

The Evolution and Characterization of the RNA Interference Pathways in Lophotrochozoa

Alessandro Formaggioni^{1,†}, Gianmarco Cavalli^{1,†}, Mayuko Hamada², Tatsuya Sakamoto², Federico Plazzi ^{1,*}, and Marco Passamonti ¹

¹Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy

²Ushimado Marine Institute, Okayama University, Okayama, Japan

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: federico.plazzi@unibo.it

Accepted: April 30, 2024

Abstract

In animals, three main RNA interference mechanisms have been described so far, which respectively mature 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 typically loaded by a specific Argonaute homolog. Moreover, other protein families play pivotal roles in the maturation of sncRNAs, like the DICER ribonuclease family, whose DICER1 and DICER2 paralogs mature respectively miRNAs and endo-siRNAs. Within 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 suggested that in all Trochozoa, the proteins related to the endo-siRNA pathway have been lost, a part of them in some phyla (i.e. Nemertea, Bryozoa, Entoprocta), while all of them in all the others. 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.

Key words: Metazoa, Lophotrochozoa, Argonaute, DICER, RNAi.

Significance

In animals, three main RNA interference mechanisms have been described so far, called miRNAs, piRNAs, and endo-siRNAs, along with proteins involved in their maturation; among animals, the distribution of these pathways has been only studied in major groups and knowledge from many phyla is poor. We inferred the evolutionary history of the proteins maturing these RNAs, namely the animal Argonaute and DICER families, including 43 lophotrochozoan species. We describe the loss of the endo-siRNA pathway during the evolution of Lophotrochozoa: some phyla exhibit the complete pathway, while other phyla show a partial processing machinery or no machinery at all—contrastingly, miRNAs were revealed the most conserved and ubiquitous mechanism of the animal RNA interference machinery.

© The Author(s) 2024. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Argonaute proteins are cytoplasmic proteins that play a key role in most of the RNA interference (RNAi) pathways. They interact with a small noncoding 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 through 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 euglenozoan 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 (Lee, Kim, 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 2 bp 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; Lee, Nakahara, et al. 2004). Through its catalytic activity, DICER1 removes the hairpin's loop, leaving a ~22 bp 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' untranslated region (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). 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. DICER2 cleaves a ~21 bp dsRNA that is loaded into an AGO-like Argonaute, called AGO2 in fruit flies (Lee, Nakahara, 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).

In insects, the siRNA-mediated RNAi activity is not restricted to endogenous dsRNAs, but DICER2 is also able to target dsRNAs of viral origin, producing exogenous siRNAs that are pivotal for the innate immune response (Schuster et al. 2019). *Caenorhabditis elegans* expresses a single DICER paralog, which processes endogenous and viral dsRNAs, but also primary miRNA structures (Welker et al. 2011). Nematodes are even capable of producing secondary siRNAs thanks to the RNA-dependent RNA polymerases (Matranga and Zamore 2007), which are absent in mammals and insects. In mammals, viral dsRNAs are targeted by RIG-I-Like receptors (Loo and Gale 2011), which induce an antiviral response through the activation of type I interferons (Isaacs et al. 1963; Schuster et al. 2019). The ability to process long dsRNAs seems to be related to the DICER helicase domain, which is functional in *C. elegans* DICER and insects' DICER2 (Welker et al. 2011; Sinha et al. 2018; Aderounmu et al. 2023). Contrastingly, the helicase function is not required for DICER proteins that are mainly involved in miRNAs maturation (i.e. insects' DICER1 and mammals' DICER; Jiang et al. 2005; Aderounmu et al. 2023).

Finally, piRNAs are 24 to 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

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 anti-sense piRNAs (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; Hiraoka 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).

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 animals (Fromm et al. 2022).

RNAi mechanisms are far less studied in Lophotrochozoa than in Deuterostomia and Ecdysozoa. This superclade includes around 14 phyla, which is notably higher than Ecdysozoa and Deuterostomia, which comprises respectively eight and three phyla (Brusca et al. 2016). Moreover, Lophotrochozoa show an astonishing variability in body plans. For medical and nutritional reasons, most of the data are restricted to Mollusca and Platyhelminthes. Parasitic Platyhelminthes (i.e. Neodermata) lack 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 a clear signature of ping-pong amplification is visible (Jehn et al. 2018), and many miRNAs have been annotated in mollusks (Fromm et al. 2022). On the other hand, proteins related to the endo-siRNA pathway are absent in Bivalvia (Rosani et al. 2016). 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 families is restricted to one Syndermata, one Brachiopoda, and two Annelida species (Fromm et al. 2022).

The phylogenetic relationships within Lophotrochozoa have been strongly debated, and they are not fully resolved yet. Morphological analyses agree to include Mollusca, Annelida, Brachiopoda, Phoronida, Nemertea, Bryozoa, Entoprocta, and Cyclophora in the Trochozoa clade (f.i.,

Kocot 2016). Some molecular analyses proved to be concordant with the Trochozoa clade (Struck et al. 2014; Laumer et al. 2015; Kocot et al. 2017; Laumer et al. 2019), but some others did not (Kocot et al. 2017; Marlétaz et al. 2019). Even the relationships among Trochozoa are far from being resolved, namely the monophyly of Polyzoa (Bryozoa + Entoprocta + Cyclophora) has regained credit recently (Khalturin et al. 2022; but see (Nesnidal et al. 2013; Kocot 2016; Bleidorn 2019; Laumer et al. 2019). Outside Trochozoa, Roussozoa (Platyhelminthes + Gastrotricha) and Chaetognathifera (Syndermata + Micrognathozoa + Gnathostomulida + Chaetognatha) are supported by most of the phylogenetic analyses (Struck et al. 2014; Laumer et al. 2015; Kocot et al. 2017; Laumer et al. 2019; Conway Morris et al. 2020).

In recent years, the increase in -omics data has made it 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. According to our results, along the Lophotrochozoa evolution, the endo-siRNA pathway has been progressively lost, starting with DICER2 in Trochozoa, followed by the loss of the fruit fly AGO2-like proteins in Phoronida, Brachiopoda, Annelida, and Mollusca. This pattern is confirmed by the distribution of DICER2 and AGO2-like proteins in the analyzed organisms. In contrast, the piRNA and miRNA pathways appeared to be conserved in almost all Lophotrochozoa.

Results

The Argonaute and DICER Phylogeny

We annotated Argonaute proteins of 43 lophotrochozoan species by analyzing 19 proteomes, 16 genomes, and 8 transcriptomes. Argonaute sequences of *Homo sapiens* (Chordata), *Drosophila melanogaster* (Arthropoda), *C. elegans* (Nematoda), and *Nematostella vectensis* (Cnidaria) were retrieved from SwissProt and included in the phylogenetic analysis as references (supplementary table S1, Supplementary Material online). Moreover, we included metazoan species whose RNAi pathways have already been studied (see Introduction), testing whether our Argonaute and DICER annotation matches results from the literature (Table 1).

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 (supplementary table S1, Supplementary Material online). The phylogenetic tree of the Argonaute superfamily supported the known

Table 1

Presence and absence of Argonaute and DICER proteins.

Phylum	Species	PIWI1	PIWI2	miAGO	siAGO	DICER1	DICER2	Data	Comp.
Mollusca ●	<i>Haliotis rufescens</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	99,4%
	<i>Mizuhopecten yessoensis</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	98,6%
	<i>Gigantopelta aegis</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	98,5%
	<i>Pecten maximus</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	98,5%
	<i>Pomacea canaliculata</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	98,2%
	<i>Crassostrea virginica</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	98,1%
	<i>Aplysia californica</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	97,8%
	<i>Ostrea edulis</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	96,8%
	<i>Lottia gigantea</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	96,5%
	<i>Mercenaria mercenaria</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	95,4%
	<i>Octopus bimaculoides</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	94,6%
	<i>Patella vulgata</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	90,5%
	<i>Biomphalaria glabrata</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	88,9%
	<i>Saccostrea glomerata</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	88,9%
<i>Acanthopleura granulata</i>	✓	✓	✓	✗	✓	✗	Assemb.	69,8%	
<i>Euprymna scolopes</i>	✓	✓	✓	✗	✓	✗	Assemb.	55,5%	
Annelida ●	<i>Alitta virens</i>	✓	✓	✓	✗	✓	✗	Assemb.	69,8%
	<i>Piscicola geometra</i>	✓	✓	✓	✗	✓	✗	Assemb.	43,7%
	<i>Helobdella robusta</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	90,2%
Brachiopoda ●	<i>Lingula anatina</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	98,6%
	<i>Hemithiris psittacea</i>	✓	✓	✓	✗	✓	✗	Transc.	90,7%
	<i>Glottidia pyramidata</i>	✗	✗	✗	✗	✓	✗	Transc.	75,1%
Phoronida ●	<i>Phoronis australis</i>	✓	✓	✓	✗	✓	✗	Assemb.	68,3%
	<i>Phoronis vancouverensis</i>	✓	✓	✗	✗	✓	✗	Transc.	84,9%
Nemertea ●	<i>Notospermus geniculatus</i>	✓	✓	✓	✓	✓	✗	Assemb.	68,4%
	<i>Lineus longissimus</i>	✓	✓	✓	✗	✓	✗	Assemb.	69,9%
	<i>Tubulanus polymorphus</i>	✓	✗	✗	✓	✓	✗	Transc.	71,8%
	<i>Malacobdella grossa</i>	✓	✓	✗	✓	✓	✗	Transc.	91,7%
	<i>Paranemertes peregrina</i>	✗	✗	✗	✓	✓	✗	Transc.	80,5%
Bryozoa ●	<i>Cryptosula pallasiana</i>	✓	✓	✓	✓	✓	✗	Assemb.	83,5%
	<i>Cristatella mucedo</i>	✗	✓	✗	✓	✓	✗	Assemb.	62,1%
	<i>Membranipora membranacea</i>	✓	✓	✓	✗	✓	✗	Assemb.	54,5%
	<i>Bugulina stolonifera</i>	✓	✗	✓	✗	✓	✗	Assemb.	54,0%
	<i>Bugula neritina</i>	✓	✗	✓	✗	✓	✗	Assemb.	50,7%
Entoprocta ●	<i>Pedicellina cernua</i>	✓	✓	✓	✓	✓	✗	Transc.	90,4%
Syndermata ●	<i>Adineta ricciae</i>	✓	✓	✓	✓	✓	✓	Assemb.	79,6%
	<i>Pomphorhynchus laevis</i>	✓	✓	✗	✗	✓	✓	Assemb.	43,2%
	<i>Brachionus asplanchnoidis</i>	✓	✓	✓	✓	✓	✗	Assemb.	85,4%
	<i>Echinorhynchus gadi</i>	✓	✓	✓	✓	✓	✓	Transc.	58,6%
Platyhelminthes ●	<i>Echinococcus granulosus</i>	✗	✗	✓	✓	✓	✓	NCBI Ref.	69,2%
	<i>Schistosoma haematobium</i>	✗	✗	✓	✓	✓	✓	NCBI Ref.	77,1%
	<i>Opisthorchis viverrini</i>	✗	✗	✓	✓	✓	✓	NCBI Ref.	64,7%
	<i>Taenia pisiformis</i>	✗	✗	✓	✓	✓	✓	Assemb.	36,3%
	<i>Schmidtea mediterranea</i>	✓	✓	✓	✓	✓	✓	Assemb.	46,5%
others	<i>Gallus gallus</i>	✓	✗	✓	✗	✓	✗	NCBI Ref.	95,3%
	<i>Danio rerio</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	99,6%
	<i>Asterias rubens</i>	✓	✓	✓	✗	✓	✗	NCBI Ref.	98,7%
	<i>Anopheles aquasalis</i>	✓	✓	✓	✓	✓	✓	NCBI Ref.	99,0%
	<i>Strongyloides ratti</i>	✗	✗	✓	✓	✓	✗	NCBI Ref.	69,2%
	<i>Acropora muricata</i>	✓	✓	✓	✓	✓	✓	Assemb.	60,4%

