The Small Noncoding RNA Processing Machinery of Two Living Fossil Species, Lungfish and Coelacanth, Gives New Insights into the Evolution of the Argonaute Protein Family

Maria Assunta Biscotti[1], Adriana Canapa[1,*], Mariko Forconi[1], Marco Gerdol[2], Alberto Pallavicini[2], Manfred Schartl[3,4], and Marco Barucca[1]

[1]Dipartimento di Scienze della Vita e dell'Ambiente, Università Politecnica delle Marche, Ancona, Italy

[2]Dipartimento di Scienze della Vita, Università di Trieste, Italy

[3]Physiological Chemistry, Biocenter, University of Wuerzburg and Comprehensive Cancer Center Mainfranken, University Clinic Wuerzburg, Germany

[4]Texas Institute for Advanced Study and Department of Biology, Texas A&M University

*Corresponding author: E-mail: a.canapa@univpm.it.

Accepted: February 15, 2017

Data deposition: The sequences were submitted to EMBL and the accession numbers were added to the manuscript.

Abstract

Argonaute (AGO) family proteins play many roles in epigenetic programming, genome rearrangement, mRNA breakdown, inhibition of translation, and transposon silencing. Despite being a hotspot in current scientific research, their evolutionary history is still poorly understood and consequently the identification of evolutionary conserved structural features should also generate useful information for better understanding their functions. We report here for the first time the transcript sets of the two subfamilies, *Ago* and *Piwi*, in the West African lungfish *Protopterus annectens* and in the Indonesian coelacanth *Latimeria menadoensis*, two key species in the evolutionary lineage leading to tetrapods. The phylogenetic analysis of 142 inferred protein sequences in 22 fully sequenced species and the analysis of microsynteny performed in the major vertebrate lineages indicate an intricate pattern for the evolution of both subfamilies that has been shaped by whole genome duplications and lineage specific gains and losses. The *argonaute* subfamily was additionally expanded by local gene duplications at the base of the jawed vertebrate lineage. The subfamily of Piwi proteins is involved in several processes such as spermatogenesis, piRNA biogenesis, and transposon repression. Expression assessment of *AGO* genes and genes coding for proteins involved in small RNA biogenesis suggests a limited activity of the *Piwi* pathway in lungfish in agreement with the lungfish genome containing mainly old and inactive transposons.

Key words: argonaute, Piwi, Coelacanth, *Latimeria menadoensis*, lungfish, *Protopterus annectens*.

Introduction

RNA interference (RNAi) is a mechanism by which small RNAs are used as a guide in a broad range of processes such as mRNA breakdown, inhibition of translation, induction of histone and DNA methylation, heterochromatin formation as well as germline development, stem-cell self-renewal, and transposon silencing (Aravin et al. 2007; Höck and Meister 2008; Ross et al. 2014; Biscotti et al. 2015). Argonaute (AGO) proteins are highly specialized in binding these small RNA molecules. Members of this protein family can be divided into Argonaute subfamily (Ago) proteins (named after the identification of "Argonaute" subfamily in *Arabidopsis thaliana*) and PIWI proteins (named accordingly after the identification of the "Piwi" subfamily in *Drosophila melanogaster*). The former, also known as Eukaryotic translation initiation factors 2C (EIF2Cs) in vertebrates, are ubiquitously expressed and bind small interfering RNAs (siRNAs) and micro RNAs (miRNAs) (Carmell et al. 2002), while the latter are mainly expressed in germline cells and interact with PIWI-interacting RNAs (piRNAs) (Hutvagner and Simard 2008).

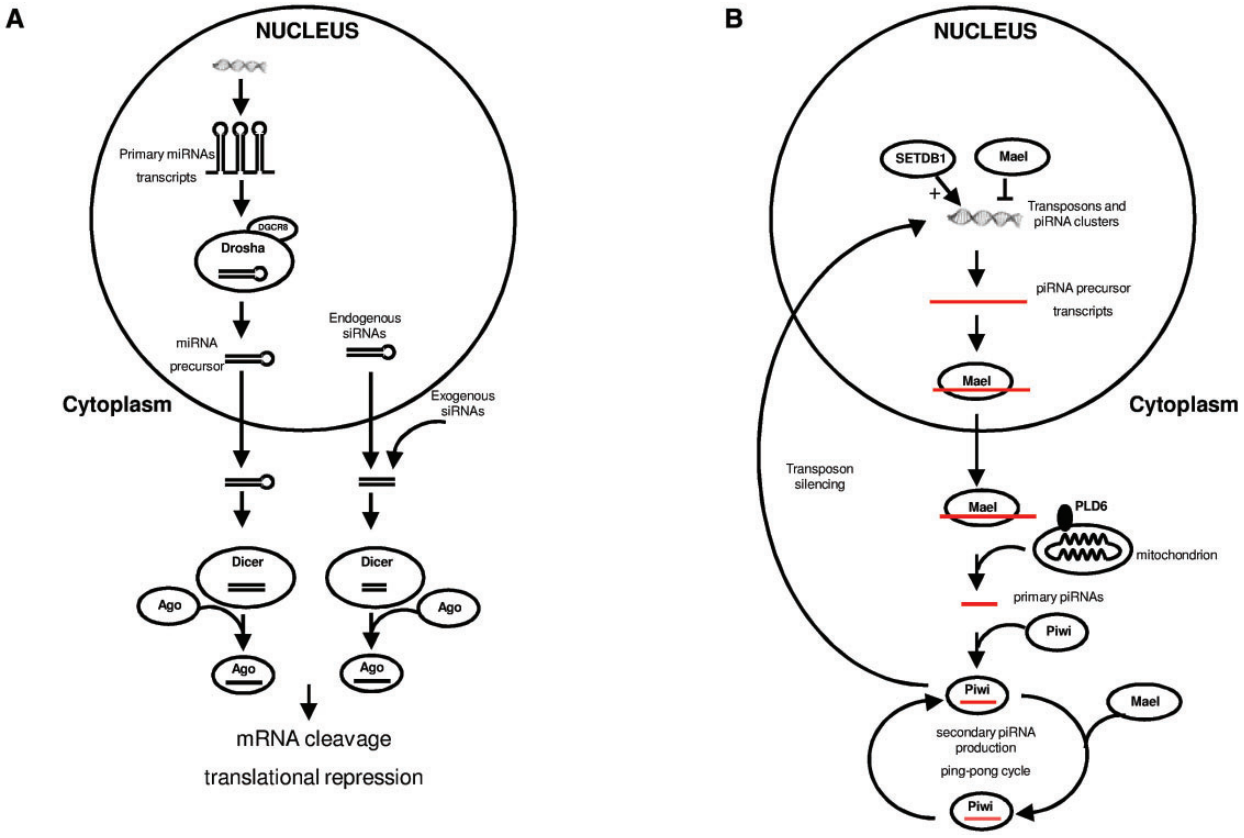

Fig. 1.—Biogenesis of small RNAs and cellular localization of proteins involved. (*A*) Biogenesis of miRNAs and siRNAs modified from Ender and Meister (2010). (*B*) Biogenesis of piRNAs modified from Ender and Meister (2010), Nishimasu et al. (2012), Weick and Miska (2014), and Pandey and Pillai (2014).

siRNAs and microRNAs (miRNAs) are produced by Dicer and Drosha and are loaded onto Ago proteins contained either in the RNA-induced initiation of transcriptional gene silencing complex (RITS) or in the RNA induced silencing complex (RISC). DiGeorge Syndrome Critical Region Gene 8 (*DGCR8*) encodes for a protein, which is a component of the microprocessor complex involved in miRNA biogenesis. This protein is required for binding the double-stranded RNA substrate and facilitates the cleavage by Drosha (Gregory et al. 2004). The small RNA molecules contained in the RITS and RISC complexes guide them to specific chromosome regions and to specific mRNAs by base-pairing interactions (fig. 1*A*).

piRNAs are produced by the slicer activity of PIWI proteins or through the nuclease PLD6 (Nishimasu et al. 2012), which is involved in primary piRNA production, and the HMG protein Maelstrom (Mael), which is involved in the primary and secondary piRNA production (Aravin et al. 2009; Castañeda et al. 2014). Moreover Mael is also a nucleo-cytoplasmic shuttling protein able to bind piRNA precursor transcripts and deliver them to cytoplasm (Pandey and Pillai 2014). This gene has been lost in teleost (Zhang et al. 2008). Recently, the deposition of H3K9me3 by the protein SETDB1 has been related with the activation of piRNA clusters, which generate the precursors of primary piRNAs (Rangan et al. 2011) (fig. 1*B*). piRNAs play a role in transposon silencing and spermatogenesis (Houwing et al. 2007; Malone and Hannon 2009; Thomson and Lin 2009; Yadav and Kotaja 2014).

The eukaryotic Argonaute proteins share a common structure characterized by a Piwi-Argonaute-Zwilli (PAZ) domain of about 120 aa, located in the N-terminal region, which binds the 3′ end of small RNAs. This domain appears to be absent in most prokaryotic Ago proteins (Wei et al. 2012). A 300 aa long PIWI domain, showing similarities to the RNase H catalytic domain, is usually conserved at the C-terminus, suggesting a possible role in the slicer activity for miRNAs and piRNAs biogenesis (McFarlane et al. 2011). The MID domain is located between the PAZ and PIWI domains and is around 120 aa long. It loads the small RNA onto Ago proteins, which then bind the 5′ end of the nucleic acid (Chen et al. 2008).

Phylogenetic analyses provided evidence for the presence of four groups corresponding to the worm-specific WAGO subfamily, the Trypanosome AGO family, the Ago subfamily, and the Piwi subfamily (Hernández and Jagus 2016). The former two are the result of lineage-specific duplications while the latter two are widespread among all kingdoms of living organisms suggesting that the last common ancestor of eukaryotes already had at least one Argonaute-like and one

Piwi-like gene which probably originated by duplication from an ancestral prokaryotic gene (Cerutti and Casas-Mollano 2006; Hutvagner and Simard 2008; Murphy et al. 2008, Swarts et al. 2014). Moreover the extant eukaryotic species display different gene numbers and lineage-specific loss or expansions of *Ago-piwi* genes (Cerutti and Casas-Mollano 2006; Höck and Meister 2008; McFarlane et al. 2011): for instance 8 *AGO* genes have been described in *Homo sapiens*, 5 genes have been recorded in *D. melanogaster*, 27 in *Caenorhabditis elegans*, 1 in *Schizosaccharomyces pombe*, and 10 in *A. thaliana* (Höck and Meister 2008; Zhou et al. 2010; Zheng 2013). However, little is known about the evolutionary diversification within each ago and piwi subfamilies across vertebrates.

