

Rapid Evolution of piRNA Pathway in the Teleost Fish: Implication for an Adaptation to Transposon Diversity

Minhan Yi¹, Feng Chen¹, Majing Luo¹, Yibin Cheng¹, Huabin Zhao², Hanhua Cheng^{1,*}, and Rongjia Zhou^{1,*}

¹Department of Genetics, College of Life Sciences, Wuhan University, Wuhan, People's Republic of China

²Department of Zoology, College of Life Sciences, Wuhan University, Wuhan, People's Republic of China

*Corresponding author: E-mail: hhcheng@whu.edu.cn; rjzhou@whu.edu.cn.

Accepted: May 14, 2014

Data deposition: The cDNA sequences for *Piwil1*, *Piwil2*, *Tdrd1*, *Tdrkh*, *Prmt5*, *Ago1*, *Ago2*, *Ago3a*, *Ago3b*, and *Ago4* of Swamp eel have been deposited at GenBank under the accessions KF663555, KF663556, KF663558, KF663549, KF663557, KF663550, KF663551, KF663552, KF663553, and KF663554.

Abstract

The Piwi-interacting RNA (piRNA) pathway is responsible for germline specification, gametogenesis, transposon silencing, and genome integrity. Transposable elements can disrupt genome and its functions. However, piRNA pathway evolution and its adaptation to transposon diversity in the teleost fish remain unknown. This article unveils evolutionary scene of piRNA pathway and its association with diverse transposons by systematically comparative analysis on diverse teleost fish genomes. Selective pressure analysis on piRNA pathway and miRNA/siRNA (microRNA/small interfering RNA) pathway genes between teleosts and mammals showed an accelerated evolution of piRNA pathway genes in the teleost lineages, and positive selection on functional PAZ (Piwi/Ago/Zwille) and Tudor domains involved in the Piwi-piRNA/Tudor interaction, suggesting that the amino acid substitutions are adaptive to their functions in piRNA pathway in the teleost fish species. Notably five piRNA pathway genes evolved faster in the swamp eel, a kind of protogynous hermaphrodite fish, than the other teleosts, indicating a differential evolution of piRNA pathway between the swamp eel and other gonochoristic fishes. In addition, genome-wide analysis showed higher diversity of transposons in the teleost fish species compared with mammals. Our results suggest that rapidly evolved piRNA pathway in the teleost fish is likely to be involved in the adaptation to transposon diversity.

Key words: teleost fish, evolution, positive selection, reproduction.

Introduction

The Piwi-interacting RNA (piRNA) pathway is a small RNA silencing system, which functions in germline specification, gametogenesis, transposon silencing, genome integrity, and stem cell maintenance across the animal phylogeny. piRNAs are endogenous small noncoding RNA molecules ranging from 26 to 31 nucleotides in length. Animals lacking piRNAs exhibit activation of transposable elements (TEs) and defects in gametogenesis (Carmell et al. 2007). Piwil1 (Piwi-like 1) and Piwil2 (Piwi-like 2) are core proteins interacted with piRNAs (Aravin et al. 2006; Girard et al. 2006; Houwing et al. 2007, 2008) and act to repress TEs both transcriptionally and posttranscriptionally. Other components involved in piRNA pathway have also been identified in both vertebrates and *Drosophila* (Ishizu et al. 2012). TDRDs proteins (Tudor domain-containing proteins) can associate with both Piwil1 and Piwil2 through their N-terminal symmetrical dimethyl arginines (sDMAs) in both mouse and zebrafish

(Chuma et al. 2006; Chen et al. 2009; Reuter et al. 2009; Huang, Houwing, et al. 2011). The protein arginine methyltransferase 5 (Prmt5) catalyzes methylation of the arginines of the sDMAs (Kirino et al. 2009; Vagin et al. 2009). Mitochondrial protein Pld6 (Phospholipase D member 6, as known as MitoPLD) is also involved in primary piRNA biogenesis in the mouse germline (Huang, Gao, et al. 2011; Watanabe et al. 2011). Genetic analyses have revealed that the mouse Ddx4 (DEAD box polypeptide 4, a.k.a. Vasa, Mvh) is essential for the piRNA ping-pong cycling (Kuramochi-Miyagawa et al. 2010). Asz1 (or Gasz) (Ankyrin repeat, SAM, and basic leucine zipper domain containing 1) plays a role in retrotransposon silencing by stabilizing Piwil2 (MILL) in nuage (Ma et al. 2009). Both Mov10l1 (Moloney leukemia virus 10-like 1) and Fkbp6 (FK506 binding protein 6) are involved in delivering piRNAs to the mouse Piwi proteins (Frost et al. 2010; Zheng et al. 2010; Xiol et al. 2012). Kinesin motor protein Kif17 (Kinesin family member 17) interacts with the