Colored dots refer to colors assigned to the different animal phyla in Figs. 1 and 2. For the four Argonaute proteins and the two DICER proteins, a green check marks the species where the protein has been annotated, and a red cross marks species where the protein is absent. The "Data" column reports the type of data, proteome annotated following the NCBI Eukaryotic Genome Annotation Pipeline ("NCBI Ref."), a genome assembly ("Assemb."), or a transcriptome ("Transc."). The last column reports the BUSCO completeness score.

main families (Fig. 1; see [supplementary fig. S1, Supplementary Material](#) online, for the uncollapsed tree with support values for each node). 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 to 4 of *H. sapiens*, and ALG1,2 of *C. elegans*. Every family was widely supported by ultrafast bootstrap approximation (UFBoot) and the SH-like approximate likelihood ratio test (SH-*alrt*) (Fig. 1).

Within each family, it was possible to identify the different Argonaute proteins. Argonaute proteins related to miRNAs were characterized by proteins like *H. sapiens* AGO1 to 4, *D. melanogaster* and *N. vectensis* AGO1, and *C. elegans* ALG1,2 and resulted in a monophyletic clade (UFBoot = 99 and SH-*alrt* = 99.1; Fig. 1). Almost all lophotrochozoan species featured a protein clustering within this clade, with very few exceptions, and at least one organism from each phylum was recovered in this clade (Fig. 1; [Table 1](#)). We will refer to this clade as the “miAGO clade.” The remaining proteins within the AGO-like family were recovered as paraphyletic with respect to the miAGO clade. This group included *D. melanogaster* AGO2, *N. vectensis* AGO2, and *C. elegans* ERGO, all proteins that target endo-siRNAs (but with some exceptions: see [Fridrich et al. 2020](#)). Actually, endo-siRNA proteins are often inferred as paraphyletic with respect to the miAGO clade ([Swarts et al. 2014](#); [Praher et al. 2017](#); [Wynant et al. 2017](#)); we will 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, and Phoronida (Fig. 1; [Table 1](#)).

PIWI-like proteins showed a pattern similar to that of the AGO-like family. PIWI2 proteins were clustered in a monophyletic clade, characterized by already annotated PIWI2 proteins, like *H. sapiens* HILI, *N. vectensis* PIWI2, and *D. melanogaster* AGO3; all the remaining PIWI-like proteins are paraphyletic with respect to PIWI2 proteins. This grade included already annotated PIWI1 proteins like *H. sapiens* HIWI, *N. vectensis* PIWI1, *C. elegans* PRG1, and *D. melanogaster* AUB and PIWI. Both PIWI-like protein groups included at least one protein from each lophotrochozoan phylum ([Table 1](#)). PIWI-like proteins were almost absent in Platyhelminthes, apart from *Schmitdea mediterranea* (Fig. 1).

Regarding the six nonlophotrochozoan species included in the analysis, the annotation of Argonaute proteins matched the expectations: miAGO proteins were retrieved

for all six species; siAGO proteins were absent in *Danio rerio*, *Gallus gallus*, and *Asterias rubens* (Deuterostomia); in all six species, we annotated PIWI1 and PIWI2 proteins, exception made for *Strongyloides ratti* (i.e. that lacked of both PIWI proteins) and, unexpectedly, *G. gallus* that lacked of PIWI2 (Fig. 1; [Table 1](#)).

The phylogenetic analysis highlighted the presence of miAGO proteins in each phylum, but only some of them featured siAGO proteins. In Arthropoda, the precursor 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, such as Cnidaria and Platyhelminthes ([Mukherjee et al. 2013](#)), 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. 2013](#)) 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; see [supplementary fig. S2, Supplementary Material](#) online, for the uncollapsed tree with support values for each node). The resulting ML tree clustered the DICER proteins into two distinct groups. Recall the position of the reference sequences, DICER1 and DICER2 sequences, are accordingly split into 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). In contrast, 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 a 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; [Table 1](#)). In line with expectations, DICER2 was found in *Acropora muricata* and *Anopheles gambiae*, while it was lacking in *D. rerio*, *G. gallus*, *A. rubens*, and *S. ratti*. In contrast, 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

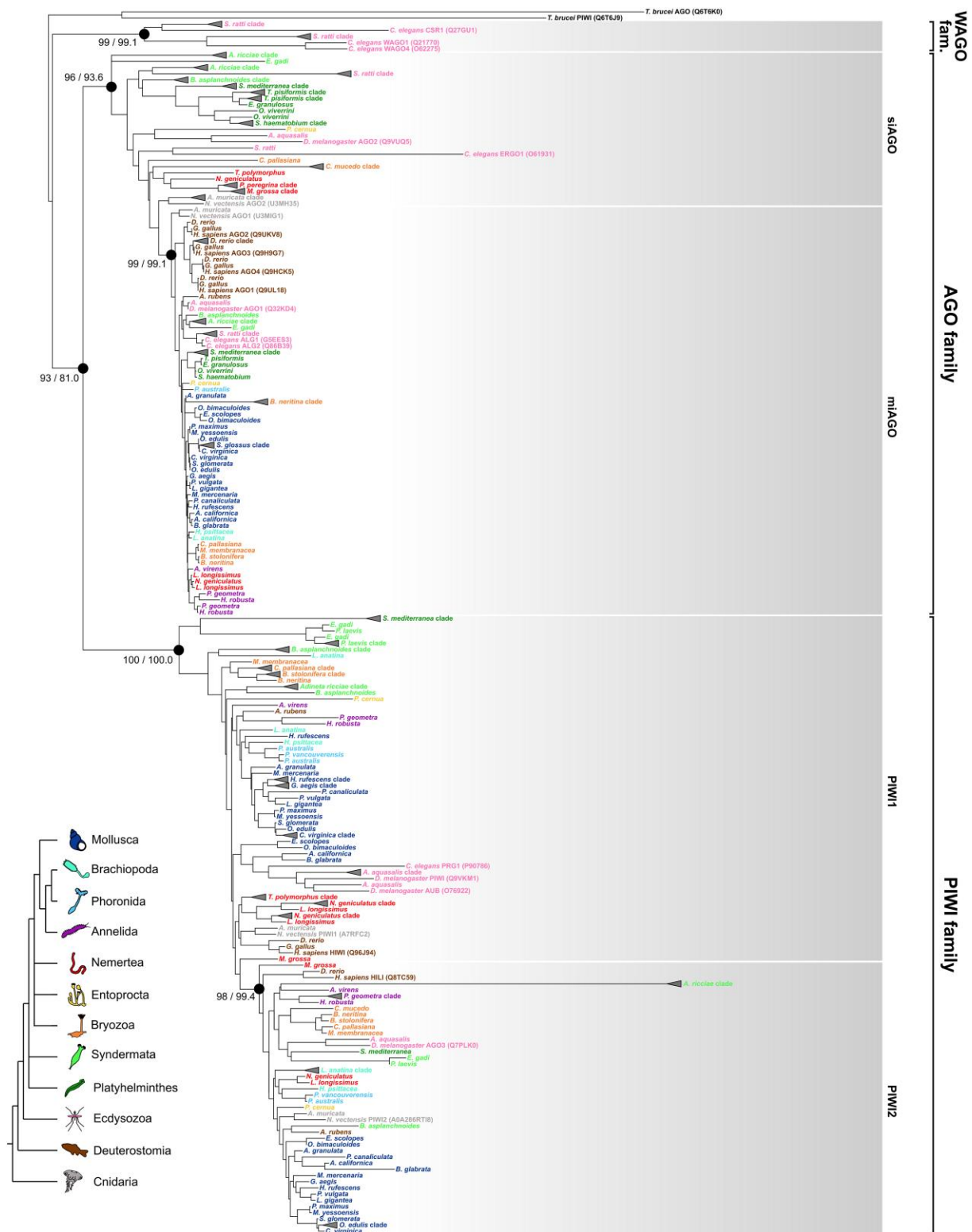


FIG. 1.—ML tree of the lophotrochozoan Argonaute proteins. For the six marked nodes, the label shows UFB00t/SH-ai1t value. Support values of the remaining nodes are shown in [supplementary fig. S1, Supplementary Material](#) online. Clades formed by paralogs of the same species were collapsed and represented 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 latest Lophotrochozoa phylogenetic analyses (Kocot et al. 2017; Bleidorn 2019; Marlétaz et al. 2019).