The members of the Argonaute family have been identified in several jawed vertebrates (Gnathostomes), from teleosts to tetrapods (Cerutti and Casas-Mollano 2006; Höck and Meister 2008; McFarlane et al. 2011). No Argonaute gene has ever been reported in the two basal sarcopterygians, the coelacanths, and lungfish. The coelacanth *Latimeria menadoensis* and the lungfish *Protopterus annectens* offer unique opportunities to explore vertebrate gene evolution and function given their key phylogenetic position in the evolutionary lineage leading to tetrapods. The genus *Latimeria* is also of interest because, despite analyses on coding genes indicated a slowly evolving genome (Amemiya et al. 2013; Nikaido et al. 2013), transposable element (TE) activity appears to be comparable to "nonliving fossil" fish and would not indicate inertia of the coelacanth genome (Bejerano et al. 2006; Xie et al. 2006; Chalopin et al. 2014; Forconi et al. 2014; Naville et al. 2014). This raises the question of TE control through piRNAs and the associated enzymes. The West African lungfish *P. annectens* is one of the six extant species of dipnoi. This taxonomic group is characterized by a large genome, reaching even 38 folds the size of the human genome (Gregory 2014; Canapa et al. 2016). The outstanding increase in genome size has been related to uncontrolled proliferation of transposable elements within lungfish genomes (Metcalfe et al. 2012; Metcalfe and Casane 2013; Canapa et al. 2016). However, analyses of the *P. annectens* transcriptome revealed a low activity of TEs (Biscotti et al. 2016). No information on the TE silencing machinery in lungfish is currently available.

We performed a comprehensive phylogenetic analysis of the AGO gene family utilizing 142 protein sequences from 22 fully sequenced species. Synteny analyses supported our phylogenetic data and indicate an intricate pattern for the evolution of both subfamilies that has been shaped by whole genome duplications (WGD) and lineage specific gains and losses. The argonaute subfamily was additionally expanded by local gene duplications at the base of the jawed vertebrates lineage. The expression levels of *AGO* transcripts as well as of genes coding for proteins involved in small RNA biogenesis were investigated to detect the activity of the silencing pathways in which these proteins are involved. The activity of the *Piwi* pathway in lungfish suggests that large parts of its genome are made up of old and inactive transposons in agreement with previous hypotheses (Metcalfe et al. 2012; Metcalfe and Casane 2013; Biscotti et al. 2016).

Materials and Methods

AGO, *PIWI*, *DGCR8*, *Dicer*, *Drosha*, *PLD6*, *SETDB1*, and *Mael* transcripts were obtained from the *L. menadoensis* (Canapa et al. 2012; Pallavicini et al. 2013) and *P. annectens* transcriptomes (Biscotti et al. 2016) (supplementary table S1, Supplementary Material online; Accession numbers from LT674425 to LT674451). The raw sequence reads of the obtained transcriptomes were deposited in the NCBI BioProject and SRA databases under the accessions PRJNA175365 and PRJNA282925, respectively. Given the high sequence identity within the genus *Latimeria* (Inoue et al. 2005; Pallavicini et al. 2013), the genome of the congeneric species *L. chalumnae* was used to obtain synteny information of members belonging to the *AGO* family in coelacanths (Amemiya et al. 2013; Nikaido et al. 2013). The syntenic positions of the corresponding genes from the other vertebrates were collected from ENSEMBL (http://www.ensembl.org) (supplementary tables S2 and S3, Supplementary Material online) and checked through Genomicus (http://www.genomicus.biologie.ens.fr/genomicus-84.01/cgi-bin/search.pl).

The correct orthology of transcripts obtained in both species was assessed by homology using NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) while for *Dicer* and *SETDB1*, given the presence of several paralogous genes, the orthology was assessed by phylogenetic analyses (supplementary figs. S1 and S2 and table S4, Supplementary Material online). PAZ, MID, and PIWI domains of members of the AGO family were inferred through InterPro (http://www.ebi.ac.uk/interpro).

The phylogenetic analyses of the AGO family was performed on amino acidic sequences using MrBayes (version 3.1; Huelsenbeck et al. 2001). Substitution models, posterior probabilities, stationarity, generations, sampling, burnin, specific tree building parameters, and rooting details are reported in the tree legend. Moreover, the Maximum likelihood was performed using MEGA7 (Kumar et al. 2016) with Jones–Taylor–Thornton (Jones et al. 1992) model and the bootstrap support for the ML tree was determined using 1000 replications. The topology of ML tree (data not shown) is similar to that obtained with Bayesian analysis. The Ago and Piwi orthologous sequences were collected from ENSEMBL or NCBI databases. *Callorhinchus milii* sequences were obtained from http://esharkgenome.imcb.a-star.edu.sg/ (Venkatesh et al. 2014). Little skate, *Leucoraja erinacea*, Ago subfamily was inferred from the transcriptome at Skatebase (Skatebase.org, Wang et al. 2012), Argonaute RISC catalytic component 1 or *EIF2C1* (AGO1): contig 19580, contig 18487, contig 349, contig 28154; Argonaute RISC catalytic component 2 or *EIF2C2* (AGO2): contig 89915; Argonaute RISC catalytic component 3 or *EIF2C3* (AGO3): contig 22246, contig 15246,

contig 11106; Argonaute RISC catalytic component 4 or *EIF2C4* (AGO4): contig 90274, contig 89949. Accession numbers of the sequences used in the phylogenetic analysis are reported in supplementary tables S2 and S3, Supplementary Material online. The accession number for the WAGO sequence of *Caenorhabditis elegans* used a outgroup is Q21770. Clustal OMEGA was used to build the alignments (http://www.ebi.ac.uk/Tools/msa/clustalo/; Sievers et al. 2011). The assignment of the *Ago* and *Piwi* genes to ohnolog families was checked using the ohnolog database (http://ohnologs.curie.fr/) by Singh et al. (2015) and the data provided in Kasahara et al. (2007) and Nakatani et al. (2007).

The expression values in *L. menadoensis* male liver, testis, and muscle, in *P. annectens* brain, liver, and gonads of male and female specimens, and in *Danio rerio* brain, liver, muscle, and gonads of female specimen and gonads of a male specimen (BioProject PRJNA255848) are reported as transcripts per million (TPM). Expression levels were calculated following the procedure described in Biscotti et al. (2016) to allow interspecies comparison.

Omega (d_N/d_S) rates were calculated with CODEML, included in the PAML 4.8 package (Yang 2007), starting from the codon-based alignment of the coding nucleotide sequences of the target genes obtained with MUSCLE (Edgar 2004). Coding sequences were retrieved from ENSEMBL for *Mus musculus*, *Loxodonta africana*, *Monodelphis domestica*, *Pelodiscus sinensis*, *Gallus gallus*, *Xenopus tropicalis*, *D. rerio*, *Lepisosteus oculatus*, and *Tetraodon nigroviridis*, from http://esharkgenome.imcb.a-star.edu.sg/ (Venkatesh et al. 2014) for *C. milii* and from SkateBase (Skatebase.org, Wang et al. 2012) for *L. erinacea* and *Scyliorhinus canicula*. The accession IDs of the sequences used for each of the 15 genes selected (*AGO1*, *AGO2*, *AGO3a*, *AGO3b*, Piwi-like RNA-mediated gene silencing 1 (*PIWIL1*), Piwi-like RNA-mediated gene silencing 2 (*PIWIL2*), Piwi-like RNA-mediated gene silencing 4 (*PIWIL4*), *Dicer*, *Drosha*, *PLD6*, *Mael*, *SETDB1*, and *DGCR8*) are reported in supplementary table S4, Supplementary Material online. Missing data (gaps) were not considered and incomplete sequences (corresponding to less than 75% of the expected length) were discarded. Only informative codons were retained with Gblocks (Talavera and Castresana 2007) and the resulting alignments were converted in a Phylip format. We used the topology of species tree from Biscotti et al. (2016) to test the null (one-ratio model) and the alternative (multiple-ratio model) model hypotheses for each gene. The alternative hypothesis assumed different omega rates for the tetrapod, Actinopterygii, lungfish and coelacanth lineages. A likelihood ratio test was used to determine the significance of the data obtained, by comparing $2\Delta\log L$ with a χ^2 distribution. The two models were considered as producing statistically significant likelihoods for *P*-values lower than 0.05. β-actin (*ACTB*) was used as control gene.

The same data set was subjected to a Tajima's Relative Rate Test (RRT) analysis (Tajima 1993), using the sequences of the three chondrichthyan species (whenever available) as outgroups. Differences in the rate of evolution of lungfish sequences (ingroup I) compared to other vertebrate species (ingroup II) were considered as significant at $Pe < 0.05$. In parallel, a Maximum Likelihood Molecular Clock analysis was performed with MEGA 7 (Kumar et al. 2016) to test the null hypothesis of an equal evolutionary rate throughout the tree. This analysis was based on a NJ tree topology, under the general time reversible model of evolution with a discrete Gamma distribution of rates across sites (supplementary table S5, Supplementary Material online).

Results

Identification of AGO Family Genes in *Latimeria* and in *Protopterus*

Six transcripts related to the AGO protein family were retrieved in the Indonesian coelacanth transcriptome (fig. 2*A*, supplementary table S1, Supplementary Material online): three belonging to the Ago subfamily (*AGO2*, *AGO3*, and *AGO4*), and three from the Piwi subfamily (*PIWIL1*, *PIWIL2*, and *PIWIL4*). With the exception of *PIWIL2*, which is incomplete at the 5′ end, all transcripts harbored a complete coding sequence (CDS). Moreover, the assembly of *PIWIL1* uncovered multiple splicing isoforms with one of them only found in liver (supplementary fig. S3, Supplementary Material online).

Our analyses revealed that the ENSEMBL gene prediction of the African coelacanth genome was accurate for *AGO1*, *AGO2*, *AGO4*, *PIWIL1*, and *PIWIL2*. Discrepancies were detected for the other two genes: *AGO3*, so far not annotated in ENSEMBL, was identified on contig JH126588: 1,908,252–1,984,196. *PIWIL4* is truncated at its 5′ end in ENSEMBL gene predictions. Using our transcriptome sequence the missing coding region was identified on a different scaffold and was therefore manually assembled (table 1, supplementary table S6, Supplementary Material online).