mouse *Piwil1* and acts in the nucleocytoplasmic transport in the chromatoid bodies (Kotaja et al. 2006). In addition, *Hen1* (*HEN1* methyltransferase homolog 1) is responsible for piRNA stability, through methylation of piRNA 3'-end of both mouse and zebrafish (Kirino and Mourelatos 2007; Kamminga et al. 2010). These piRNA pathway genes, which are evolutionarily conserved across the animal phylogeny, act jointly to protect the genome from invasive TEs in the germline (Ishizu et al. 2012).

A substantial amount of piRNAs is derived from TEs, which play a crucial role in silencing TEs (Aravin et al. 2007). TEs are selfish genetic elements that can mobilize within a genome, potentially leading to serious genome damage. piRNAs act as guardians of the genome and protect it from invasive TEs in the germline cell (Kalmykova et al. 2005). In the process of gametogenesis, derepression of even a single active TE may cause infertility, a disaster from a Darwinian point of view.

The types of transposons vary dramatically in vertebrate species, ranging from a few to over a hundred (Malone and Hannon 2009). In *Drosophila*, there are approximately 150 different TE types, which have different expression, replication, and mobilization strategies. There are much more TE types in teleost fishes than that in mammals, whereas some types are absent from the mammalian genomes (Aparicio et al. 2002; Volff et al. 2003). Around 20 clades of retrotransposons are present in pufferfish genomes, whereas only six clades were detected in human and mouse genomes (Volff et al. 2003). Notably, fish genomes contain over 30 types of LINE-1 retrotransposons, each of which is active, whereas mammalian genomes possess only one (Furano et al. 2004). About 4.8% transcriptome of the platyfish is derived from TE sequences which represent 40 different families, suggesting that many TEs are still active (Schartl et al. 2013). DNA transposons in fish species are often active, whereas they are inactive in mammals (Bohne et al. 2012). Although TE family numbers are lower in mammals than in fish species, TE copy numbers are generally much higher in mammals than in fishes (Volff 2005). Therefore, TEs in fish genomes appear to have undergone a higher turnover than in the mammalian genomes (Duvernell et al. 2004), which is similar to that observed in *Drosophila* (Eickbush and Furano 2002).

In mouse, piRNAs are mainly derived from three classes of retrotransposons: LTR, LINE, and SINE (Aravin et al. 2007, 2008). However, zebrafish piRNAs have been observed to match to both retrotransposons and DNA transposons (Houwing et al. 2007, 2008) and are divided into more than 20 groups (Zhou et al. 2010). Diverse piRNAs may provide an adaptive defense to repress diversified transposons in teleosts. These observations prompt us to test whether high diversity and turnover rate of TEs in fish have driven the evolution of piRNA pathway genes, which is involved in TE silencing.

On the basis of their variety and abundance, TEs present an imposing threat to the germline cells and their genomes. With regard to the germline cells, they have a special need to

protect their genome, because genetic information must be faithfully transmitted to offspring. Thus, it is crucial to repress TE activity to ensure the integrity of the germline genomes. As a result, piRNA machinery was suggested to have undergone adaptive evolution, indicating that TE activity has driven the evolution of piRNA pathway (Malone and Hannon 2009). Thus it seems that as a result from a coevolutionary arms race, there is a process of reciprocal adaptation between piRNA machinery and TE activities. Several recent studies observed evidences of selection in the piRNA pathway in *Drosophila* (Vermaak et al. 2005; Castillo et al. 2011; Kolaczowski et al. 2011; Simkin et al. 2013). So far, studies on adaptive evolution of piRNA pathway have been largely confined to the *Drosophila* species. As such, adaptive evolution of piRNA pathway remains unknown in vertebrates.