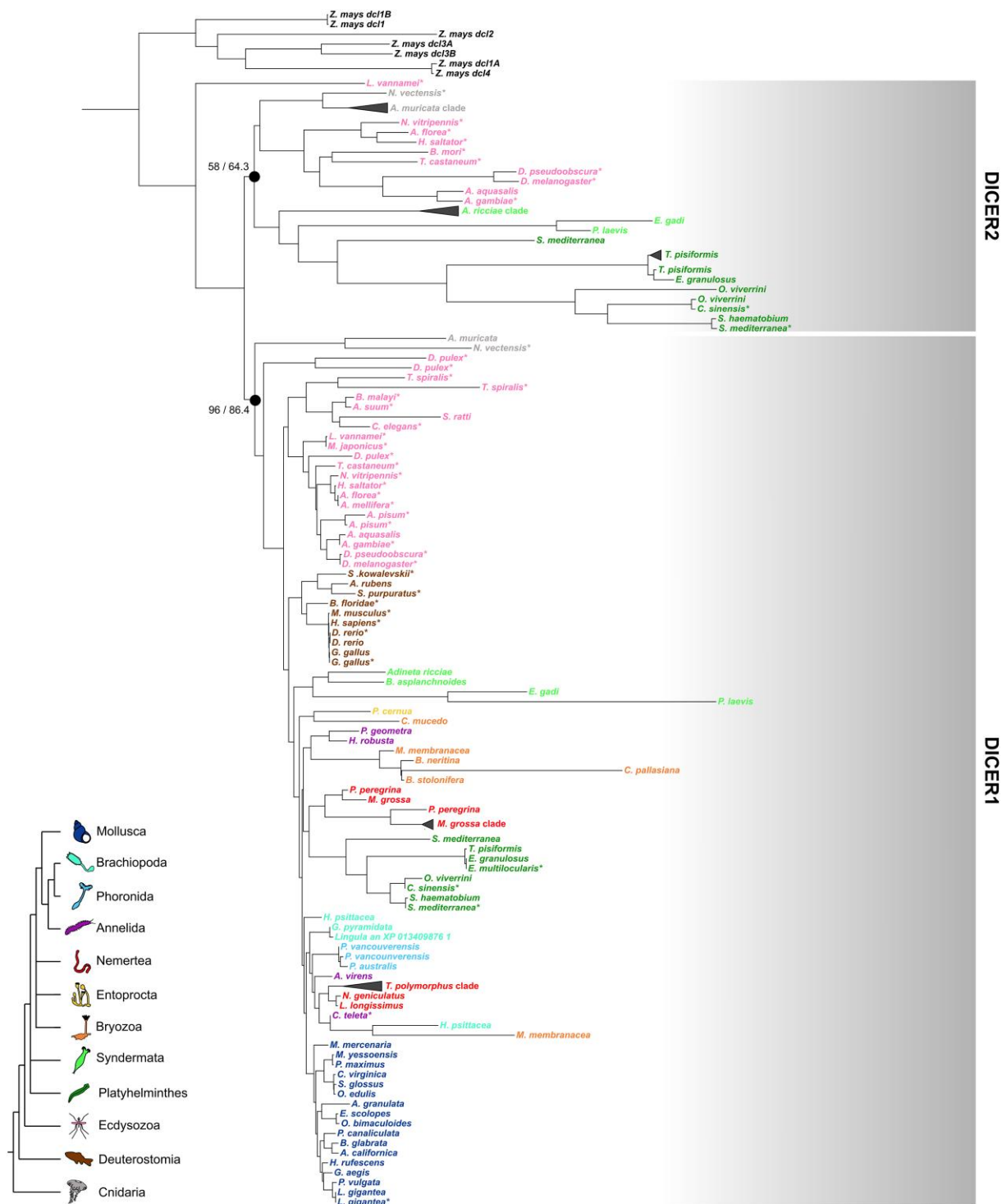


FIG. 2.—ML tree of the lophotrochozoan DICER proteins. For two marked nodes, it is reported respectively the UFBoot and the SH-*alt* value. Support values of the remaining nodes are shown in the [supplementary fig. S2, Supplementary Material](#) online. Clades formed by paralogs of the same species were collapsed and represented with a triangle. Reference 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 latest Lophotrochozoa phylogenetic analyses (Kocot et al. 2017; Bleidorn 2019; Marlétaz et al. 2019).

Downloaded from <https://academic.oup.com/gbe/article/16/5/evae098/7666253> by guest on 26 July 2024

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 (supplementary fig. S3, Supplementary Material online). We also estimated the ratio of nonsynonymous to synonymous substitution rates (ω) along the AGO family tree and the DICER family tree: we confirmed that, during the evolution of the DICER and AGO families, purifying selection on DICER1 (LRT = 34.34, P -value = 4.62×10^{-9} ; supplementary fig. S4B, Supplementary Material online) and miAGO (LRT = 204.64, P -value = 0; supplementary fig. S4A, Supplementary Material online) has intensified compared with 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 (Table 1). 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 (Table 1). 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 underrepresented phyla (i.e. Nemertea, Bryozoa, Syndermata). The annotation of genome assemblies with BRAKER has heavily relied on protein databases (see Materials and Methods); thus, it may be possible that this method produced better annotations for clades like Mollusca, Annelida, or Platyhelminthes.

Domain Characterization in Argonaute and DICER Proteins

The annotation of Argonaute and DICER proteins mostly relied on the annotation of peculiar domains (see Materials and Methods). However, other domains characterize the two protein families. Thus, we built domain profiles of Argonaute and DICER domains from multiple sequences alignments. Profiles were aligned against the annotated proteins to evaluate the domain composition of lophotrochozoan Argonaute and DICER proteins.

Besides the PIWI and PAZ domains, Argonaute proteins are also characterized by N-terminal (N) and the MID domain. We failed to annotate the two domains in most of PIWI proteins; in particular, the MID domain was not annotated in any Syndermata and Bryozoa PIWI1 protein, but also in each lophotrochozoan phylum, we reported at least one species without the domain (supplementary table S2,

Supplementary Material online). Additionally, each Syndermata PIWI2 protein lacked not only the MID domain but also the N domain. Similarly, Bryozoa PIWI2 proteins did not contain the MID domain either (supplementary tables S2 and S3, Supplementary Material online). On the other hand, in almost all miAGO proteins both domains were annotated (*Hemithiris psittacea* being the only exception), while in siAGO proteins, almost all Platyhelminthes lacked the MID domain. After localizing the domain position in the Argonaute alignment, we examined whether certain species lacked the domains entirely. This was determined by assessing whether their sequence in that portion of the alignment was either entirely absent (i.e. with most sites being gaps) or significantly degenerated (with most sites containing amino acids, but the sequence being too degenerated for accurate domain annotation). In most cases, the sequences were found to be degenerated (supplementary table S4, Supplementary Material online).

Regarding DICER domain composition, we assessed the presence of the Helicase (Hel) and PAZ domain. The Hel domain was absent in most of Lophotrochozoa DICER2 proteins, with the only exception of two Syndermata paralogs (supplementary table S5, Supplementary Material online). The PAZ domain also resulted absent from most of Lophotrochozoa DICER2 proteins, exception made for one *S. mediterranea* DICER2 paralog (supplementary table S6, Supplementary Material online). Among DICER1 proteins, Hel and PAZ domains were annotated in all Mollusca species, while in other lophotrochozoan phyla, the annotation of both domains was restricted to some species (supplementary tables S5 and S6, Supplementary Material online). Notably, the two domains resulted completely absent in Bryozoa. Contrastingly with the Argonaute analysis, DICER domains, when not recovered, were completely missing. According to the structure of DICER, the PAZ and Hel domains are toward the N-terminal end, with the Hel being the first domain of the protein (Mukherjee et al. 2013). Accordingly, all the lophotrochozoan DICER proteins lacking either the two domains or solely the Hel domain were found to be truncated at the N-terminus. The only exceptions were some DICER2 paralogs of the syndermatan *Adineta ricciae*: while five paralogs resulted truncated, three of them displayed degeneration only (supplementary table S4, Supplementary Material online).