Seven AGO transcripts displaying a complete CDS were detected in the West African lungfish (fig. 2*B*, table 1, supplementary table S1, Supplementary Material online): four belonging to the Ago subfamily (*AGO1-4*) and three to Piwi subfamily (*PIWIL1*, *PIWIL2*, and *PIWIL4*).

Piwi-like RNA-mediated gene silencing 3 (*PIWIL3*) was neither retrieved from the transcriptomes of lungfish and coelacanth nor from the noneutherian vertebrate genomes scrutinized here. This is in line with earlier data that *piwil3* is a eutherian-specific novelty subsequently lost in mouse (Murchinson et al. 2008).

The prediction of conserved protein domains revealed the presence of a GAGE domain in PIWIL1 of both species, a feature which has never been described before in any other organism. This domain is common to proteins from the GAGE family, which so far reported have been only in humans and which are characterized by an antigenic peptide recognized by cytotoxic T-cells (Zendman et al. 2002). Our analysis shows

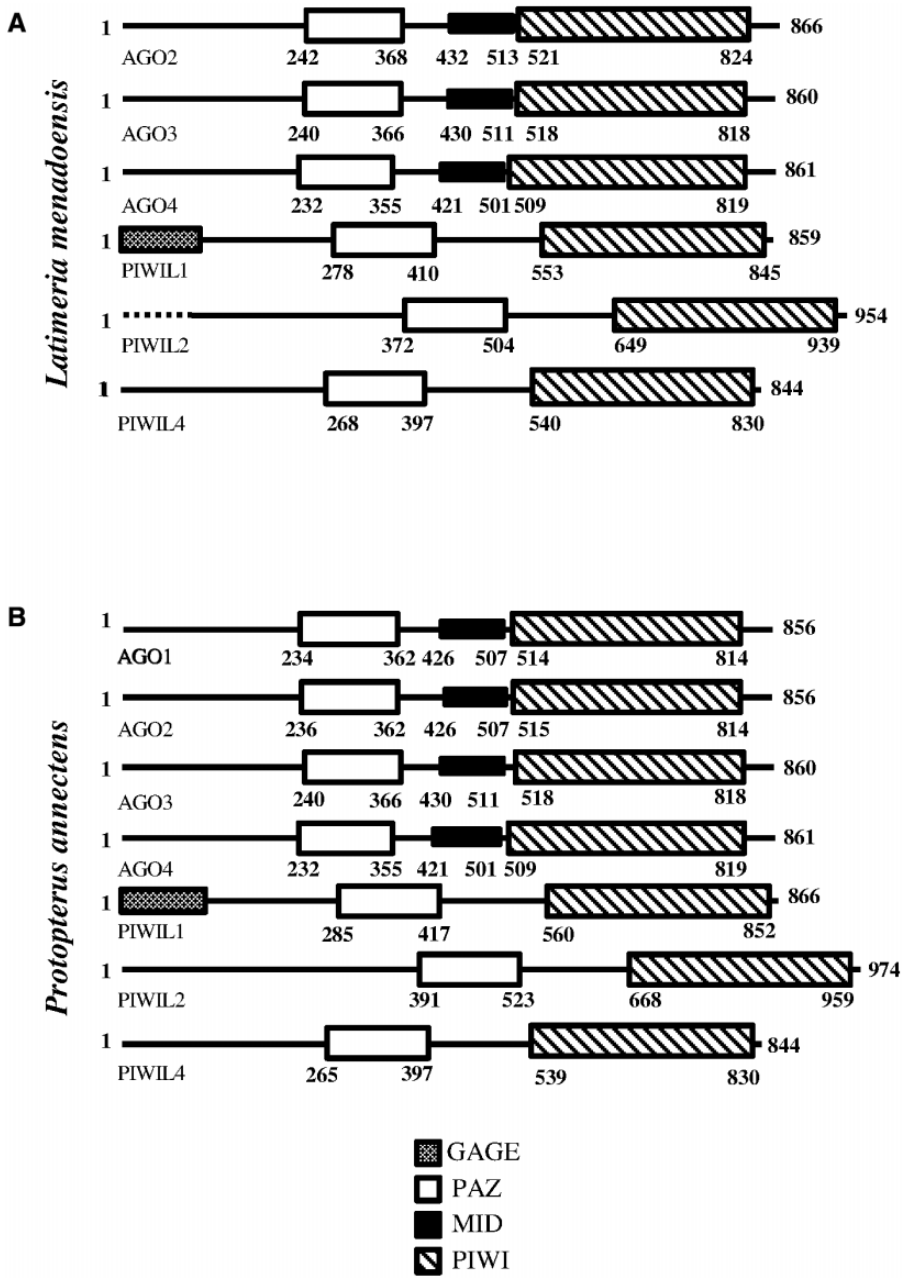

Fig. 2.—Schematic representation of AGO proteins. (*A*) Schematic representation of AGO members identified in the *L. menadoensis* transcriptome. (*B*) Schematic representation of AGO members identified in the *P. annectens* transcriptome. Dashed lines indicate a region of the *piwil2* sequence which was not identified in the *L. menadoensis* transcriptome but which is present in the *L. chalumnae* genome.

that this domain is present in all PIWIL1 proteins from Actinopterygii and Sarcopterygii.

Phylogeny and Microsynteny Conservation of the AGO Family

A phylogenetic analysis was performed to gain further information into the evolutionary history of the AGO family (fig. 3). The analysis revealed two main clades, one corresponding to *Ago* genes and the other to *Piwi* genes allowing the attribution of the identified sequences of coelacanth and lungfish to the two subfamilies.

For the Ago subfamily, the analysis revealed four distinct clades representing AGO1, AGO2, AGO4, and AGO3 with the *Ciona intestinalis* argonaute 2 sequence located external. All clades are supported by high values of posterior probability. Each group presents a similar topology: the chondrichthyan sequences are external and tied to a node, which includes Actinopterygii and Sarcopterygii. In all Ago sub-trees the

Table 1

Presence and Copy Number of *AGO* Genes in Representative Gnathostomes

Class	Species	Common Name	*AGO1*	*AGO2*	*AGO3*	*AGO4*	*PIWIL1*	*PIWIL2*	*PIWIL3*	*PIWIL4*
Chondrichthyes	*Callorhinchus milii*	Elephant shark	✓	✓	✓	✓	✓	✓		✓
Actinopterygians	*Danio rerio*	Zebrafish	✓	✓	✓✓	✓	✓	✓		
	Gadus morhua	Cod	✓	✓	✓✓	✓	✓	✓		
	Gasterosteus aculeatus	Stickleback	✓	✓	✓✓	✓	✓	✓		
	Oreochromis nilotica	Tilapia	✓	✓	✓✓	✓	✓	✓		
	Lepisosteus oculatus	Spotted gar	✓✓	✓		✓	✓	✓		
	Oryzias latipes	Medaka	✓	✓	✓✓	✓	✓	✓		
	Xiphophorus maculatus	Platyfish	✓	✓	✓	✓	✓	✓		
Sarcopterygians	*Anolis carolinensis*	Anole lizard	✓	✓	✓	✓	✓✓	✓		
	Bos taurus	Cow	✓	✓	✓	✓	✓	✓	✓	✓
	Canis familiaris	Dog	✓	✓	✓	✓	✓	✓	✓	✓
	Gallus gallus	Chicken	✓	✓	✓	✓	✓	✓		
	Homo sapiens	Man	✓	✓	✓	✓	✓	✓	✓	✓
	Latimeria	**Coelacanth**	✓	✓	✓	✓	✓	✓		✓
	Mus musculus	Mouse	✓	✓	✓	✓	✓	✓	✓	✓
	Ornithorhynchus anatinus	Platypus	✓	✓	✓	✓	✓✓	✓		
	Protopterus annectens	**West African lungfish**	✓	✓	✓	✓	✓	✓		✓
	Sarcophylus harrisii	Tasmanian devil	✓	✓	✓	✓	✓	✓	✓	✓
	Xenopus tropicalis	Western clawed frog	✓	✓	✓	✓	✓	✓		✓

NOTE.—Check signs indicate the presence of a gene/transcript in public databases. Multiple check signs indicate gene duplication events in the species. The species analysed in this work are in bold.

sequences of *Protopterus* and *Latimeria* form a sister group of tetrapods, with the exception of AGO3 where these sequences are located basal to the Actinopterygian clade. In teleosts AGO3 presents a duplication as result of the teleost-specific WGD. Indeed the two *Ago3* genes are located on chromosomes 11 and 16 in medaka and on chromosomes 16 and 19 in zebrafish. These two chromosome pairs are derived from the Teleost-specific Genome Duplication (TSGD) as described in Kasahara et al. (2007) and Nakatani et al. (2007).

The microsynteny analysis of *Ago* subfamily members evidenced a higher conservation in Sarcopterygians than teleosts. In shark and sarcopterygians *AGO4, AGO1*, and *AGO3* belong to the same cluster while in ray-finned fishes this cluster is disrupted. The microsynteny of *Ago* genes in Chondrichthyes and Sarcopterygii suggests that this arrangement was already present in the common ancestor of Gnathostomes.

In all analyzed species *AGO2* is located in a separate genomic region *leading* to the *hypothesis* that *AGO2* and the *AGO1/3/4* ancestral gene might be derived from a genome duplication event (fig. 4). The presence of only one *Ago* gene in *Ciona* which is located outside the other Ago clades in the tree suggests that *AGO2* and the *AGO 1/3/4* ancestor would be the result of 1R or 2R WGD and then local gene duplications generated *AGO1*, *AGO3*, and *AGO4*. This view is also supported by the results retrieved from the ohnolog database (Singh et al. 2015) that reports *AGO1/3/4* and *AGO2* as belonging to a same ohnolog family (families composed of paralogous genes derived from WGD events).

Phylogenetic analysis for the Piwi subfamily showed the presence of three distinct clades related to PIWIL1/PIWIL3, PIWIL2, and PIWIL4. PIWIL1/3, and PIWIL4 are sister groups and PIWIL2 is an external branch to both. Of the two Piwi sequences of *C. intestinalis* one is external to the other Piwi sequences and one is a sister branch of the Piwil2 clade.

In the PIWIL1, PIWL2, and PIWIL4 branches the topology is similar: the chondrichthyan, coelacanth and lungfish sequences are always external to tetrapods.