To test a possible association between piRNA pathway evolution and TE diversity in the teleost fish species, we conducted a comparatively evolutionary analysis for piRNA pathway genes between the teleost fish species and mammals, and compared these genes with Ago genes that function in miRNA/siRNA pathway. The swamp eel (*Monopterus albus*) that belongs to the order Synbranchiformes is a protogynous fish, which can transform its sex from female via intersex into male (Zhou et al. 2001; Cheng et al. 2003). We tested whether the selective pressure of piRNA pathway differs between swamp eel (a kind of hermaphroditic fish) and the other fish species. We found an accelerated evolution of piRNA pathway genes in the teleost lineage. Notably, we detected many positively selected sites distributed on functional domains of piRNA pathway genes on the ancestral branch of teleosts, suggesting adaptive evolution of piRNA pathway genes as organisms evolve. In addition, we observed that five piRNA pathway genes in the swamp eel evolved more rapidly than other fish species, and their high expression level in three types of gonads (ovary, ovotestis, and testis) during sex reversal suggests the crucial role of piRNA pathway genes during the sex reversal. To further test whether high diversity and turnover rate of TEs exist in fish, we observed various types of fish TEs in comparison with mammals, suggesting a link of TEs diversity to the rapid evolution of piRNA pathway in fish species. Thus, this study highlights that rapidly evolved piRNA pathway is likely to adapt to transposon diversity in the teleost fish lineages.

Materials and Methods

Data Preparation

The coding sequences of 18 genes (*Piwil1*, *Piwil2*, *Tdrd1*, *Tdrkh*, *Pld6*, *Ddx4*, *Asz1*, *Mov1011*, *Fkbp6*, *Kif17*, *Prmt5*, *Hen1*, *Ago1*, *Ago2*, *Ago3a*, *Ago3b*, *Ago4*, and β -actin) in 29 vertebrates (21 mammals and 8 fishes) were obtained from Ensembl (www.ensembl.org, last accessed May 31, 2014) and National Center for Biotechnology Information

(NCBI; <http://www.ncbi.nlm.nih.gov>, last accessed May 31, 2014) (supplementary table S1, Supplementary Material online). Five miRNA/siRNA pathway genes (*Ago1*, *Ago2*, *Ago3a*, *Ago3b*, and *Ago4*), belonging to Ago subfamily in the Argonaute family, and β -actin were used as control genes. Two analyses were performed to ensure orthology among genes. First, we blasted these genes in GenBank (<http://www.ncbi.nlm.nih.gov>, last accessed May 31, 2014) to ensure the high hits. Second, ClustalW is used to align the orthologous genes (Chenna et al. 2003) and to check their conservation among vertebrate species. A few coding sequences in some species (supplementary table S1, Supplementary Material online) are not available in Ensembl or NCBI because of unfinished genome sequencing. Some partial sequences with large gaps or sequencing errors were discarded in this study.

Phylogenetic Analysis

The protein sequences of 30 vertebrates were aligned with ClustalW program (Chenna et al. 2003) and manual adjustments. The maximum-likelihood tree was constructed using PhyML 3.0 (Guindon and Gascuel 2003; Guindon et al. 2010) with LG model (Le and Gascuel 2008). To root the tree, lamprey *Piwil1* was used as an outgroup. The phylogeny was analyzed using 100 replicates. The neighbor joining (NJ) trees using bootstrap analysis with 1,000 replicates were constructed using MEGA 5 (Tamura et al. 2011) with Poisson model.

Synonymous and Nonsynonymous Distance

The coding sequences of 18 genes were aligned separately with ClustalW (Chenna et al. 2003) and manual adjustments. The sequences of each gene from both mammals and teleosts were divided into two groups (one group: 21 mammals, and the other: 9 teleost fish species). Synonymous (d_s) and nonsynonymous (d_n) were calculated using the modified Nei–Gojobori method (Zhang et al. 1998) implemented in MEGA 5 (Tamura et al. 2011). The ω ratios (d_n/d_s) were calculated to compare the evolutionary distance for each gene between two groups.

Branch Model Test

To test various selection pressure acting on different clades, we used branch model based on the ω ratio (d_n/d_s) in the codeml in PAML 4 package (Yang 2007). A species tree of 30 vertebrates (from Ensembl and teleost phylogeny; Chen and Mayden 2010) was used to determine the evolutionary relationship. We first tested the one-ratio model as the null hypothesis, in which the entire tree has one ω ratio. Two-ratio model was then used (a background ratio ω_0 and a different ratio ω_1 on the foreground branch). The significance for the likelihood ratio tests (LRTs) was tested using χ^2 analysis. $2\Delta\ln L$

was used (twice the difference of log likelihood between two models) in this analysis.