We also evaluated the conservation of the DECH box within the Hel domain. The amino acid composition of the DECH box (i.e. aspartic acid, glutamic acid, cysteine, and histidine) resulted conserved in Syndermata, Nematoda, Phoronida, and Brachiopoda. Most Annelida and Gastropoda showed an aspartic acid instead of a glutamic acid on the second position (resulting in DDCH); the DECH box resulted further mutated in Bivalvia (i.e. ENCH in Ostreida, DHCQ in Pectinida, DDCH in *Mercenaria mercenaria*), in *Biomphalaria glabrata*

(DNCH), and in Cephalopoda (ECSN). Platyhelminthes also reported a highly diverged DECH box ([supplementary fig. S5, Supplementary Material online](#)).

Looking for the Endo-siRNA Signature in Small RNA Libraries

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 overlap by 19 bases. Similarly, piRNAs produced by the ping-pong cycle go in pairs that overlap by 10 bases (Antoniewski 2014; Khanal et al. 2022). Thus, siRNAs and piRNAs have a unique signature that can be identified in sncRNA libraries by looking for overlapping pairs of reads. We retrieved eight sncRNA libraries from lophotrochozoan and nonlophotrochozoan species (namely *D. rerio*, *Apostichopus japonicus*, *A. muricata*, *A. gambiae*, *D. melanogaster*, *S. mediterranea*, *Schistosoma japonicum*, *Crassostrea gigas*), and we sequenced the sncRNA pool of *Notospermus geniculatus* (Nemertea) to include a species with an incomplete endo-siRNA pathway (i.e. presence of siAGOs, but absence of DICER2) in our analysis. 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 read 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), species equipped with DICER2 (i.e. *D. melanogaster*, *A. gambiae*, *A. muricata*, *S. mediterranea*, and *S. japonicum*) reported a Z-score higher in the 19-overlap group than species without DICER2 (*D. rerio*, *A. japonicus*, *C. gigas*, *N. geniculatus*; Fig. 3). Some species also reported a sharp increase in the Z-score for the 10-overlap group for 21 bp long reads only, namely *N. geniculatus*, *A. japonicus*, *C. gigas*, and *S. mediterranea*. A 10-base overlap would correspond to the piRNA signature, although the length of piRNAs in Arthropoda ranges from 25 to 30 bp. Overall, considering the phylogenomic analysis, all the species that exhibited a complete siRNA pathway (using DICER2 and siAGO as a proxy) showed a high Z-score in the 19-overlap group (Fig. 3). Finally, we investigated the range of action of the three sncRNA types in three metazoan species (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 ([supplementary fig. S6A, Supplementary Material online](#)). In terms of reads per million (RPM), in all three species, miRNAs are the most expressed and their length ranges from 20 to 25 nc. In contrast,

piRNAs were annotated in the in the 24 to 30 nc length range. In most cases, we were not able to discern the endo-siRNA signal from noise, but *A. gambiae* was the species that showed the highest expression levels ([supplementary fig. S6A, Supplementary Material online](#)). We also compared ovary and somatic tissue sncRNA libraries. As expected, in all three species, piRNAs is the class more expressed in the ovaries ([supplementary fig. S6B, Supplementary Material online](#)). Overall, in the three species, we did not report notable differences in terms of sncRNA length or differential expression in somatic/ovarian tissue among piRNA and miRNA types.

The annotation of the sRNA types was performed even for the newly sequenced small RNA libraries of *N. geniculatus*. Following the guidelines of Fromm et al. (2015), we identified 154 bona fide miRNA genes. Each miRNA gene has been annotated by blasting its preliminary structure against the MirGeneDB pre-miRNA database (Fromm et al. 2022) and checking the conservation of the seed region. According to the annotation results, 63 *N. geniculatus* miRNA genes were included in 29 miRNA families already described in other metazoan species ([supplementary table S7, Supplementary Material online](#)). The remaining 91 miRNA genes did not exhibit significant similarities with the pre-miRNAs in the database ([supplementary materials, Supplementary Material online](#)). To assess the conservation of the novel miRNAs and increase the reliability of our predictions, we aligned the pre-miRNAs against the *Lineus longissimus* genome (i.e. the most closely related species to *N. geniculatus* with an assembled genome). We identified at least 28 novel miRNA genes that are shared between *N. geniculatus* and *L. longissimus*. Based on their preliminary sequences and seed regions, we clustered these 28 novel miRNA genes into 17 novel families ([supplementary table S8, Supplementary Material online](#)). Comparing the expression levels of the three different types of small RNAs, considering small RNAs with a length ranging between 20 and 24 nucleotides, miRNAs resulted the most expressed even in *N. geniculatus*. Conversely, piRNAs resulted expressed mostly in the length range 25 to 29 nc ([supplementary fig. S7, Supplementary Material online](#)). As for *C. gigas* ([supplementary fig. S6B, Supplementary Material online](#)), we were not able to discern the endo-siRNA signal from noise.

Discussion

Most Lophotrochozoans Have Conventional miRNA and piRNA Pathways

RNAi pathways play a central role in many molecular aspects, from mRNA regulation to defense mechanisms, and Argonaute proteins are the key RNAi players in all eukaryotes. All the phylogenetic analyses agree to divide the

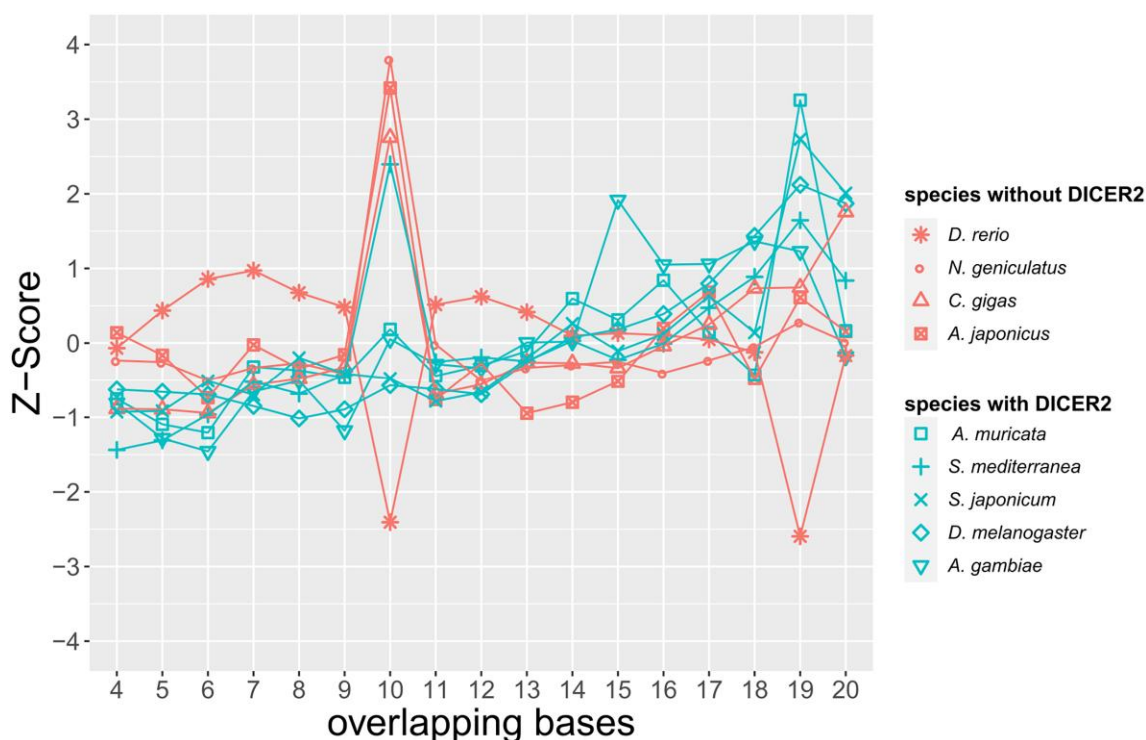


Fig. 3.—Evaluating the siRNA signature in sncRNA libraries. The plot reports the Z-score between the number of pairs of all possible overlaps. A Z-score > 1 means that pairs overlapping of that length are at least a standard deviation more numerous than the mean of all the overlaps. The species are sorted by the presence/absence of DICER2.

eukaryotic Argonaute superfamily into 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 animals; moreover, the PIWI and the WAGO families are restricted to animals or even nematodes, respectively (Swarts et al. 2014). It is still uncertain how the four families emerged during eukaryote evolution. For instance, several eukaryotic clades have a miRNA-like 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 identified 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 (Table 1). The same pattern has been observed in the DICER phylogeny, with DICER2 proteins being paraphyletic with respect to DICER1 proteins, which clustered in a well-supported monophyletic clade (Fig. 2) (Mukherjee et al. 2013). Concordantly, DICER2 and DICER1 are related to siAGO and miAGO proteins, respectively. 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 lophotrochozoans showed two distant related PIWI proteins (i.e. one AUB-like and one AGO3-like; Fig. 1; Table 1), 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 already been 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 three clades.

The miRNA pathway is the most ubiquitous RNAi pathway among Metazoa, and almost all species reported a miAGO and a DICER1 protein. For these proteins, we

even detected lower root-to-tip distances and a decrease in 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 to 25 nt long, they are not more expressed in the ovaries than in somatic tissue, and overall, they are by far the most expressed sncRNA class.

The conservation is also reflected in the domain composition; most of miAGO and DICER1 proteins included all the domains. However, a novel pattern has been observed in the DECH box. The DECH box is a motif present in many helicase domains, and it coordinates ATP hydrolysis (Yerukhimovich et al. 2018). Although it is conserved between distant related DICER proteins (f.i., *H. sapiens* DICER1, *D. melanogaster* DICER2, *C. elegans* DICER; [supplementary fig. S5, Supplementary Material](#) online), its conservation does not imply that the Hel domain is active, since the *H. sapiens* DICER has been proved to work in an ATP-independent manner (Liu et al. 2018). The Hel domain is likely inactive also in Mollusca and Annelida, where the DECH box diverges of one or two amino acids, but it remains conserved in Phoronida, Brachiopoda, and Nematoda, where it may be still active ([supplementary fig. S5, Supplementary Material](#) online).