The microsynteny of *PIWIL1* is conserved between sarcopterygians, *L. oculatus*, and teleosts (fig. 5). The genomic arrangement of genes flanking *PIWIL2* and *PIWIL4* is less conserved. Indeed in *L. oculatus*, *C. milii*, and Sarcopterygians *PIWIL2* has a different upstream gene while in spotted gar microsynteny represents an intermediated pattern sharing the downstream gene with both shark and Sarcopterygians and the upstream gene with teleosts. The microsyntenic arrangement of *PIWIL4* is conserved in Chondrichthyes and Sarcopterygians suggesting that it was already present in the common ancestor of Gnathostomes and that the lack in the basal fish *L. oculatus* and in teleosts is specific to this lineage.

Even the genes belonging to the Piwi subfamily could be derived from WGD events as suggested from the results of the ohnolog database that, however, do not support an origin for *PIWIL2* from these events.

Expression of AGO Gene Family in *Latimeria* and in *Protopterus*

The evaluation of *Ago* gene transcripts revealed expression in all examined tissues (fig. 6*A* and *B*); while *Piwi* gene expression is restricted to gonad tissues with the exception of *Latimeria*

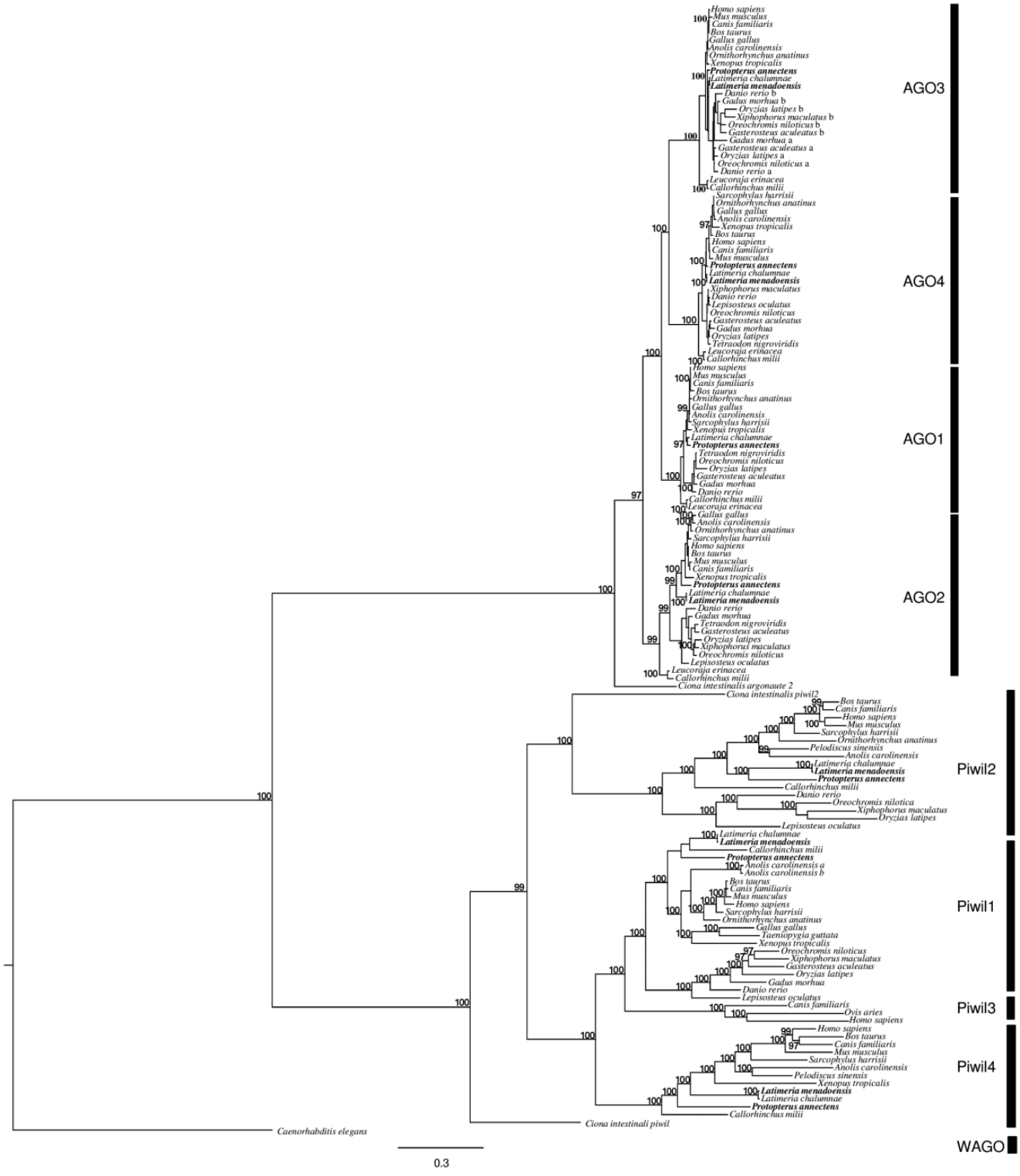

Fig. 3.—Phylogenetic tree of the AGO family. Bayesian inference. Amino-acidic model applied (mixed: Jones = with posterior probability 1.00, standard deviation 0.00). 6,000,000 generations, 15,000 as the burning. Stationarity defined as when the standard deviation of split frequencies reached 0.007, PSRF = 1.000. Mid-point rooting. The sequences in bold were obtained in this work. Black bars on the right represent the paralogy groups. Numbers close to nodes indicate posterior probability values. Only values > 95 are reported.

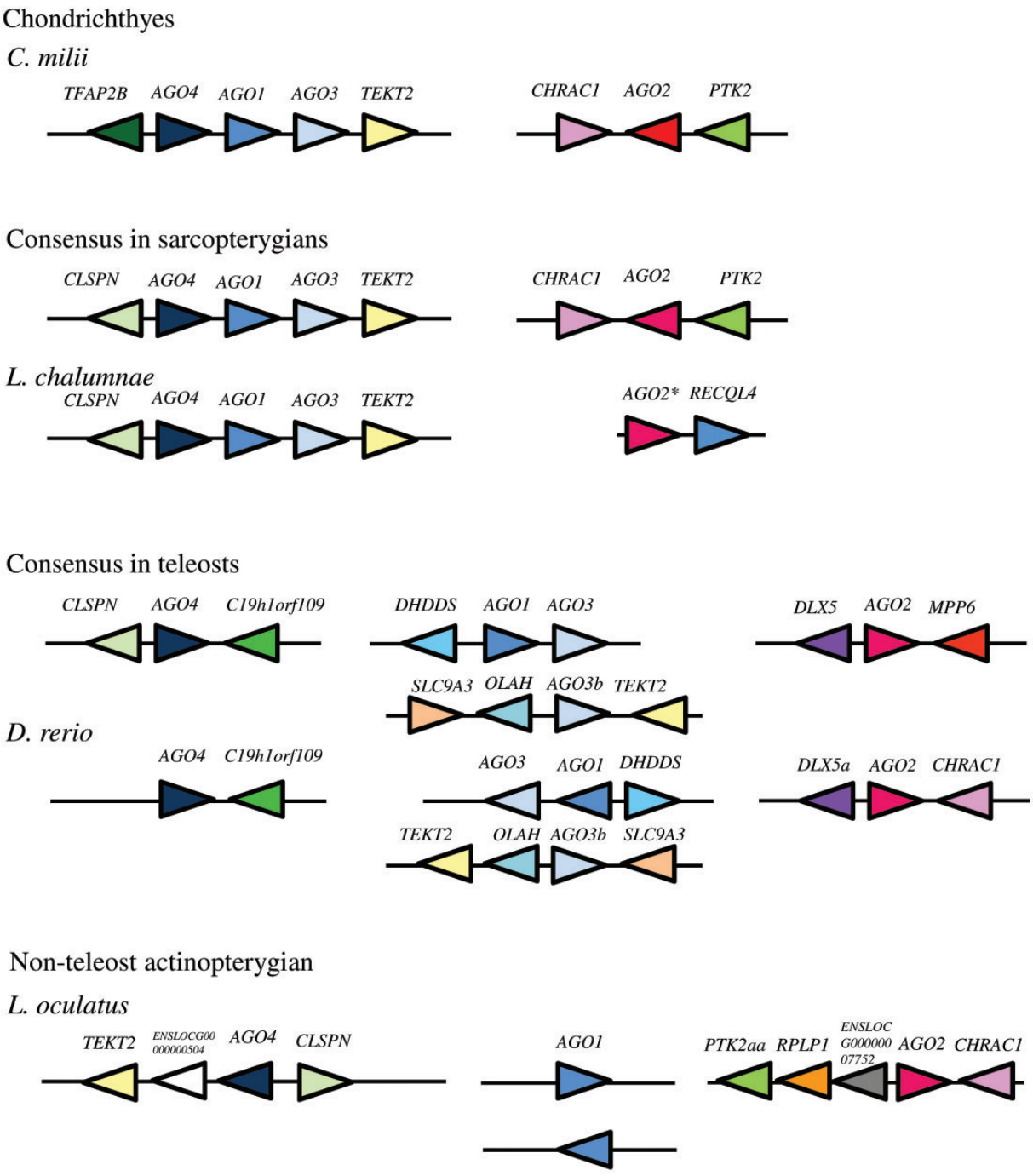

Fig. 4.—Synteny conservation of *Ago* genes in Gnathostomes. Arrowheads indicate 5′–3′ gene direction. Lines underneath genes indicate syntenic arrangement. *Indicates that this gene is annotated on multiple scaffolds (the mapping of the complete *L. menadoensis AGO2* cDNA to its African congener genome allowed to identify the 5′ UTR at the beginning of another scaffold JH127518, a region close to Protein tyrosine kinase 2 (*PTK2*), the tetrapod downstream flanking gene). Syntenic maps are reported for *C. milii*, *L. chalumnae*, *D. rerio* and *L. oculatus*. Consensus syntenic maps are reported for sarcopterygian and teleost clades, respectively. Genomic localizations of *Ago* genes in the analyzed vertebrate species are reported in supplementary table S2, Supplementary Material online.

PIWIL1 that additionally shows expression in liver due to a specific isoform resulting from alternative splicing (fig. 6*C* and *D*, supplementary fig. S3, Supplementary Material online).

In *L. menadoensis*, the expression levels of *Ago* and *Piwi* genes in three tissues are particularly high for *AGO2* and *PIWIL1* in testis (fig. 6*A* and *C*).

