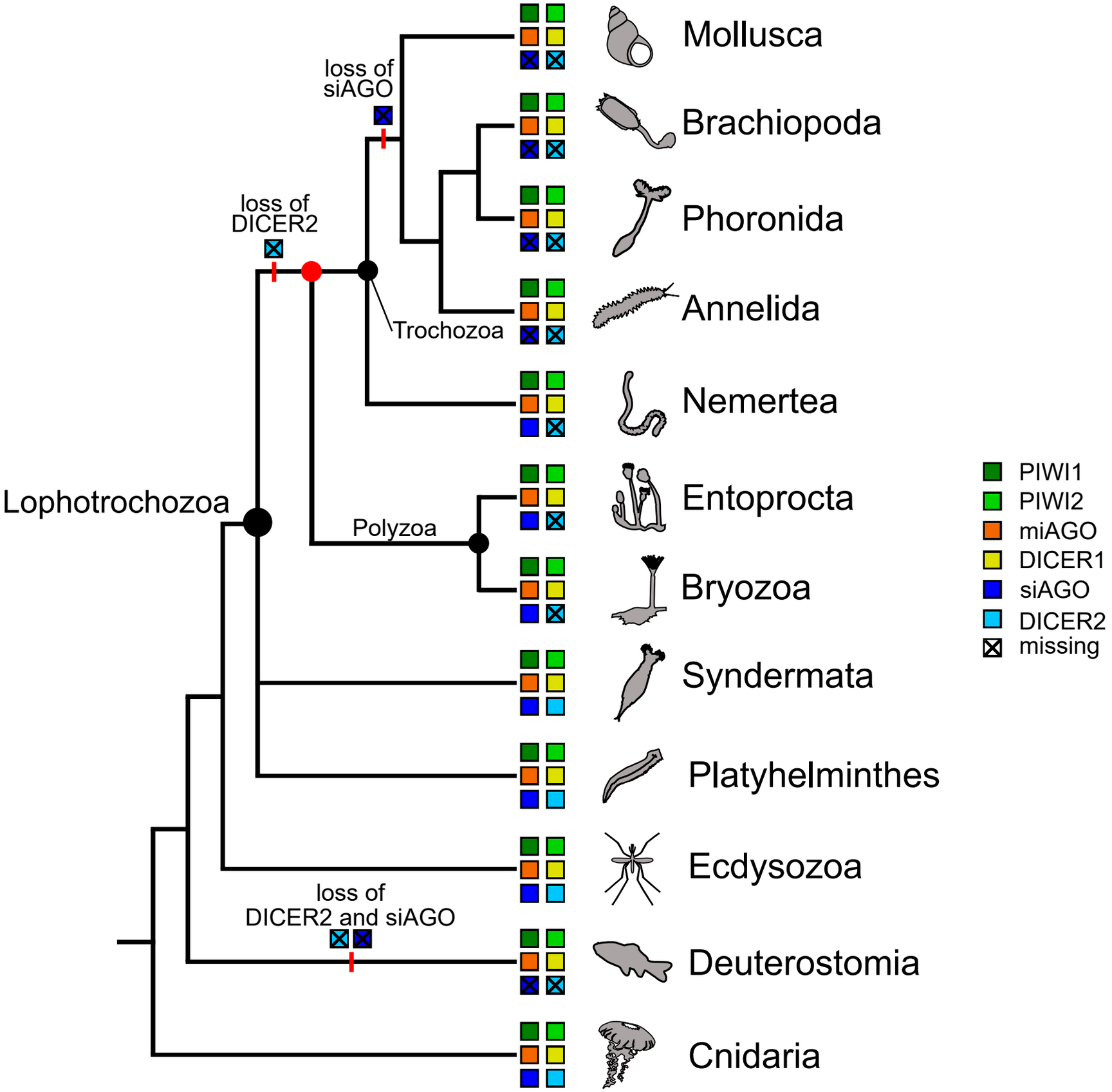
612 Zdobnov EM, Kuznetsov D, Tegenfeldt F, Manni M, Berkeley M, Kriventseva E V. 2021. OrthoDB in 2020:
613 evolutionary and functional annotations of orthologs. *Nucleic Acids Res* 49:D389–D393.

614


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 (Q62275)
99 / 99.1
A. ricciae clade
E. gadi
A. ricciae clade
S. ratti clade
96 / 93.6
B. asplanchnoides clade
S. mediterranea clade
T. pisiformis clade
T. pisiformis clade
E. granulosus
O. viverrini
O. viverrini
S. haematobium clade
P. cernua
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)
99 / 99.1
A. rubens
A. aquasalis
D. melanogaster AGO1 (Q32KD4)
B. asplanchnoides
A. ricciae clade
E. gadi
S. ratti clade
C. elegans ALG1 (G5EES3)
C. elegans ALG2 (Q86B39)
S. mediterranea clade
T. pisiformis
E. granulosus
O. viverrini
S. haematobium
P. cernua
P. australis
A. granulata
B. neritina clade
O. bimaculoides
Euprymna sc g41641.t1
O. bimaculoides
P. maximus
M. yessoensis
O. edulis
S. glossus clade
C. virginica
C. virginica
S. glomerata
O. edulis
G. aegis
P. vulgata
L. gigantea
M. mercenaria
P. canaliculata
H. rufescens
A. californica
A. californica
B. glabrata
H. patiaceа
L. anatina
C. pallasiana
M. membranacea
B. stolonifera
B. neritina
A. virens
L. longissimus
N. geniculatus
L. longissimus
P. geometra
H. robusta
P. geometra
H. robusta
93 / 81.0
S. mediterranea clade
E. gadi
P. laevis
E. gadi
P. laevis clade
B. asplanchnoides clade
100 / 100.0
L. anatina
M. membranacea
C. pallasiana clade
B. stolonifera clade
B. neritina
Adineta ricciae clade
B. asplanchnoides
P. cernua
A. virens
A. rubens
P. geometra
H. robusta
L. anatina
H. rufescens
H. patiaceа
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
O. 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)
A. ricciae clade
98 / 99.4
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. anatina clade
N. geniculatus
L. longissimus
H. patiaceа
P. vancouverensis
P. australis
P. cernua
A. muricata
N. vectensis PIWI2 (A0A286RTI8)
B. asplanchnoides
A. rubens
E. scolopes
O. bimaculoides
A. granulata
P. canaliculata
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
loss of siAGO
loss of DICER2
Trochozoa
Lophotrochozoa
Polyzoa
loss of DICER2 and siAGO
PIWI1
PIWI2
miAGO
DICER1
siAGO
DICER2
missing