The Evolution of the Endo-siRNA Pathway in Lophotrochozoa

Pathways maturing endo-siRNAs are deeply diverse between metazoan clades. In the fruit fly, the miRNA and the endo-siRNA pathways are separated, having a specific Argonaute and DICER protein for each pathway. In *C. elegans*, a single DICER protein is responsible for the maturation of miRNAs and endo-siRNAs: endo-siRNAs are then loaded by ERGO-1 (but also other Argonaute proteins; Han et al. 2009). RNA-dependent RNA polymerases amplify the mechanism through the production of secondary siRNAs, loaded by WAGO proteins (Billi et al. 2014). In mammals, even if they lack siAGO proteins, endo-siRNAs are loaded on AGO2 (Watanabe et al. 2008), but their maturation has not been elucidated yet. Endo-siRNAs have been also reported in early diverging animals, where they are loaded by a specific siAGO (Fridrich et al. 2020).

Overall, small RNAs produced by long endogenous dsRNAs have been described in all Metazoa, but the maturation pathway evolved differently in different clades. This pathway has been certainly overlooked in Lophotrochozoa. An endo-siRNA pathway is likely to exist in most of early diverging Lophotrochozoa, since most of Platyhelminthes and Syndermata included a siAGO and DICER2 protein (Figs. 1

and 2). This endo-siRNA pathway looks like the one of insects and cnidarians, where two distinct DICER and AGOs interact with two distinct small RNA types (at least for most small RNAs, see Fridrich et al. 2020). However, lophotrochozoan DICER2 proteins showed notable differences in the domain composition compared with cnidarian or insect ones. Most of them lack the Hel and PAZ domain, which are pivotal for DICER2. The PAZ domain recognizes target dsRNAs and binds their 3' end, while the Hel domain, which seems to be inactive in all metazoan DICER1 (Aderounmu et al. 2023), allows the translocation of DICER2 along the target dsRNA, producing siRNAs processively (Kandasamy and Fukunaga 2016).

Nevertheless, we detected small RNAs with the peculiar endo-siRNA signature (i.e. 21 bp small RNA pairs with an overlap of 19 bases; Fig. 3) in the small RNA transcriptomes of *S. mediterranea* and *S. japonicum* (Platyhelminthes). Thus, it is possible that the DICER2 of early diverging Lophotrochozoa can still mature dsRNAs without the PAZ and the Hel domain. When the Hel domain is experimentally inactivated in DICER2 of *C. elegans* or in *D. melanogaster*, the protein loses the ability of translocase along dsRNAs, but it is still able to target dsRNAs and produce endo-siRNAs (Welker et al. 2011; Sinha et al. 2018). At the same time, the inactivation of the PAZ domain leads to the production of siRNAs of altered length (Kandasamy and Fukunaga 2016). Overall, Platyhelminthes and Syndermata DICER2 might still work, and our results show that (Fig. 3), but the lack of the two domains might affect them in fidelity and efficiency.

Within Trochozoa, we did not detect DICER2, but some species belonging to Nemertea, Bryozoa, and Entoprocta possess siAGO proteins. These three phyla have occasionally been placed as sister group of all the other Trochozoa (Kocot 2016; Kocot et al. 2017; Laumer et al. 2019; Khalturin et al. 2022). Thus, the endo-siRNA pathway would have been progressively lost during the evolution of Lophotrochozoa. The first step has been the loss of DICER2 in the ancestor of Trochozoa. Then, the loss of siAGO proteins in the ancestor of Mollusca, Brachiopoda, Phoronida, and Annelida followed (Fig. 4). On the other hand, many other analyses do not place Entoprocta, Ectoprocta, and Nemertea at the base of Trochozoa (Laumer et al. 2015; Marlétaz et al. 2019); in that scenario, siAGO proteins would have been lost multiple times, depending on the phylogenetic relationships between phyla. Finally, the absence of a complete endo-siRNA pathway in Trochozoa is confirmed by the analysis of small RNA transcriptomes: we did not detect small RNAs with the endo-siRNA signature in species with an uncomplete endo-siRNA pathway (Fig. 3), namely *C. gigas* (Mollusca) and *N. geniculatus* (Nemertea), but also *D. rerio* and *A. japonicus* (Deuterostomia).

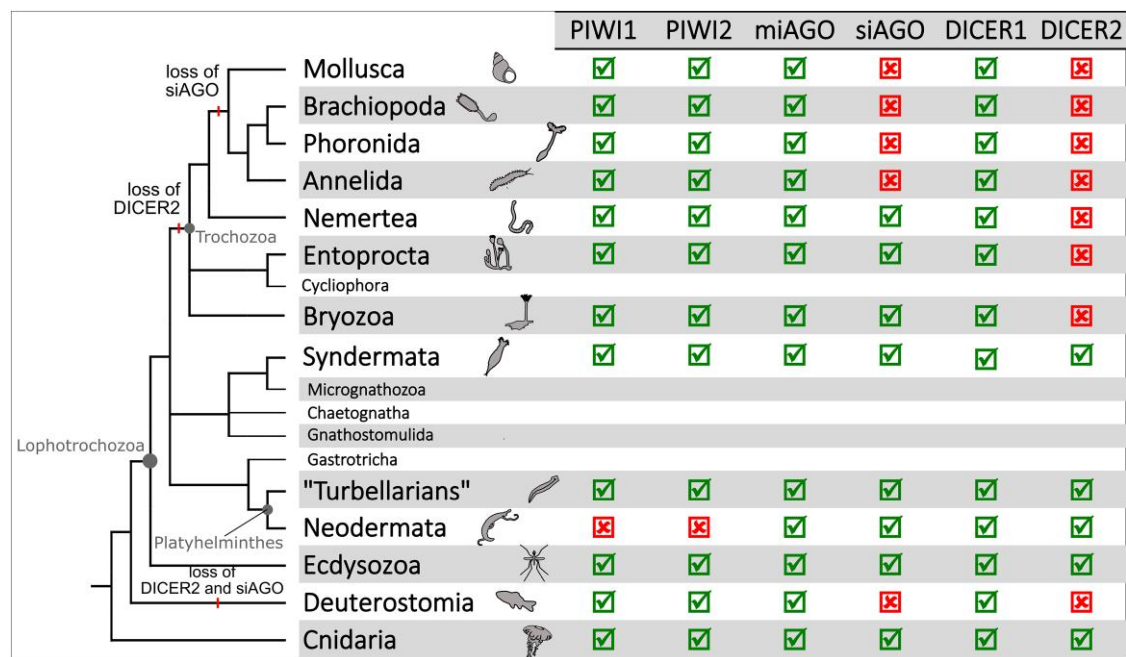


Fig. 4.—The loss of DICER2 and siAGO along the Metazoa evolution. The lophotrochozoan phylogenetic tree is reconstructed according to the latest Lophotrochozoa phylogenetic analyses (Kocot et al. 2017; Bleidorn 2019; Marlétaz et al. 2019). For each protein, it is reported its presence, with a check, or the absence, with a cross, in each metazoan clade.

Unraveling the Evolution of RNAi Pathways in Lophotrochozoa

In this study, we highlighted notable differences between Lophotrochozoa and other Metazoa RNAi pathways. Platyhelminthes and Syndermata have maintained an endo-siRNA machinery, but DICER2 diverges considerably from the DICER2 protein of other metazoan clades. Nevertheless, Platyhelminthes are able to produce endo-siRNAs (Fig. 3). Since the endo-siRNA pathway is highly diverse among different Metazoa clades, it is likely that also in early diverging Lophotrochozoa, the endo-siRNA pathway has evolved in a unique way, not completely comparable with other metazoan pathways.

An even more unique condition was described in Nemertea, Entoprocta, and Ectoprocta. These clades show an intermediate state during the loss of the endo-siRNA pathway in Lophotrochozoa. The absence of DICER2 proteins may preclude a functional endo-siRNA pathway. However, this pathway proved to be very flexible. It is possible that the siAGO protein of these clades has evolved to load small RNAs matured from other pathways. Like in *C. elegans*, DICER1 may mature miRNA as well as endo-siRNA. Having said that, there are multiple scenarios where siAGO proteins may be involved, also considering the number of unconventional RNAi pathways that have been described so far (Yang and Lai 2011).

Finally, the loss of siAGO proteins and, more generally, of the endo-siRNA pathway, possibly in all Trochozoa, has

been strongly supported by the joint phylogenetic analysis and analysis of sncRNA libraries. The pattern is comparable with that obtained from Deuterostomia, where the absence of the canonical (i.e. involving both DICER and siAGOs) endo-siRNA pathway has already been reported using the lack of annotated siAGO proteins as a proxy (Wynant et al. 2017). Our analysis supports this hypothesis, since neither siAGO nor DICER2 homologs were retrieved in deuterostomes (Figs. 1 and 2). Nevertheless, in mammals, endo-siRNAs are processed by the same DICER and AGO proteins that process miRNAs (Watanabe et al. 2008; Svobodova et al. 2016). Similarly, in Trochozoa, endo-siRNAs may be matured by the miRNA pathway or other proteins related to the RNAi mechanism. As for mammals, immunoprecipitation or knockout experiments might elucidate whether Argonaute proteins, as well as other protein families, can interact with other sncRNAs types in addition to piRNAs and miRNAs. All these findings are also deeply linked to the characterization of the innate immune system. In mammals, the interferon pathway has replaced RNAi mechanisms in the role of viral defense (Isaacs et al. 1963; Loo and Gale 2011; Schuster et al. 2019). An interferon defense mechanism has been described also in Mollusca (Huang et al. 2017; Qiao et al. 2021). Thus, in Lophotrochozoa, an interferon system might have evolved coincidentally with the loss of the endo-siRNA pathway. Further comparative analyses might characterize the evolution of these

pathways and elucidate the mechanisms that control the innate immune system in Lophotrochozoa.