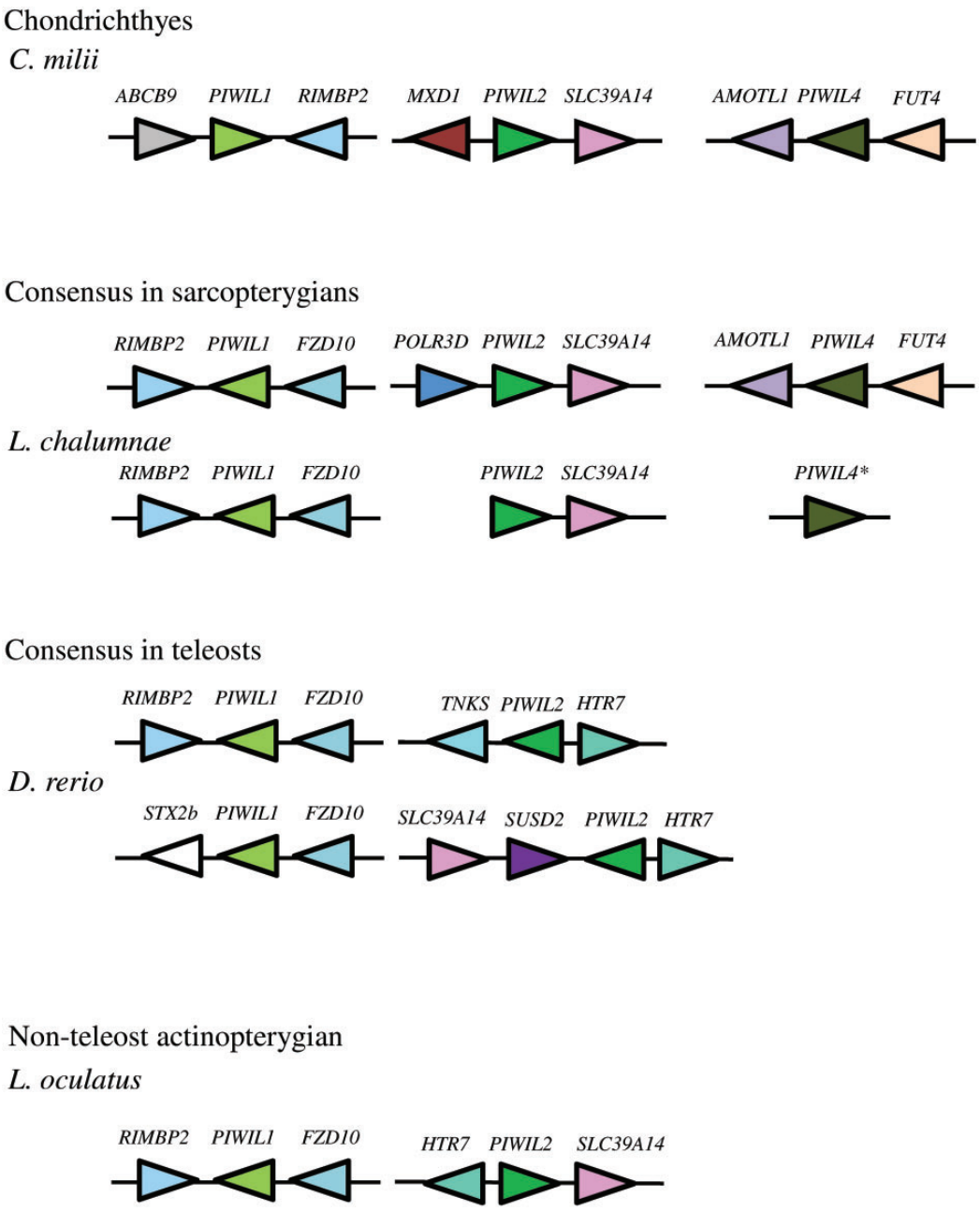

Fig. 5.—Synteny conservation of *Piwi* genes in Gnathostomes. Arrow heads indicate gene direction. Lines underneath genes indicate syntenic arrangement. * indicates that this gene is annotated on multiple scaffolds (JH128514, JH130604, JH134549). Syntenic maps are reported for *C. milii*, *L. chalumnae*, *D. rerio*, and *L. oculatus*. In addition, consensus syntenic maps are reported for the sarcopterygian and teleost clades, respectively. Genomic localizations of *Piwi* genes in the analyzed vertebrate species are reported in supplementary table S3, Supplementary Material online.

In *P. annectens*, *Ago* genes show comparable expression in brains and livers of both sexes and in male gonads. In female gonads, *AGO1*, *AGO2*, and *AGO4* have a significantly high activity, with the exception of *AGO3*, which in turn displays an expression level similar to other tissues (fig. 6*B*). In male gonads all three *Piwi* genes are expressed with the mature male gonad showing higher levels than the immature gonad. However, in female gonads *PIWIL4* is not expressed and *PIWIL1* shows values about twenty fold higher than *PIWIL2* (fig. 6*D*).

In coelacanth and in the mature male lungfish the expression patterns of the *Piwi* genes are similar, while the different expression values observed in the two lungfish male specimens could be related to the developmental stage (Kowalczykiewicz et al. 2012) (fig. 6*C* and *D*).

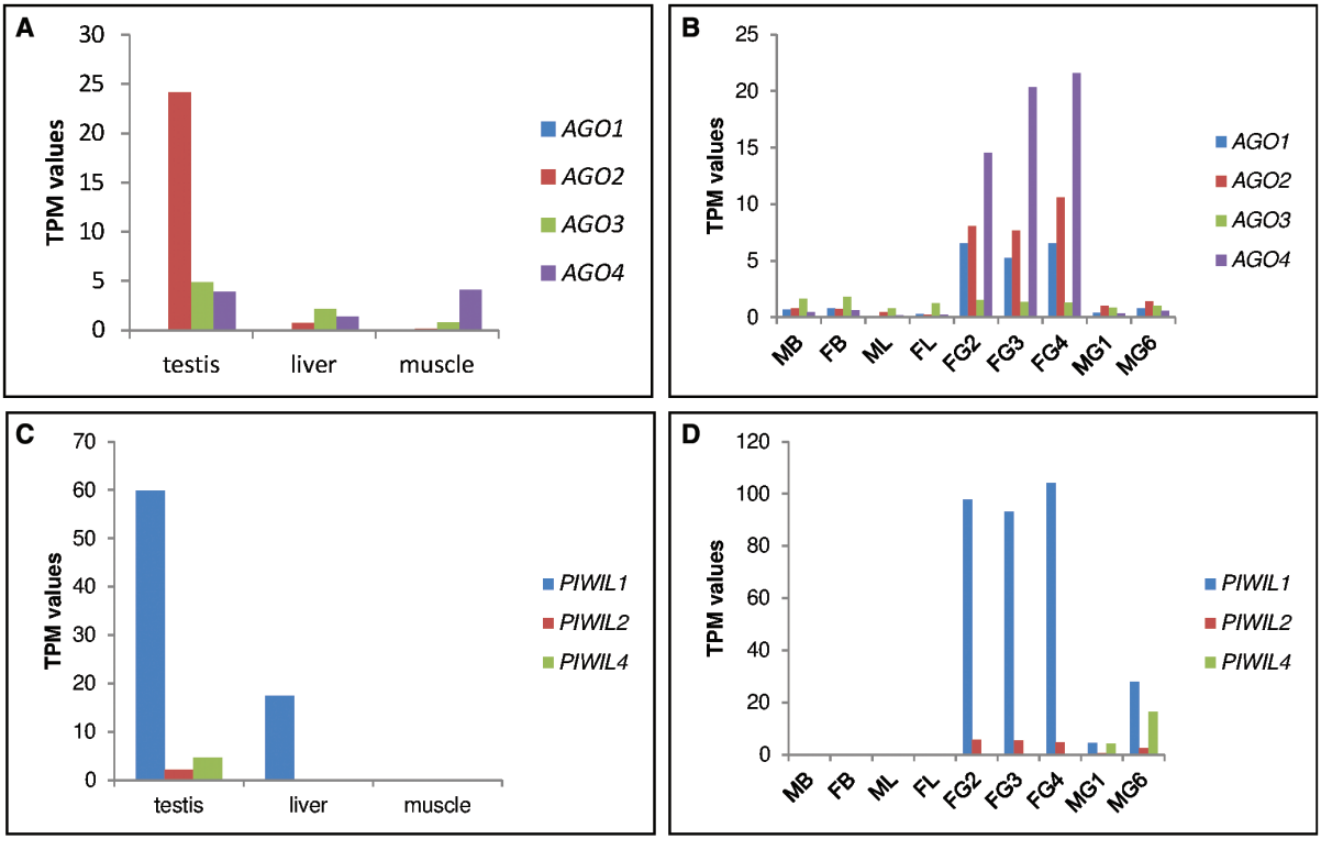

Fig. 6.—Expression levels of *AGO* genes. (*A*) Expression values of *Ago* genes in the male liver, testis, and muscle transcriptomes of *L. menadoensis*. (*B*) Expression values of *Ago* genes in the transcriptomes obtained from brain, liver, and gonad tissues of male and female specimens of *P. annectens*. (*C*) Expression values of *Piwi* genes in the male liver, testis, and muscle transcriptomes of *L. menadoensis*. (*D*) Expression values of *Piwi* genes in the transcriptomes obtained from brain, liver, and gonad tissues of male and female specimens of *P. annectens*.

Identification and Expression of Genes Involved in Small RNA Processing

Six transcripts coding for proteins involved in small RNAs production were investigated in both species. In *Latimeria Drosha*, *Mael*, *PLD6*, and *SETDB1* show complete coding sequence while *DGCR8* and the manually assembled *Dicer* transcript have a few nucleotides missing from the CDS. In *P. annectens* complete CDSs were retrieved for *DGCR8*, *Drosha*, *Mael*, and *SETDB1* while *Dicer* and *PLD6* are incomplete at their 3′ and 5′ ends, respectively.

Expression analysis revealed that in *Latimeria* all six genes are expressed in testis and muscle even though *PLD6* and *Mael* show very weak expression in the latter. In *P. annectens* *DGCR8*, *Drosha* and *SETDB1* show expression in all analyzed tissues; *Mael* is significantly expressed only in gonads; *Dicer* is expressed with very low levels in all tissues with the exception of male and female livers; *PLD6* was detected only in male gonads and with very low values also in female gonads. Notably the expression levels (except for *DGCR8*) in male mature gonad of lungfish are higher than in the immature specimen (fig. 7).