Detection of Positive Selection

To detect positively selected sites, the branch-site model in the program codeml in PAML 4 was used (Yang 2007). We set the ancestral branch of teleosts as foreground and the other lineages as background branch. The test 2 in PAML 4 was used to compare two models. In model A, three ω ratios ($0 < \omega_0 < 1$, $\omega_1 = 1$, $\omega_2 > 1$) are set for foreground and two ω ratios ($0 < \omega_0 < 1$, $\omega_1 = 1$) for background. Null model parameters are the same as in model A, whereas $\omega_2 = 1$. LRTs were calculated to test the significance of differences between two models. If the test shows positive selection, the sites under selection were analyzed according to high Bayes empirical Bayes posterior probabilities (BEB PPs) (Yang et al. 2005). All positively selected sites were aligned to the protein sequences and functional domains in the Pfam database (Coghill et al. 2008; Punta et al. 2012).

Sliding Window Analysis

To analyze distribution of all sites with the ω values (d_n/d_s) across the *Piwil1* sequence, we performed sliding window analysis of the ω values of each interval estimated by SWAAP (Pride 2000) with window size 60 bp and step size 9 bp. The *Piwil1* sequences from 20 mammals and 8 teleosts were used in the analysis.

Homology Modeling

Three-dimensional (3D) protein structures of four zebrafish proteins (*Piwil1*-PAZ domain, *Piwil2*-PAZ domain, *Tdrd1*-the third Tudor domain, *Tdrkh*-Tudor domain) were built using the SWISS-MODEL server (<http://swissmodel.expasy.org>, last accessed May 31, 2014) (Schwede et al. 2003) with solved 3D homologous proteins as templates. The crystal structures of templates were obtained from the Protein Data Bank (PDB). The PDB ID (identity) are as follows: 3O7V (human *Piwil1*-PAZ) (Tian et al. 2011), 3O7X (human *Piwil2*-PAZ) (Tian et al. 2011), 4B9W (mouse *Tdrd1*-the third Tudor domain) (Mathioudakis et al. 2012), 3FDR (human *Tdrkh*-Tudor) (Chen et al. 2009). The models were visualized by Swiss-PDB Viewer v3.7 and rendered with POV-Ray v3.6 (Persistence of Vision Ray Tracer, USA). The positively selected sites distributed on these domains were highlighted as red spheres to show their locations in the protein structures.

Degenerate Polymerase Chain Reaction, Rapid Amplification of cDNA Ends, and Reverse Transcription Polymerase Chain Reaction

The swamp eels (*M. albus*) were obtained from markets in Wuhan, China. Their sexes were confirmed by microscopic analysis of gonad sections (ovary, ovotestis, and testis). Total

RNAs were isolated from adult tissues using TRIZOL (Invitrogen, CA). SMART cDNAs were reversely transcribed from the RNAs using the SMART cDNA synthesis kit following protocol (Clontech, CA). Degenerate PCR was used to amplify the conserved regions of ten genes (*Piwil1*, *Piwil2*, *Tdrd1*, *Tdrkh*, *Prmt5*, *Ago1*, *Ago2*, *Ago3a*, *Ago3b*, and *Ago4*) using multiple degenerate primers (supplementary table S2, Supplementary Material online). PCR was performed in a 20 μ l reaction mix containing 10 mM Tris-HCl, pH 8.3, 1.5 mM MgCl₂, 50 mM KCl, 150 mM dNTP, 0.2 μ M each primer, and 1 unit of Taq DNA polymerase. Based on the conserved regions sequences, 3'-RACE of these genes was performed using common primer CDSIII: 5'-ATTCTAGA GCGCGAGGCGCCGACATGd(T)₃₀N₁N-3' (N = A, G, C, or T; N₁ = A, G, or C) and gene-specific primers (supplementary table S3, Supplementary Material online). 5'-RACE of these genes were performed using common SMARTIII

primer: 5'-AAGCAGGGTTCACGCAGAGTGGCCATTACGGC CGGG-3' and gene-specific primers (supplementary table S3, Supplementary Material online). All sequences were cloned and sequenced. Semiquantitative RT-PCR was performed for five Ago genes (*Ago1*, *Ago2*, *Ago3a*, *Ago3b*, and *Ago4*) and six piRNA pathway genes (*Piwil1*, *Piwil2*, *Tdrd1*, *Tdrkh*, *Ddx4*, and *Prmt5*). *Hprt* was used as a control. The cDNAs from adult tissues were amplified using gene-specific PCR primers (supplementary table S4, Supplementary Material online).