Materials and Methods

Annotation of Argonaute and DICER Proteins

Argonaute and DICER proteins were annotated for a wide range of omics data. Initially, we analyzed all lophotrochozoan proteomes annotated through the National Center for Biotechnology Information (NCBI) Eukaryotic Genome Annotation Pipeline (Thibaud-Nissen et al. 2016). To increase the sampling in underrepresented clades, we selected 17 assemblies (supplementary table S9, Supplementary Material online) and predicted gene models using the BRAKER2 automated pipeline (Stanke et al. 2008; Hoff et al. 2019; Brůna et al. 2021). To enhance the DICER and Argonaute model predictions, we enriched the Metazoa OrthoDB database provided by BRAKER2 with Argonaute and DICER sequences annotated from the lophotrochozoan NCBI proteomes. We collapsed all isoforms, retaining the longest ones, using the Perl script `agat_sp_keep_longest_isoform.pl` (Dainat et al. 2019). When few genome assemblies were available for a given phylum, we searched for Argonaute and DICER proteins in transcriptomes (Table 1). We trimmed the reads with Trimmomatic-0.39 (Bolger et al. 2014) using 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 by locally aligning the transcripts against the nonredundant protein database (Sayers et al. 2022) with DIAMOND blastp (Camacho et al. 2009; Buchfink et al. 2015) and discarding all transcripts with a nonmetazoan 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 6 other metazoan species. Coding sequences were translated into amino acid sequences, and then, Argonaute and DICER proteins were annotated as follows: 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 searched the profiles against each ($-e$ -value $10e^{-6}$). Only proteins with both domains annotated were considered Argonaute proteins and retained for phylogenetic analysis.

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

round of filtering, we retrieved the ribonuclease 3 domain alignment from Pfam (accession: pf14622.9), the only domain shared between all metazoan DICERs (Mukherjee et al. 2013); Scanning the sequences with HMMER ($-$ -value $10e^{-6}$), we selected only those containing the ribonucleases 3 domain. However, orthologs of the endoribonuclease DROSHA were possibly included among the annotated DICERs at this stage. Therefore, we downloaded the DROSHA orthologs from OrthoDB (reference: 9211at3208) (Zdobnov et al. 2021), and we built a custom data set with both DICER from Mukherjee and colleagues (2013) and DROSHA sequences. We locally aligned the set of putatively annotated DICER proteins against this data set. We retained proteins whose best five hits were all with DICER orthologs and 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 data sets comprising all annotated sequences from proteomes, genomes, and transcriptomes, with the addition of reference sequences chosen from SwissProt (supplementary table S1, Supplementary Material online) or the bilaterian annotated set for the DICER data set. Data sets were aligned with MAFFT v7.508 (Katoh and Standley 2013), using the options `--maxiterate 1,000 --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`. Additionally, another masking tool, ClipKIT (Steenwyk et al. 2020), was used to assess the impact of the masking step on the phylogenetic analyses. Gblocks resulted in being more conservative than ClipKIT, masking most of the alignment columns (supplementary table S10, Supplementary Material online). However, the ML trees inferred from the two alignments showed no difference between each other regarding the presence of each Argonaute (i.e. miAGO, siAGO, PIWI1,2) or DICER protein in each lophotrochozoan species (supplementary table S10, Supplementary Material online). Therefore, only the alignment obtained with Gblocks was used for downstream analyses.

The ML trees were inferred with IQ-TREE (Nguyen et al. 2015) using the predefined protein mixture model LG + C20 + R4. To assess the robustness of the clades, we calculated the UFBoot with 1,000 bootstrap replicates (Hoang et al. 2018) and the SH-*alrt* 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 branches belonging to DICER1 and miAGO clades as foreground. All 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, *A. gambiae* DICER2, and *Tribolium castaneum* and *Brugia malayi* DICER1. Those proteins were removed during the selection analysis.

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

The annotation of the N (accession: pfam16486) and MID (accession: pfam16487) domains for Argonaute proteins and the Hel and PAZ domains for DICER proteins was made using HMMER, setting the e-value cutoff as the lowest e-value among outgroup sequences. The Hel and PAZ profiles were built from sequences downloaded from UniProt (Bateman et al. 2021): *D. melanogaster* DICER1 (accession: Q9VCU9), *D. melanogaster* DICER2 (accession: A1ZAW0), *N. vectensis* DICER1 (accession: U3MHS9), *Mytilus gallus* DICER (accession: A0A140H129), *C. elegans* DICER (accession: P34529), and *H. sapiens* DICER (accession: Q9UPY3).

N. geniculatus sncRNA Library Sequencing and Analysis of sncRNA Libraries

Six specimens (three males and three females) of the nemertean *N. 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 s; gonads were then dissected in 7% MgCl₂·H₂O on ice and stored in RNAlater (Thermo Fisher Scientific Inc., Waltham, USA), following the manufacturer's instructions. Total RNA was extracted using a standard chloroform: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 small RNA libraries that were sequenced on an Illumina HiSeq 2500 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 alongside publicly available sncRNA libraries from five other species. The libraries were selected and downloaded from the Sequence Read Archive (SRA) provided by NCBI (supplementary table S11, Supplementary Material online). Where multiple samples from the same project were available, libraries were pooled together, to obtain a single fastq file for each species. Adapters and low-quality bases were removed from reads using Cutadapt v3.9.7 (Martin 2011), with the options -e 0.2 -O 5 --quality-cutoff 6 --discard-untrimmed. Trimmed reads were mapped on the reference genome using Bowtie (Langmead et al. 2009), allowing up to 100 multiple alignments (-m 100). The distribution of overlaps between reads was estimated using the python script *overlapping_reads.py* ([https://](https://github.com/ARTbio/tools-artbio/blob/master/tools/small_rna_signatures/overlapping_reads.py)

github.com/ARTbio/tools-artbio/blob/master/tools/small_rna_signatures/overlapping_reads.py; Antoniewski 2014).

For *C. gigas*, *D. rerio*, *A. gambiae*, and *N. geniculatus*, 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 et al. 2019), along with the annotated miRNAs of up to five closely related species. The novel predicted miRNAs were evaluated following the criteria established by Fromm et al. (2015) and discarding novel miRNAs with a STAR sequence coverage lower than five reads. Putative siRNA and piRNA pairs were predicted based on read overlaps (i.e. siRNA pairs overlap of the read length —2, piRNA pairs with an overlap of 10 nucleotides), calculated using *overlapping_reads.py*. Pairs of siRNAs and piRNAs were discarded when one of the paired small RNAs had a coverage lower than 5, the logarithmic ratio of the pair exceeded 1.5, or the pair mapped on a miRNA region. Differential expression of sncRNAs between somatic tissues and ovaries was tested with edgeR (Robinson et al. 2010) with a generalized linear model and a quasi-likelihood *F*-test (Lund et al. 2012). We chose a stringent *P*-value threshold of 0.001 to consider a small RNA as significantly differentially expressed. Novel miRNA genes in *N. geniculatus* were assigned to known or novel miRNA families. We locally aligned the preliminary structure of miRNA genes against the full set of MiRGeneDB pre-miRNA using blastn. Novel genes were assigned to the best hit miRNA family only if they shared the same seed (supplementary table S7, Supplementary Material online). The remaining miRNA genes were clustered into novel miRNA families by blasting the pre-miRNAs against each other and comparing the seed sequence: miRNAs were clustered in the same family if they blasted against each other and the seed sequence differed by up to one nucleotide (supplementary table S8, Supplementary Material online).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We would like to thank the anonymous reviewers for their thoughtful comments and constructive feedback, which have significantly improved the quality of this manuscript.

Author Contributions

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. 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.

Funding

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

Data Availability

The data underlying this article are available in the article and in its online [supplementary material](#). Newly sequenced raw reads were uploaded to the SRA under the BioProject PRJNA942081. Python, bash, and R scripts written for the differential expression of miRNAs, piRNAs, and siRNAs have been uploaded on a GitHub repository (https://github.com/AlessandroFormaggioni/argo_Metazoa/tree/main/DE_pipeline). If further data supporting this article are required, they will be made available upon reasonable request to the corresponding author.