To test the correlation between transposon activity and the expression of genes involved in the silencing machinery, we analyzed also *D. rerio*, a teleost with high recent transposon activity (Chalopin et al. 2015). The comparison of the expression values, with the exception of *Mael* loss in teleosts, evidenced a higher expression in *D. rerio* compared to the two living fossil species (fig. 7).

Rates of Molecular Evolution

The evolutionary rates of 14 genes involved in piRNA and siRNA/miRNA pathways were evaluated through dN/dS, Tajima's RRT, and a Maximum Likelihood Molecular Clock analyses for each gene in Actinopterygii, coelacanth, lungfish and tetrapods. The molecular clock test revealed that genes involved in small RNA processing, as well *Ago* and *Piwi*, are unlikely to have evolved under molecular clock constraints (supplementary table S5, Supplementary Material online). Overall, *Piwi* genes as well as genes involved in small RNA processing evolve more rapidly than *Ago* genes with the exception of *AGO1* in coelacanth (fig. 8). Moreover in Actinopterygii, coelacanth, and lungfish the dN/dS ratio of

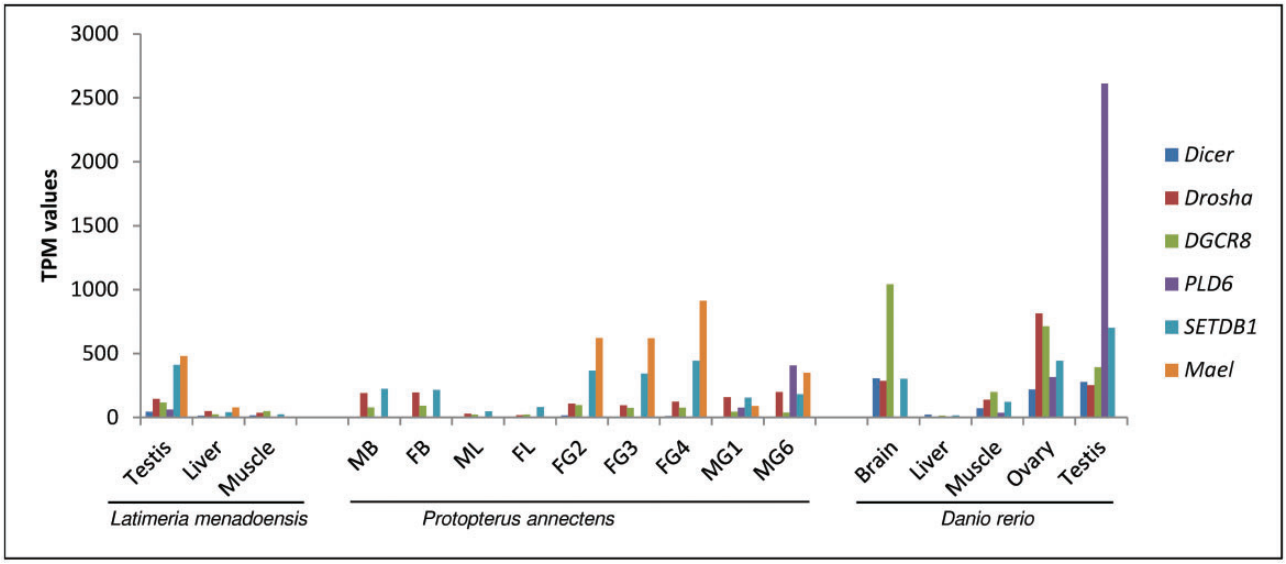

Fig. 7.—Expression levels of genes involved in small RNA biogenesis. Expression values of genes coding for protein involved in small RNA production in *L. menadoensis*, *P. annectens*, and *D. rerio*.

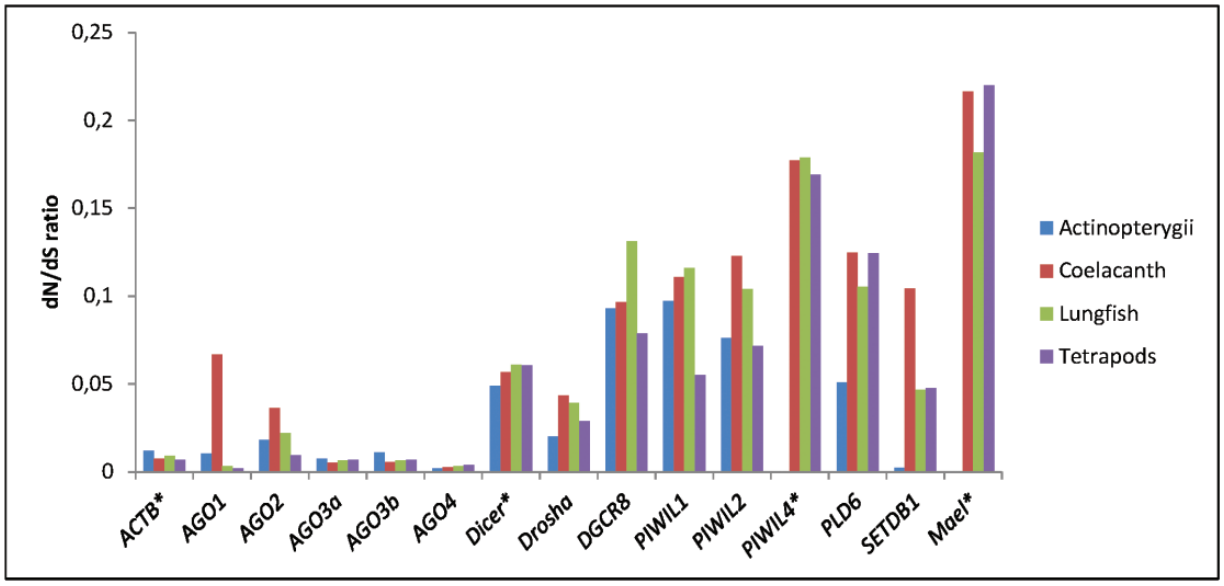

Fig. 8.—dN/dS ratio for 14 genes involved in piRNA and miRNA/siRNA pathways. dN/dS (omega) ratios observed in Actinopterygii, coelacanth, lungfish, and tetrapods. An asterisk marks genes where the alternative (multiple rates) model did not perform significantly better than the null (single rate) model.

PIWIL1 is significantly higher than in tetrapods while coelacanth, lungfish, and tetrapods present higher values than Actinopterygii for *PLD6* and *SETDB1*. However for *ACTB*, *Dicer*, *PIWIL4*, and *Mael* genes the alternative model did not fit sequence data significantly better than the null model, indicating that these genes evolved with similar dN/dS ratios in different lineages.

Discussion

AGO Family Evolution in Gnathostomes

The integration of our synteny conservation and phylogenetic data allows to draw now a clearer picture of the evolutionary history of the *AGO* family of the Gnathostomes. The presence

of seven genes in *Latimeria* and *Protopterus* is shared by Chondrichthyes and tetrapods suggesting that the common ancestor of Gnathostomes had these genes. Moreover in Actinopterygians PIWIL4 is absent and AGO3 underwent duplication.

The phylogenetic analyses, microsyntenic studies and the evidence that *Ago* genes are located in ohnolog regions (Kasahara et al. 2007; Nakatani et al. 2007, Singh et al. 2015) suggest that the *AGO*-like ancestral gene has undergone a WGD in Gnathostomes leading to the paralogs *AGO2* and the ancestral gene of *AGO1/3/4*. In the latter, two successive duplications led to the formation of *AGO1* and *AGO3/4* that in turn duplicated to generate the *AGO3* and *AGO4* genes. This scenario is in agreement with the microsynteny data showing that these three genes are arranged in a cluster (fig. 4). A further gene duplication of *AGO3* occurred in the teleost lineage. The occurrence of this duplication is probably linked to the TGD (Meyer and Schartl 1999) and the conservation of both genes and the close dN/dS values may suggest a sub-functionalization for *AGO3a* and *b*.

The *Ago* microsynteny analysis showed a parallel retention of the *loci* positions between Chondrichthyes and Sarcopterygians suggesting an ancestral condition in these two groups while the pattern in Actinopterygians represents a derived condition. Such similarity between Elasmobranchs and Sarcopterygii at the gene and genomic level, in comparison to the faster evolving modern teleosts, was also noted for other genes (Mulley and Holland 2010; Venkatesh et al. 2014).

The phylogenetic analysis for the *Piwi* subfamily, composed of three genes (*PIWIL1*, *PIWIL2*, and *PIWIL4*) in Gnathostomes, suggests that the ancestral gene *PIWI-like* has undergone a duplication that led to *PIWIL2* and to an ancestral form of *PIWIL1/4*. The occurrence of *PIWIL1* and *PIWIL4* seems to be the result of a WGD event. Another interesting aspect is related to the presence of *PIWIL4* in the sarcopterygian lineages and its absence in actinopterygians. In fact, the phylogenetic analysis as well as the microsynteny analysis suggest that this gene was already present in the common ancestor of Chondrichthyes and Osteichthyes, and consequently its absence in actinopterygians is due to gene loss in this lineage. The topology of the phylogenetic tree reveals a difference in branch lengths for the two subfamilies, namely generally longer branches for the *Piwi* sequences. The molecular clock test revealed that *Piwi* genes as well *Ago* genes and those involved in small RNA processing are unlikely to have evolved under molecular clock constraints which is consistent with the fast evolutionary rates of teleost genomes (Ravi and Venkatesh 2008; Amemiya et al. 2013; Venkatesh et al. 2014). While the results of Tajima's RRT were mostly consistent with previous reports, they also revealed that such a fast evolutionary rate also involves some components of the small RNA processing machinery in the nonteleost fish *L. oculatus* (namely, AGO2, AGO4, PIWIL1, and PIWIL2), whereas others (Dicer, Drosha, DGCR8, and Pld6) specifically experienced a faster evolutionary rate (compared to lungfish) in the teleost lineage. Overall the greater accumulation of mutations in the Piwi proteins compared with Ago proteins is due to a different mutation rate. The greatest divergence within the Piwi subfamily may be linked to a request for greater functional flexibility, given the wide range of functions attributed to these proteins. In particular in teleosts the rapid evolution of Piwi and proteins involved in the piRNA pathway might be linked to the higher diversity of transposons in this lineage (Yi et al. 2014). Our dN/dS analysis confirms the slow evolution of *Ago* genes with the exception of *AGO1* in coelacanth and evidences higher values of dN/dS ratio for *Piwi* genes and for genes involved in siRNA/miRNA and piRNA processing. In particular the higher values of *PIWIL1* in coelacanth and lungfish compared to tetrapods could be due to the high variability of transposable elements in these lineages (Forconi et al. 2014; Biscotti et al. 2016) as also noted in teleosts (Yi et al. 2014).