TE Analysis

To analyze diversity of TEs in fish genomes, we combined homology-based and de novo approaches to identify TEs in genomes of six teleost fish species. Zebrafish (*Danio rerio*), tetraodon (*Tetraodon nigroviridis*), fugu (*Takifugu rubripes*),

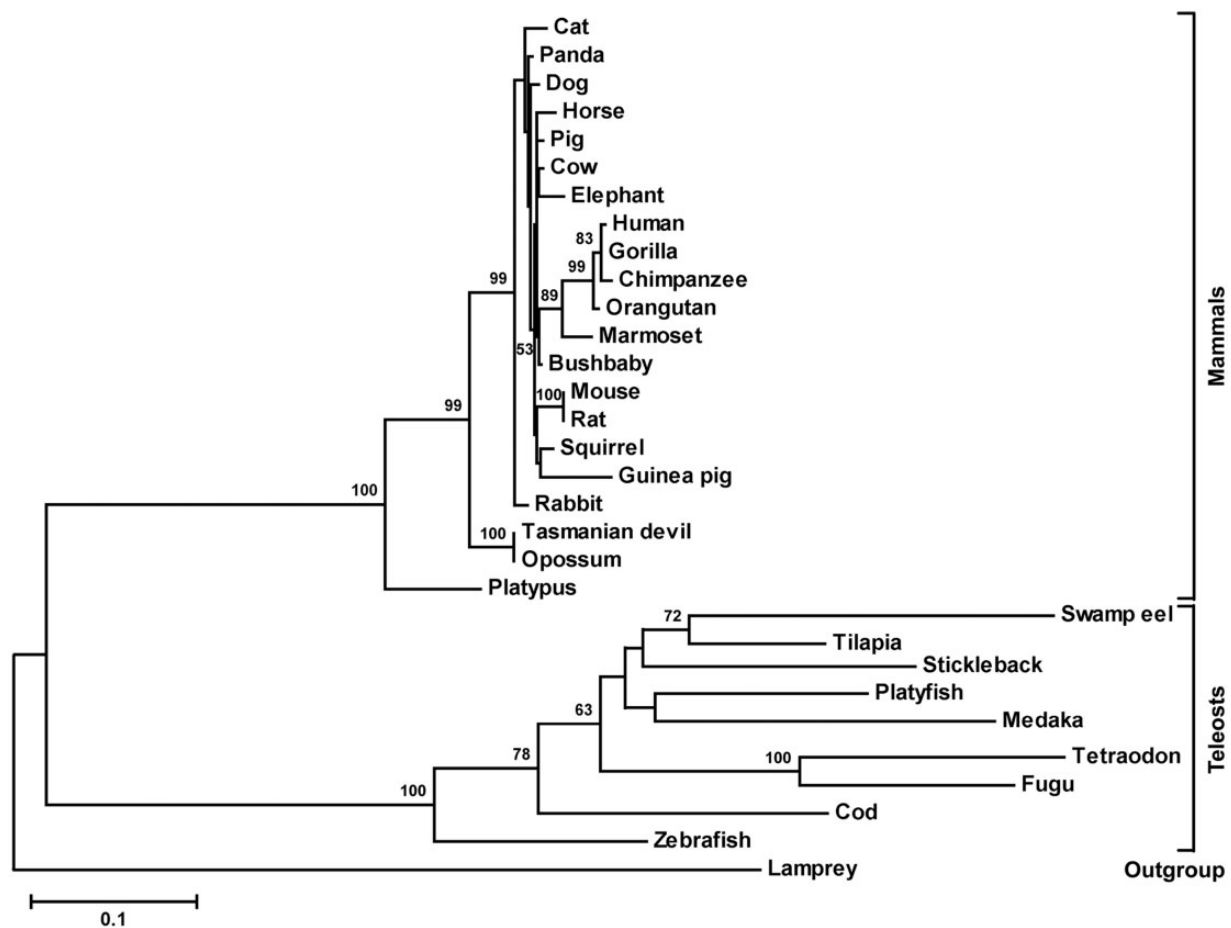


FIG. 1.—Phylogenetic tree of *Piwil1* in 30 vertebrates. Phylogenetic analysis was performed with PhyML and the lamprey (an order of jawless fish, also a very ancient lineage of vertebrates) *Piwil1* was used as an outgroup. Bootstrap values (percentage) were estimated using bootstrap analysis with 100 replicates and the values above 50% are shown next to the nodes. The number of amino acid substitutions per site was used to assess evolutionary distances. The distance scale (0.1) is shown. The accession numbers of all sequences in 30 vertebrates are listed in supplementary table S1, Supplementary Material online.

medaka (*Oryzias latipes*), and stickleback (*Gasterosteus aculeatus*) genome data were downloaded from Ensembl release 63. Swamp eel (*M. albus*) genome (DDBJ/EMBL/GenBank under the accession AONE00000000) was sequenced in our laboratory recently. The homology-based approach utilized database Rebase (release 16.02) with RepeatMasker (version 3.2.9) (Tarailo-Graovac and Chen 2009). The de novo approach utilized two prediction programs (RepeatModeler

version 1.0.4 [Price et al. 2005] and LTR-FINDER version 1.0.5) to build the de novo repeat libraries based on the genome sequences. The multicopy genes and contaminations were removed from the libraries. Then, the RepeatMasker was used again to find repeats in these repetitive sequence libraries. Finally, we combined all the results generated by these methods and analyzed the diversity of TEs in fish genomes.