Literature Cited

- Aderounmu AM, Aruscavage PJ, Kolaczowski B, Bass BL. Ancestral protein reconstruction reveals evolutionary events governing variation in Dicer helicase function. *eLife* 2023;12:e85120. <https://doi.org/10.7554/eLife.85120>.
- Antoniewski C. Computing siRNA and piRNA overlap signatures. *Animal endo-SiRNAs*. 1173. New York, NY: Humana Press; 2014. p. 135–146.
- Bartel DP. Metazoan MicroRNAs. *Cell* 2018;173(1):20–51. <https://doi.org/10.1016/j.cell.2018.03.006>.
- Bateman A, Martin M-J, Orchard S, Magrane M, Agivetova R, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bursteinas B, et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49(D1):D480–D489. <https://doi.org/10.1093/nar/gkaa1100>.
- Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 2001;409(6818):363–366. <https://doi.org/10.1038/35053110>.
- Billi AC, Fischer SEJ, Kim JK. Endogenous RNAi pathways in *C. elegans*. In: The *C. elegans* Research Community, ed. *WormBook*. 2014. doi:10.1895/wormbook.1.170.1. <http://www.wormbook.org>.
- Bleidorn C. Recent progress in reconstructing lophotrochozoan (spiralian) phylogeny. *Org Divers Evol*. 2019;19(4):557–566. <https://doi.org/10.1007/s13127-019-00412-4>.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 2007;128(6):1089–1103. <https://doi.org/10.1016/j.cell.2007.01.043>.
- Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform*. 2021;3(1):lqaa108. <https://doi.org/10.1093/nargab/lqaa108>.
- Brusca R, Moore W, Shuster S. *Invertebrates*. Third Edition. Sunderland: Sinauer Associates; 2016.
- Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59–60. <https://doi.org/10.1038/nmeth.3176>.
- Calcino AD, Fernandez-Valverde SL, Taft RJ, Degnan BM. Diverse RNA interference strategies in early-branching metazoans. *BMC Evol Biol*. 2018;18(1):160. <https://doi.org/10.1186/s12862-018-1274-2>.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10(1):421. <https://doi.org/10.1186/1471-2105-10-421>.
- Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 2000;17(4):540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>.
- Conway Morris S, Smith RDA, Cuthill JFH, Bonino E, Lerosey-Aubril R. A possible Cambrian stem-group gnathiferan-chaetognath from the Weeks Formation (Miaolingian) of Utah. *J Paleontol*. 2020;94(4):624–636. <https://doi.org/10.1017/jpa.2020.4>.
- Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, et al. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* 2008;453(7196):798–802. <https://doi.org/10.1038/nature07007>.
- Dainat J, Hereñú D, Pucholt P. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format. 2019. <https://github.com/NBISweden/AGAT>.
- Fontenla S, Rinaldi G, Smircich P, Tort JF. Conservation and diversification of small RNA pathways within flatworms. *BMC Evol Biol*. 2017;17(1):215. <https://doi.org/10.1186/s12862-017-1061-5>.
- Fontenla S, Rinaldi G, Tort JF. Lost and found: piwi and Argonaute pathways in flatworms. *Front Cell Infect Microbiol*. 2021;11:653695. <https://doi.org/10.3389/fcimb.2021.653695>.
- Fridrich A, Modepalli V, Lewandowska M, Aharoni R, Moran Y. Unravelling the developmental and functional significance of an ancient Argonaute duplication. *Nat Commun*. 2020;11(1):6187. <https://doi.org/10.1038/s41467-020-20003-8>.
- Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res*. 2012;40(1):37–52. <https://doi.org/10.1093/nar/gkr688>.
- Fromm B, Billipp T, Peck LE, Johansen M, Tarver JE, King BL, Newcomb JM, Sempere LF, Flatmark K, Hovig E, et al. A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. *Annu Rev Genet*. 2015;49(1):213–242. <https://doi.org/10.1146/annurev-genet-120213-092023>.
- Fromm B, Høy E, Domanska D, Zhong X, Aparicio-Puerta E, Ovchinnikov V, Umu S, Chabot P, Kang W, Aslanzadeh M, et al. MirGeneDB 2.1: toward a complete sampling of all major animal phyla. *Nucleic Acids Res*. 2022;50(D1):D204–D210. <https://doi.org/10.1093/nar/gkab1101>.
- Garcia Silva MR, Tosar JP, Frugier M, Pantano S, Bonilla B, Esteban L, Serra E, Rovira C, Robello C, Cayota A, et al. Cloning, characterization and subcellular localization of a *Trypanosoma cruzi* Argonaute protein defining a new subfamily distinctive of trypanosomatids. *Gene* 2010;466(1-2):26–35. <https://doi.org/10.1016/j.gene.2010.06.012>.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644–652. <https://doi.org/10.1038/nbt.1883>.
- Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP. Early origins and evolution of

- microRNAs and piwi-interacting RNAs in animals. *Nature* 2008;455(7217):1193–1197. <https://doi.org/10.1038/nature07415>.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59(3):307–321. <https://doi.org/10.1093/sysbio/syq010>.
- Han T, Manoharan AP, Harkins TT, Bouffard P, Fitzpatrick C, Chu DS, Thierry-Mieg D, Thierry-Mieg J, Kim JK. 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in *Caenorhabditis elegans*. *Proc Natl Acad Sci USA*. 2009;106(44):18674–18679. <https://doi.org/10.1073/pnas.0906378106>.
- Hirakata S, Siomi MC. piRNA biogenesis in the germline: from transcription of piRNA genomic sources to piRNA maturation. *Biochim Biophys Acta Gene Regul Mech*. 2016;1859(1):82–92. <https://doi.org/10.1016/j.bbagrm.2015.09.002>.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018;35(2):518–522. <https://doi.org/10.1093/molbev/msx281>.
- Höck J, Meister G. The Argonaute protein family. *Genome Biol*. 2008;9(2):210. <https://doi.org/10.1186/gb-2008-9-2-210>.
- Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. *Methods Mol Biol*. 2019;1962:65–95. https://doi.org/10.1007/978-1-4939-9173-0_5.
- Huang B, Zhang L, Du Y, Xu F, Li L, Zhang G. Characterization of the mollusc RIG-I/MAVS pathway reveals an archaic antiviral signalling framework in invertebrates. *Sci Rep*. 2017;7(1):8217. <https://doi.org/10.1038/s41598-017-08566-x>.
- Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet*. 2011;12(2):99–110. <https://doi.org/10.1038/nrg2936>.
- Isaacs A, Cox RA, Rotem Z. Foreign nucleic acids as the stimulus to make interferon. *Lancet*. 1963;282(7299):113–116. [https://doi.org/10.1016/S0140-6736\(63\)92585-6](https://doi.org/10.1016/S0140-6736(63)92585-6).
- Iwakawa H, Tomari Y. Life of RISC: formation, action, and degradation of RNA-induced silencing complex. *Mol Cell*. 2022;82(1):30–43. <https://doi.org/10.1016/j.molcel.2021.11.026>.
- Jehn J, Gebert D, Pipilescu F, Stern S, Kiefer JST, Hewel C, Rosenkranz D. PIWI genes and piRNAs are ubiquitously expressed in mollusks and show patterns of lineage-specific adaptation. *Commun Biol*. 2018;1(1):137. <https://doi.org/10.1038/s42003-018-0141-4>.
- Jiang F, Ye X, Liu X, Fincher L, McKearin D, Liu Q. Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev*. 2005;19(14):1674–1679. <https://doi.org/10.1101/gad.1334005>.
- Kandasamy SK, Fukunaga R. Phosphate-binding pocket in Dicer-2 PAZ domain for high-fidelity siRNA production. *Proc Natl Acad Sci USA*. 2016;113(49):14031–14036. <https://doi.org/10.1073/pnas.1612393113>.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–780. <https://doi.org/10.1093/molbev/mst010>.
- Khalturin K, Shunatova N, Shchenkov S, Sasakura Y, Kawamitsu M, Satoh N. Polyzoa is back: the effect of complete gene sets on the placement of Ectoprocta and Entoprocta. *Sci Adv*. 2022;8(26):eabo4400. <https://doi.org/10.1126/sciadv.abo4400>.
- Khanal S, Zancanela BS, Peter JO, Flynt AS. The small RNA universe of *Capitella teleta*. *Front Mol Biosci*. 2022;9:802814. <https://doi.org/10.3389/fmolb.2022.802814>.
- Kocot KM. On 20 years of Lophotrochozoa. *Org Divers Evol*. 2016;16(2):329–343. <https://doi.org/10.1007/s13127-015-0261-3>.
- Kocot KM, Struck TH, Merkel J, Waits DS, Todt C, Brannock PM, Weese DA, Cannon JT, Moroz LL, Lieb B, et al. Phylogenomics of Lophotrochozoa with consideration of systematic error. *Syst Biol*. 2017;66(2):256–282. <https://doi.org/10.1093/sysbio/syw079>.
- Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res*. 2019;47(D1):D155–D162. <https://doi.org/10.1093/nar/gky1141>.
- Kuhn C-D, Joshua-Tor L. Eukaryotic Argonautes come into focus. *Trends Biochem Sci*. 2013;38(5):263–271. <https://doi.org/10.1016/j.tibs.2013.02.008>.
- Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25. <https://doi.org/10.1186/gb-2009-10-3-r25>.
- Laumer C, Bekkouche N, Kerbl A, Goetz F, Neves R, Sørensen M, Kristensen R, Hejnol A, Dunn C, Giribet G, et al. Spiralian phylogeny informs the evolution of microscopic lineages. *Curr Biol*. 2015;25(15):2000–2006. <https://doi.org/10.1016/j.cub.2015.06.068>.
- Laumer CE, Fernández R, Lemer S, Combosch D, Kocot KM, Riesgo A, Andrade SCS, Sterrer W, Sørensen MV, Giribet G, et al. Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proc R Soc Biol Sci*. 2019;286(1906):20190831. <https://doi.org/10.1098/rspb.2019.0831>.
- Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Rådmark O, Kim S, et al. The nuclear RNase III Drosha initiates microRNA processing. *Nature* 2003;425(6956):415–419. <https://doi.org/10.1038/nature01957>.
- Lee Y, Kim M, Han J, Yeom K-H, Lee S, Baek SH, Kim VN. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*. 2004;23(20):4051–4060. <https://doi.org/10.1038/sj.emboj.7600385>.
- Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, Carthew RW. Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* 2004;117(1):69–81. [https://doi.org/10.1016/S0092-8674\(04\)00261-2](https://doi.org/10.1016/S0092-8674(04)00261-2).
- Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Tóth KF. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev*. 2013;27(4):390–399. <https://doi.org/10.1101/gad.209841.112>.
- Liu Z, Wang J, Cheng H, Ke X, Sun L, Zhang QC, Wang H-W. Cryo-EM structure of human Dicer and its complexes with a pre-miRNA substrate. *Cell* 2018;173(5):1191–1203.e12. <https://doi.org/10.1016/j.cell.2018.03.080>.
- Loo Y-M, Gale M. Immune signaling by RIG-I-like receptors. *Immunity* 2011;34(5):680–692. <https://doi.org/10.1016/j.immuni.2011.05.003>.
- Lund SP, Nettleton D, McCarthy DJ, Smyth GK. Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Stat Appl Genet Mol Biol*. 2012;11(5):fj/sagmb.2012.11.issue-5/1544-6115.1826/1544-6115.1826.xml. <https://doi.org/10.1515/1544-6115.1826>.
- Malone CD, Hannon GJ. Small RNAs as guardians of the genome. *Cell* 2009;136(4):656–668. <https://doi.org/10.1016/j.cell.2009.01.045>.
- Marlétaz F, Peijnenburg KTCA, Goto T, Satoh N, Rokhsar DS. A new spiralian phylogeny places the enigmatic arrow worms among gnathiferans. *Curr Biol*. 2019;29(2):312–318. <https://doi.org/10.1016/j.cub.2018.11.042>.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17(1):10. <https://doi.org/10.14806/ej.17.1.200>.
- Matranga C, Zamore PD. Small silencing RNAs. *Curr Biol*. 2007;17(18):R789–R793. <https://doi.org/10.1016/j.cub.2007.07.014>.

- Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar G, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. Pfam: the protein families database in 2021. *Nucleic Acids Res.* 2021;49(D1):D412–D419. <https://doi.org/10.1093/nar/gkaa913>.
- Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 2013;41(12):e121–e121. <https://doi.org/10.1093/nar/gkt263>.
- Mohn F, Handler D, Brennecke J. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science* 2015;348(6236):812–817. <https://doi.org/10.1126/science.aaa1039>.
- Moran Y, Agron M, Praher D, Technau U. The evolutionary origin of plant and animal microRNAs. *Nat Ecol Evol.* 2017;1(3):0027. <https://doi.org/10.1038/s41559-016-0027>.
- Moran Y, Praher D, Fredman D, Technau U. The evolution of microRNA pathway protein components in Cnidaria. *Mol Biol Evol.* 2013;30(12):2541–2552. <https://doi.org/10.1093/molbev/mst159>.
- Mukherjee K, Campos H, Kolaczowski B. Evolution of animal and plant dicers: early parallel duplications and recurrent adaptation of antiviral RNA binding in plants. *Mol Biol Evol.* 2013;30(3):627–641. <https://doi.org/10.1093/molbev/mss263>.
- Nesnidal MP, Helmkampf M, Meyer A, Witek A, Bruchhaus I, Ebersberger I, Hankeln T, Lieb B, Struck TH, Hausdorf B, et al. New phylogenomic data support the monophyly of Lophophorata and an Ectoproct-Phoronid clade and indicate that Polyzoa and Kryptozoa are caused by systematic bias. *BMC Evol Biol.* 2013;13(1):253. <https://doi.org/10.1186/1471-2148-13-253>.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–274. <https://doi.org/10.1093/molbev/msu300>.
- Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet.* 2019;20(2):89–108. <https://doi.org/10.1038/s41576-018-0073-3>.
- Praher D, Zimmermann B, Genikhovich G, Columbus-Shenkar Y, Modepalli V, Aharoni R, Moran Y, Technau U. Characterization of the piRNA pathway during development of the sea anemone *Nematostella vectensis*. *RNA Biol.* 2017;14(12):1727–1741. <https://doi.org/10.1080/15476286.2017.1349048>.
- Qiao X, Wang L, Song L. The primitive interferon-like system and its antiviral function in molluscs. *Dev Comp Immunol.* 2021;118:103997. <https://doi.org/10.1016/j.dci.2021.103997>.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139–140. <https://doi.org/10.1093/bioinformatics/btp616>.
- Rosani U, Pallavicini A, Venier P. The miRNA biogenesis in marine bivalves. *PeerJ.* 2016;4:e1763. <https://doi.org/10.7717/peerj.1763>.
- Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau D, Connor R, Funk K, Kelly C, Kim S, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2022;50(D1):D20–D26. <https://doi.org/10.1093/nar/gkab1112>.
- Schuster S, Miesen P, van Rij RP. Antiviral RNAi in insects and mammals: parallels and differences. *Viruses* 2019;11(5):448. <https://doi.org/10.3390/v11050448>.
- Shabalina S, Koonin E. Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol.* 2008;23(10):578–587. <https://doi.org/10.1016/j.tree.2008.06.005>.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Sinha NK, Iwasa J, Shen PS, Bass BL. Dicer uses distinct modules for recognizing dsRNA termini. *Science* 2018;359(6373):329–334. <https://doi.org/10.1126/science.aag0921>.
- Song J-J, Smith SK, Hannon GJ, Joshua-Tor L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 2004;305(5689):1434–1437. <https://doi.org/10.1126/science.1102514>.
- Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntentically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics.* 2008;24(5):637–644. <https://doi.org/10.1093/bioinformatics/btn013>.
- Steenwyk JL, Buida TJ, Li Y, Shen X-X, Rokas A. ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol.* 2020;18(12):e3001007. <https://doi.org/10.1371/journal.pbio.3001007>.
- Strasser JFH, Irisarri I, Williams TA, Burki F. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nat Commun.* 2021;12(1):1879. <https://doi.org/10.1038/s41467-021-22044-z>.
- Struck TH, Wey-Fabrizius AR, Golombek A, Hering L, Weigert A, Bleidorn C, Klebow S, Iakovenko N, Hausdorf B, Petersen M, et al. Platyzoan paraphyly based on phylogenomic data supports a noncoelomate ancestry of spiralia. *Mol Biol Evol.* 2014;31(7):1833–1849. <https://doi.org/10.1093/molbev/msu143>.
- Svoboda E, Kubikova J, Svoboda P. Production of small RNAs by mammalian dicer. *Pflugers Arch.* 2016;468(6):1089–1102. <https://doi.org/10.1007/s00424-016-1817-6>.
- Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin EV, Patel DJ, van der Oost J. The evolutionary journey of Argonaute proteins. *Nat Struct Mol Biol.* 2014;21(9):743–753. <https://doi.org/10.1038/nsmb.2879>.
- Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, et al. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 2008;453(7194):534–538. <https://doi.org/10.1038/nature06904>.
- Tarver JE, Sperling EA, Nailor A, Heimberg AM, Robinson JM, King BL, Pisani D, Donoghue PCJ, Peterson KJ. miRNAs: small genes with big potential in metazoan phylogenetics. *Mol Biol Evol.* 2013;30(11):2369–2382. <https://doi.org/10.1093/molbev/mst133>.
- Thibaud-Nissen F, DiCuccio M, Hlavina W, Kimchi A, Kitts PA, Murphy TD, Pruitt KD, Souvorov A. P8008 the NCBI Eukaryotic Genome Annotation Pipeline. *J Anim Sci.* 2016;94(suppl_4):184–184. <https://doi.org/10.2527/jas2016.94supplement4184x>.
- Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 2008;453(7194):539–543. <https://doi.org/10.1038/nature06908>.
- Weick E-M, Miska EA. piRNAs: from biogenesis to function. *Development* 2014;141(18):3458–3471. <https://doi.org/10.1242/dev.094037>.
- Welker N, Maity T, Ye X, Aruscavage P, Krauchuk A, Liu Q, Bass B. Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode. *Mol Cell.* 2011;41(5):589–599. <https://doi.org/10.1016/j.molcel.2011.02.005>.
- Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol.* 2015;32(3):820–832. <https://doi.org/10.1093/molbev/msu400>.
- Wheeler BM, Heimberg AM, Moy VN, Sperling EA, Holstein TW, Heber S, Peterson KJ. The deep evolution of metazoan microRNAs. *Evol Dev.* 2009;11(1):50–68. <https://doi.org/10.1111/j.1525-142X.2008.00302.x>.

- Wu J, Yang J, Cho WC, Zheng Y. Argonaute proteins: structural features, functions and emerging roles. *J Adv Res.* 2020;24:317–324. <https://doi.org/10.1016/j.jare.2020.04.017>.
- Wynant N, Santos D, Vanden Broeck J. The evolution of animal Argonautes: evidence for the absence of antiviral AGO Argonautes in vertebrates. *Sci Rep.* 2017;7(1):9230. <https://doi.org/10.1038/s41598-017-08043-5>.
- Yang J-S, Lai EC. Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. *Mol Cell.* 2011;43(6):892–903. <https://doi.org/10.1016/j.molcel.2011.07.024>.
- Yerukhimovich MM, Marohnic CC, Frick DN. Role of the conserved DECH-box cysteine in coupling hepatitis C virus helicase-catalyzed ATP hydrolysis to RNA unwinding. *Biochemistry* 2018;57(43):6247–6255. <https://doi.org/10.1021/acs.biochem.8b00796>.
- Zdobnov EM, Kuznetsov D, Tegenfeldt F, Manni M, Berkeley M, Kriventseva EV. OrthoDB in 2020: evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* 2021;49(D1):D389–D393. <https://doi.org/10.1093/nar/gkaa1009>.

Associate editor: Selene Fernandez-Valverde