Expression Patterns of AGO Family Genes

Ago genes were expressed in all examined tissues of coelacanths and lungfish similar to other organisms (Zhou et al. 2010; McFarlane et al. 2011; Meister 2013). The different expression levels of *Ago* genes observed here might be related to the interaction of Ago proteins with different subsets of miRNA as demonstrated for the human orthologs (Azuma-Mukai et al. 2008). In *Latimeria* and *Protopterus*, the higher expression of Ago genes in gonads might indicate a role in gametogenesis. Indeed, in several organisms Ago proteins are involved in transcriptional gene regulation processes during oogenesis (Watanabe et al. 2008; Azzam et al. 2012; Leebonoi et al. 2015) and spermatogenesis (Borges et al. 2001; Nonomura et al. 2007; González-González et al. 2008; Leebonoi et al. 2015). In the shrimp *Penaeus monodon*, Ago proteins and in particular AGO4 have been also related to transposon silencing in gonads, but this mechanism is not yet understood (Leebonoi et al. 2015).

The expression pattern of *Dicer*, *Drosha*, and *DGCR8* genes, involved in siRNA and miRNA biogenesis, reveals a higher activity of *Dicer* and *DGCR8* in *Latimeria* compared to *P. annectens*. *AGO1*, *AGO2*, *AGO4*, and *Dicer* show a gonad activity in female lungfish, although the low values of *Dicer* suggest that its function could be replaced by the slicing activity of *AGO2* (Meister et al. 2004).

In male lungfish, the expression of *Ago* genes was not correlated with the developmental stage of the gonad, differently from *Drosha*, *PLD6*, *SETDB1*, and *Mael*. However it cannot be ruled out that this pattern is due to individual conditions or low sample size.

Besides expression of a *PIWIL1* specific alternative splicing isoform in *Latimeria, Piwi* genes displayed a gonad specific expression pattern in *L. menadoensis* and in *P. annectens*. The liver expression in the coelacanth

suggests that Piwi proteins might also play an important and different role outside gonadal tissues. Also in other organisms the expression of *Piwi* genes has been observed in somatic tissues (Yan et al. 2011; Lee et al. 2011; Kowalczykiewicz et al. 2012).

The expression in males is in line with other data that suggest an involvement of Piwi proteins in spermatogenesis (Deng and Lin 2002; Qiao et al. 2002; Kuramochi-Miyagawa et al. 2004; Carmell et al. 2007; ; Zhou et al. 2010; Chen et al. 2012; Zhao et al. 2012; Kowalczykiewicz et al. 2012).

In *P. annectens*, we demonstrated that *Piwi* genes are expressed in female gonads. The expression of these genes in ovaries is common to other oviparous species such as zebrafish (Houwing et al. 2007, 2008), medaka (Zhao et al. 2012) and chicken (Kim et al. 2012). However, expression of *Piwi* genes and in particular of *PIWIL1* in female gonads has been reported in pig with values significantly lower than those detected in testis (Zhou et al. 2010; Kowalczykiewicz et al. 2012) and has been detected in adult platypus and human ovaries (Lim et al. 2013) suggesting that the role of these proteins could be independent from the reproduction strategy. Overall, the function of Piwi proteins in the ovary is still not completely understood and investigation on a larger number of organisms should be useful to obtain further insights (Ma et al. 2014).

A role of Piwi proteins in piRNA biogenesis has been recognized early on (Aravin et al. 2007; O'Donnell and Boeke 2007). They have been associated with primary and secondary piRNA processing (Aravin et al. 2007; Brennecke et al. 2007; Lim et al. 2013). Primary piRNAs derive from the cleavage of long transcripts produced from genomic loci called piRNA clusters. The enzyme involved is the endonuclease PLD6/MITOPLD/Zucchini. These piRNAs initiate the production of secondary piRNAs which self-amplify through a Ping-Pong pathway involving Piwi proteins (Siomi et al. 2011). The protein Mael is also involved in this pathway (Aravin et al. 2009), but not in teleost fish, where this gene has been lost (Zhang et al. 2008). The coexpression of *Mael* and *PIWIL2* in Sertoli cells and granulosa cells has been related to involvement of the encoded proteins in secondary piRNAs processing (Aravin et al. 2009; Lim et al. 2013; Kowalczykiewicz et al. 2014).

Expression of the *PLD6* gene was detected in male gonads of both species but not in female lungfish suggesting that its activity might be replaced by other enzymes. Conversely, the coexpression of *Mael* and *PIWIL2* in all the samples supports a conserved role in secondary piRNA biogenesis in the basal sarcopterygians. The lower expression of *Mael* in lungfish could indicate that the production of secondary piRNAs is less active than in coelacanths. The weak production of piRNAs in lungfish is also supported by the lower expression of the methyltransferase *SETDB1*. In *Drosophila*, this enzyme is involved in the deposition of H3K9me3 that actives piRNA cluster transcription with consequent production of piRNAs that control transposon activity (Rangan et al. 2011). Recent studies have indicated that Mael is also essential for Piwi-mediated silencing of transposons in *Drosophila* (Klenov et al. 2011; Sienski et al. 2012). Indeed another important function of Piwi proteins in gonads is their involvement in transposon silencing to ensure genome integrity.

Coelacanths and lungfish are two organisms of outstanding interest in evolutionary biology because of their living fossil status. Indeed their morphology remained highly similar to their ancestors that lived 400 MYA. Moreover in both species the morphological stasis reflects also a molecular stasis at the coding-gene level (Amemiya et al. 2013; Biscotti et al. 2016). In this respect, it is interesting to analyze the activity of transposable elements, which are considered as major drivers of genome shaping (Warren et al. 2015; Canapa et al. 2016). Despite slow evolution of coding sequences, mobile elements show an opposite trend in both species (Chalopin et al. 2014; Forconi et al. 2014; Biscotti et al. 2016) and the huge expansion of the lungfish genome has been related to the accumulation of mobile elements (Metcalfe et al. 2012; Metcalfe and Casane 2013; Biscotti et al. 2016; Canapa et al. 2016). In this context, it is interesting to consider *piwi* gene activity given their involvement in transposon silencing (Houwing et al. 2007, 2008; Aravin et al. 2007; Kuramochi-Miyagawa et al. 2008; Siomi et al. 2011). Despite the high number of expected transposable elements responsible for the enormous size of the lungfish genomes (Metcalfe et al. 2012; Metcalfe and Casane 2013), which is about 38 fold of the coelacanth genome (Makapedua et al. 2011), transposable elements showed lower expression values in *P. annectens* than in coelacanths (Biscotti et al. 2016). The expression levels of the Piwi pathway genes in *P. annectens*, comparable with those observed in *Latimeria* and lower than those observed in *D. rerio*, could be due to the fact that Piwis have a limited role in transposon silencing. This finding is in agreement with the hypothesis of a lungfish genome made up of mainly nonactive mobile elements (Metcalfe et al. 2012; Metcalfe and Casane 2013; Biscotti et al. 2016). Consequently, most of the mobile elements in lungfish genome may not need to be silenced by the Piwi pathway. However, as also observed for *Latimeria* (Amemiya et al. 2013; Forconi et al. 2014), the activity of transposable elements (Biscotti et al. 2016) in a lineage with apparent morphological stasis, as typical feature of the "living fossil status", evidences that the lungfish genome is not completely inert from the evolutionary point of view.

Similarily, the evolutionary rates of the genes for the small noncoding RNA processing machinery are not considearbly different for the lungfish and coelacanth orthologs. Thus, the status of being a living fossil on the morphological level, despite being somehow linked to decreased rates of moleular

evolution in protein-coding genes (Amemiya et al. 2013), does not necessarily implicate a low genomic plasticity.

Supplementary Material

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

Authors Contributions

M.A.B., A.C., M.B., and M.F. involved in the identification and characterization of *AGO* genes, phylogenetic analyses, microsynteny, generation of the lungfish expression data, and manuscript writing; M.S. helped in the generation of the lungfish expression data and contributed to interpretation of the data and manuscript writing; A.P. and M.G. subjected in expression analyses and dN/dS analyses. All authors have made contributions to conception and design and have given final approval of the version to be published.

Acknowledgments

This work was supported by a grant from Ministero della Ricerca e dell'Istruzione. We thank the anonymous reviewers for their constructive comments, which helped us to improve the manuscript.