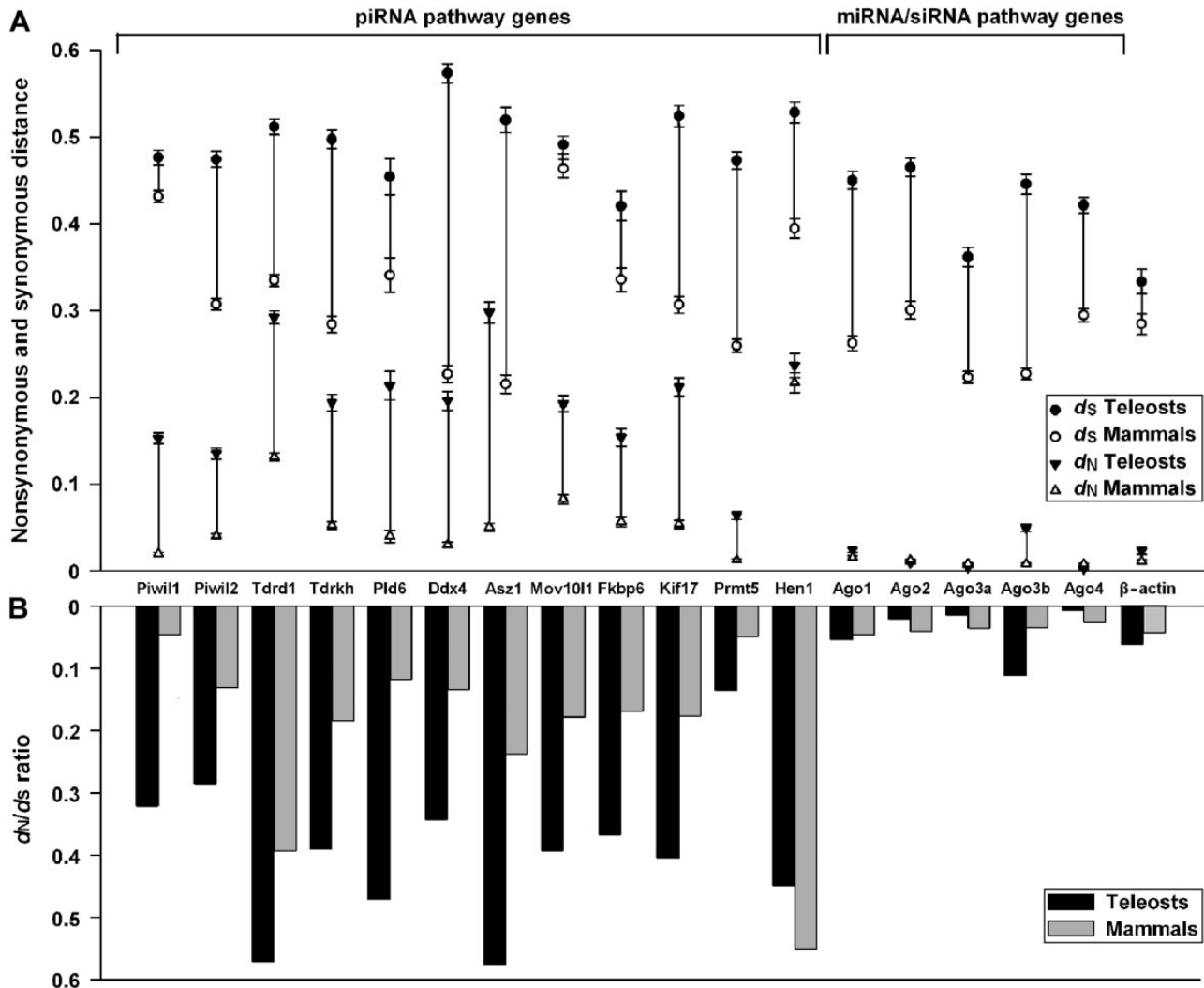


Fig. 2.—Nonsynonymous (d_N) and synonymous (d_S) nucleotide distance, and d_N/d_S ratio for 18 genes between teleosts and mammals. The first 12 genes (*Piwi1*, *Piwi2*, *Tdrd1*, *Tdrkh*, *Pld6*, *Ddx4*, *Asz1*, *Mov10l1*, *Fkbp6*, *Kif17*, *Prmt5*, and *Hen1*) are involved in piRNA pathways. Five Ago genes (*Ago1*, *Ago2*, *Ago3a*, *Ago3b*, and *Ago4*), which are involved in miRNA/siRNA pathways, and β -actin were used as control genes. (A) Nonsynonymous (d_N) and synonymous (d_S) nucleotide distance between teleosts and mammals. Straight lines between two circles (d_S) or two triangles (d_N) show distance between teleosts and mammals. All values are mean \pm SE. (B) d_N/d_S ratio between teleosts and mammals. Black and gray columns show d_N/d_S ratios of different genes in teleosts and mammals, respectively. Gene full names are as follows: *Piwi1* (Piwi-like 1), *Piwi2* (Piwi-like 2), *Tdrd1* (Tudor and KH domain containing protein), *Pld6* (Phospholipase D member 6), *Ddx4* (DEAD box polypeptide 4), *Asz1* (Ankyrin repeat, SAM and basic leucine zipper domain containing 1), *Mov10l1* (Moloney leukemia virus 10-like 1), *Fkbp6* (FK506 binding protein 6), *Kif17* (Kinesin family member 17), *Prmt5* (Protein arginine methyltransferase 5), *Hen1* (HEN1 methyltransferase homolog 1), *Ago1* (Argonaute RISC catalytic component 1), *Ago2* (Argonaute RISC catalytic component 2), *Ago3a* (Argonaute RISC catalytic component 3a), *Ago3b* (Argonaute RISC catalytic component 3b), and *Ago4* (Argonaute RISC catalytic component 4). The d_N/d_S ratios of piRNA pathway genes are significantly larger than those of miRNA/siRNA pathway genes ($P < 0.01$, t -test).

Annotation of Repeat-Associated piRNAs

The zebrafish piRNA sequences were from Ketting's study (Houwing et al. 2007) together with our data (Zhou et al. 2010). Nonredundant piRNA sequences were used to search against TE superfamilies using Repeat-Masking (Kohany et al. 2006) with default parameters in Repbase (Jurka 2000). Percentage for each TE superfamily was calculated with respect to total TE-associated piRNAs. A total of 97,666 zebrafish piRNAs were analyzed using the annotated TE database in Repbase.