Literature Cited

Amemiya CT, et al. 2013. The African coelacanth genome provides insights into tetrapod evolution. Nature 496:311–316.
Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. Science 316:744–747.
Aravin AA, et al. 2009. Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. PLoS Genet. 5:e1000764.
Azzam G, Smibert P, Lai EC, Liu J-L. 2012. Drosophila Argonaute 1 and its miRNA biogenesis partners are required for oocyte formation and germline cell division. Dev Biol. 365:384e94.
Azuma-Mukai A, et al. 2008. Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. Proc Natl Acad Sci U S A. 105:7964–7969.
Bejerano G, et al. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature 441:87–90.
Brennecke J, et al. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128:1089–1103.
Biscotti MA, Canapa A, Forconi M, Olmo E, Barucca M. 2015. Transcription of tandemly repetitive DNA: functional roles. Chromosome Res. 23:463–477.
Biscotti MA, et al. 2016. The lungfish transcriptome informs about molecular evolution events at the transition from water to land. Sci Rep 6:21571.
Borges F, Pereira PA, Slotkin RK, Martienssen RA, Becker JD. 2001. MicroRNA activity in the Arabidopsis male germline. J Exp Bot 62:1611e20.
Canapa A, et al. 2012. Composition and phylogenetic analysis of vitellogenin coding sequences in the Indonesian Coelacanth *Latimeria menadoensis*. J Exp Zool B Mol Dev Evol. 318:404–416.
Canapa A, Barucca M, Biscotti MA, Forconi M, Olmo E. 2016. Transposons. Genome size, and evolutionary insights in animals. Cytogentic Genome Res. DOI:10.1159/000444429
Carmell MA, Xuan Z, Zhang MQ, Hannon GJ. 2002. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev. 16:2733–2742.
Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH. 2007. Miwi2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell 12:503–514.
Castañeda J, et al. 2014. Reduced pachytene piRNAs and translation underlie spermiogenic arrest in Maelstrom mutant mice. EMBO J. 33:1999–2019.
Cerutti H, Casas-Mollano JA. 2006. On the origin and functions of RNA-mediated silencing: from protists to man. Curr Genet. 50:81–99.
Chalopin D, et al. 2014. Evolutionary active transposable elements in the genome of the coelacanth. J Exp Zool B Mol Dev Evol. 322:322–333.
Chalopin D, Naville M, Plard F, Galiana D, Volff JN. 2015. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome Biol Evol. 7:567–580.
Chen PY, et al. 2008. Strand-specific 5′-O-methylation of siRNA duplexes controls guide strand selection and targeting specificity. RNA 14:263–274.
Chen R, et al. 2012. Cloning of the quail PIWI gene and characterization of PIWI binding to small RNAs. PLoS One 7:e51724.
Deng W, Lin H. 2002. Miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. Dev Cell 2:819–830.
Ender C, Meister G. 2010. Argonaute proteins at a glance. J Cell Sci. 123:1819–1823.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.
Forconi M, et al. 2014. Transcriptional activity of transposable elements in coelacanth. J Exp Zool B Mol Dev Evol. 322:379–389.
Gregory RI, et al. 2004. The Microprocessor complex mediates the genesis of microRNAs. Nature 432:235–240.
Gregory TR. 2014. Animal Genome Size Database. http://www.genomesize.com.
González-González E, López-Casas PP, del Mazo J. 2008. The expression patterns of genes involved in the RNAi pathways are tissue-dependent and differ in the germ and somatic cells of mouse testis. Biochim Biophys Acta. 1779:306–311.
Hernández G and Jagus R, editors. 2016. Evolution of the protein synthesis machinery and its regulation. Switzerland: Springer Press.
Höck J, Meister G. 2008. The Argonaute protein family. Genome Biol. 9:210.
Houwing S, et al. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell 129:69–82.
Houwing S, Berezikov E, Ketting RF. 2008. Zili is required for germ cell differentiation and meiosis in zebrafish. EMBO J. 27:2702–2711.
Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310–2314.
Hutvagner G, Simard MJ. 2008. Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol. 9:22–32.
Inoue JG, Miya M, Venkatesh B, Nishida M. 2005. The mitochondrial genome of Indonesian coelacanth *Latimeria menadoensis* (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. Gene 349:227–235.
Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282.
Kasahara M, et al. 2007. The medaka draft genome and insights into vertebrate genome evolution. Nature 447:714–719.
Kim TH, et al. 2012. Conserved functional characteristics of the PIWI family members in chicken germ cell lineage. Theriogenology 78:1948–1959.
Klenov MS, et al. 2011. Separation of stem cell maintenance and transposon silencing functions of Piwi protein. Proc Natl Acad Sci U S A. 108:18760–18765.

Kowalczykiewicz D, Pawlak P, Lechniak D, Wrzesinski J. 2012. Altered expression of porcine Piwi genes and piRNA during development. PLoS One 7:e43816.
Kowalczykiewicz D, et al. 2014. Characterization of *Sus scrofa* small non-coding RNAs present in both female and male gonads. PLoS One 9:e113249.
Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 33:1870–1874.
Kuramochi-Miyagawa S, Kimura T, Ijiri TW, Isobe T, Asada N. 2004. Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. Development 131:839–849.
Kuramochi-Miyagawa S, et al. 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev. 22:908–917.
Lee EJ, et al. 2011. Identification of piRNAs in the central nervous system. RNA 17:1090–1099.
Leebonoi W, Sukthaworn S, Panyim S, Udomkit A. 2015. A novel gonad-specific Argonaute 4 serves as a defense against transposons in the black tiger shrimp Penaeus monodon. Fish Shellfish Immunol 42:280–288.
Lim SL, et al. 2013. Conservation and expression of PIWI-interacting RNA pathway genes in male and female adult gonad of amniotes. Biol Reprod 89:136.
Ma X, et al. 2014. Piwi is required in multiple cell types to control germline stem cell lineage development in the Drosophila ovary. PLoS One 9:e90267.
Makapedua MD, et al. 2011. Genome size, GC percentage and 5mC level in the Indonesian coelacanth *Latimeria menadoensis*. Mar Genomics 4:167–172.
Malone CD, Hannon GJ. 2009. Small RNAs as guardians of the genome. Cell 136:656–668.
McFarlane L, et al. 2011. Expansion of the Ago gene family in the teleost clade. Dev Genes Evol. 221:95–104.
Meister G, et al. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell 15:185–197.
Meister G. 2013. Argonaute proteins: functional insights and emerging roles. Nat Rev Genet. 14:447–459.
Metcalfe CJ, Filée J, Germon I, Joss J, Casane D. 2012. Evolution of the Australian lungfish (*Neoceratodus forsteri*) genome: a major role for CR1 and L2 LINE elements. Mol Biol Evol. 29:3529–3539.
Metcalfe CJ, Casane D. 2013. Accommodating the load: The transposable element content of very large genomes. Mob Genet Elements 3:e24775.
Meyer A, Schartl M. 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr Opin Cell Biol. 11:699–704.
Mulley JF, Holland PW. 2010. Parallel retention of Pdx2 genes in cartilaginous fish and coelacanths. Mol Biol Evol. 27:2386–2391.
Murphy D, Dancis B, Brown JR. 2008. The evolution of core proteins involved in microRNA biogenesis. BMC Evol Biol. 8:92.
Nakatani Y, Takeda H, Kohara Y, Morishita S. 2007. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res. 17:1254–1265.
Naville M, Chalopin D, Volff JN. 2014. Interspecies insertion polymorphism analysis reveals recent activity of transposable elements in extant coelacanths. PLoS One 9:e114382.
Nikaido M, et al. 2013. Coelacanth genomes reveal signatures for evolutionary transition from water to land. Genome Res. 23:1740–1748.
Nishimasu H, et al. 2012. Structure and function of Zucchini endoribonuclease in piRNA biogenesis. Nature 491:284–287.
Nonomura K, et al. 2007. A germ cell specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. Plant Cell 19:2583e94.
O'Donnell KA, Boeke JD. 2007. Mighty Piwis defend the germline against genome intruders. Cell 129:37–44.
Pallavicini A, et al. 2013. Analysis of the transcriptome of the Indonesian coelacanth *Latimeria menadoensis*. BMC Genomics 14:538.
Pandey RR, Pillai RS. 2014. Primary piRNA biogenesis: caught up in a Maelstrom. EMBO J. 33:1979–1980.
Qiao D, Zeeman AM, Deng W, Looijenga LH, Lin H. 2002. Molecular characterization of hiwi, a human member of the piwi gene family whose overexpression is correlated to seminomas. Oncogene 21:3988–3999.
Rangan P, et al. 2011. piRNA production requires heterochromatin formation in Drosophila. Curr Biol. 21:1373–1379.
Ravi V, Venkatesh B. 2008. Rapidly evolving fish genomes and teleost diversity. Curr Opin Genet Dev. 18:544–550.
Ross RJ, Weiner MM1, Lin H. 2014. PIWI proteins and PIWI-interacting RNAs in the soma. Nature 505:353–359.
Singh PP, Arora J, Isambert H. 2015. Identification of ohnolog genes originating from whole genome duplication in early vertebrates, based on synteny comparison across multiple genomes. PLoS Comput Biol. 11:e1004394.
Sievers F, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 7:539.
Sienski G, Dönertas D, Brennecke J. 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell 151:964–980.
Siomi MC, Sato K, Pezic D, Aravin AA. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol. 12:246–258.
Swarts DC, et al. 2014. The evolutionary journey of Argonaute proteins. Nat Struct Mol Biol. 21:743–753.
Tajima F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135:599–607.
Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 56:564–577.
Thomson T, Lin H. 2009. The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. Annu Rev Cell Dev Biol. 25:355–376.
Venkatesh B, et al. 2014. Elephant shark genome provides unique insights into gnathostome evolution. Nature 505:174–179.
Wang Q, et al. 2012. Community annotation and bioinformatics workforce development in concert - Little skate genome annotation workshops and jamborees. Database 2012:bar064.
Warren IA, et al. 2015. Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. Chromosome 23:505–531.
Watanabe T, et al. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature 453:539–e543.
Wei KF, Wu LJ, Chen J, Chen YF, Xie DX. 2012. Structural evolution and functional diversification analyses of argonaute protein. J Cell Biochem. 113:2576–2585.
Weick EM, Miska EA. 2014. piRNAs: from biogenesis to function. Development 141:3458–3471.
Xie X, Kamal M, Lander ES. 2006. A family of conserved noncoding elements derived from an ancient transposable element. Proc Natl Acad Sci U S A. 103:11659–11664.
Yi M, et al. 2014. Rapid evolution of piRNA pathway in the teleost fish: implication for an adaptation to transposon diversity. Genome Biol Evol. 6:1393–1407.
Yadav RP, Kotaja N. 2014. Small RNAs in spermatogenesis. Mol Cell Endocrinol 382:498–508.
Yan Z, et al. 2011. Widespread expression of piRNA-like molecules in somatic tissues. Nucleic Acids Res. 39:6596–6607.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
Zendman AJ, Van Kraats AA, Weidle UH, Ruiter DJ, Van Muijen GN. 2002. The XAGE family of cancer/testis-associated genes: alignment and expression profile in normal tissues, melanoma lesions and Ewing's sarcoma. Int J Cancer 99:361–369.
Zhang D, Xiong H, Shan J, Xia X, Trudeau VL. 2008. Functional insight into Maelstrom in the germline piRNA pathway: a unique domain homologous to the DnaQ-H 3'-5' exonuclease, its lineage-specific expansion/loss and evolutionarily active site switch. Biol Direct. 3:48.
Zhao H, Duan J, Cheng N, Nagahama Y. 2012. Specific expression of Olpiwi1 and Olpiwi2 in medaka (*Oryzias latipes*) germ cells. Biochem Biophys Res Commun. 418: 592–417.
Zheng Y. 2013. Phylogenetic analysis of the Argonaute protein family in platyhelminths. Mol Phylogenet Evol. 66:1050–1054.
Zhou X, Guo H, Chen K, Cheng H, Zhou R. 2010. Identification, chromosomal mapping and conserved synteny of porcine Argonaute family of genes. Genetica 138:805–812.

Associate editor: Marta Barluenga