Results

piRNA Pathway Genes Evolve More Rapidly in Teleosts than in Mammals

To gain insights into evolutionary dynamics of piRNA pathway in vertebrate phylogeny, we first performed phylogenetic analysis of Piwil1 protein, which is a core component of piRNA machinery in vertebrates. Comparative analysis of amino acid substitution rates among 30 vertebrate species, including 21 mammals and 9 teleost fish species, showed longer branch length of the Piwil1 in teleosts than in mammalian species (similar results in other piRNA pathway proteins were also observed), suggesting higher substitution rate in fish branches (fig. 1). To further investigate accelerated evolution of piRNA pathway in teleosts, using the modified Nei-Gojobori method (Zhang et al. 1998), we compared the synonymous and nonsynonymous nucleotide substitution distances of 12 piRNA pathway genes together with 5 miRNA/siRNA pathway genes between two groups (mammals vs. teleosts). The d_N/d_S value differences between two groups were markedly higher in piRNA genes (except for *Hen1*) than in Ago genes (fig. 2A). This suggested that piRNA pathway genes in teleosts undertook more nonsynonymous substitutions than in mammals. To analyze evolutionary rates in piRNA pathway genes between mammals and teleosts, we computed the d_N/d_S values for each gene in two groups (fig. 2B). We observed that d_N/d_S ratios of piRNA pathway genes were higher than that of Ago genes ($P < 0.01$), suggesting piRNA pathway genes evolved more rapidly than miRNA/siRNA pathway genes. In addition, the d_N/d_S ratios of 12 piRNA pathway genes except for *Hen1* in teleosts were markedly higher than that in mammals. These results suggested that evolutionary rates of piRNA pathway genes were higher in teleosts than in mammals. *Ago3b*, a teleost-specific duplicated gene of *Ago3* (McFarlane et al. 2011), is an exception. Gene duplication leads to a relaxation of selection and consequently an elevated rate of molecular evolution for the duplicated genes (Kondrashov et al. 2002). Therefore, the d_N/d_S ratio of *Ago3b* in teleosts is higher than in mammals.

To confirm these results obtained from distance analysis, we further performed branch model analysis by the codeml algorithm from the PAML 4 package (Yang 2007).

Table 1

Branch model tests for piRNA pathway genes and Ago genes

Gene	2 ΔL^a	P Value ^b	ω_0^c	ω_1^d
<i>Piwil1</i>	712.356	6.150E-157	0.022	0.165
<i>Piwil2</i>	48.940	2.639E-12	0.081	0.137
<i>Tdrd1</i>	8.499	0.003	0.253	0.300
<i>Tdrkh</i>	33.835	5.999E-09	0.126	0.210
<i>Pld6</i>	52.876	3.552E-13	0.036	0.131
<i>Ddx4</i>	76.679	2.011E-18	0.074	0.175
<i>Asz1</i>	35.661	2.348E-09	0.135	0.283
<i>Mov10l1</i>	60.794	6.336E-15	0.099	0.179
<i>Fkbp6</i>	31.633	1.862E-08	0.094	0.185
<i>Kif17</i>	28.705	8.428E-08	0.063	0.105
<i>Prmt5</i>	30.149	4.001E-08	0.031	0.059
<i>Hen1</i>	25.956	3.492E-07	0.360	0.230
<i>Ago1</i>	0.402	0.526	0.021	0.019
<i>Ago2</i>	27.078	1.954E-07	0.022	0.009
<i>Ago3a</i>	45.255	1.729E-11	0.025	0.006
<i>Ago3b</i>	9.874	0.001	0.025	0.038^e
<i>Ago4</i>	79.917	3.905E-19	0.018	0.003
<i>β-actin</i>	3.359	0.066	0.013	0.018

NOTE.—Higher ω values are in bold.

^aTwice the difference of likelihood values between two nested models (one-ratio model vs. two-ratio model).

^bP values significant at the 1% threshold after Bonferroni correction in bold.

^c ω_0 indicates a ratio of the mammalian clade (background).

^d ω_1 indicates a different ratio of the teleost fish clade (foreground).

^e*Ago3b* is a fish specific copy of *Ago3*.

Comparative analysis between the two-ratio model and one-ratio model showed that ω_1 (d_N/d_S) values for all piRNA pathway genes except *Hen1* in the teleost lineage were significantly larger than ω_0 in mammals (table 1). In contrast, ω_1 for *Ago2*, *Ago3a*, and *Ago4* in the teleost lineage were significantly smaller than ω_0 in the mammal lineages. The ω of *Ago1* and *β -actin* have no significant difference between mammals and fishes. These results from branch model analysis were consistent with synonymous and nonsynonymous distance analysis. These data suggested that piRNA pathway genes evolved more rapidly in teleosts than in mammals.

Adaptive Evolution in Functional Domains of piRNA Pathway Genes

Although ω ratios of piRNA pathway genes were smaller than 1 at whole gene level, it cannot be judged that positive selection did not occur in piRNA pathway genes. To examine the possible positive selection occurred in piRNA pathway, we used the branch-site model (Yang 2007) to test whether there was positive selection on the ancestral branch of the teleost lineages. Branch-site model test showed evidence of positive selection for most piRNA pathway genes (table 2). There were more positively selected sites in Piwi proteins and TDRDs proteins than other piRNA pathway proteins, for example, 13, 12, 9, and 7 positively selected sites with BEB PPs higher than 95% were detected in *Piwil1*, *Piwil2*, *Tdrd1*, and *Tdrkh*, respectively (table 2).

Table 2

Detection of positive selection for piRNA pathway genes and Ago genes

Gene	2ΔL ^a	P value ^b	Estimates of the Parameters in the Modified Model A ^c	Positively Selected Sites ^d
<i>Piwil1</i>	12.997	0.0003	$\omega_0 = 0.068$; $\omega_2 = 40.906$ $P_0 = 0.861$; $P_1 = 0.052$ $P_{2a} = 0.080$; $P_{2b} = 0.004$	<u>73K</u> , <u>158A</u> , <u>259S</u> , <u>285R</u> , <u>297C</u> , <u>357V</u> , <u>363V</u> , <u>448S</u> , <u>517G</u> , <u>531Y</u> , <u>535Q</u> , <u>555V</u> , <u>684L</u> , <u>692S</u> , <u>850N</u> , <u>852D</u>
<i>Piwil2</i>	17.351	3.106E-05	$\omega_0 = 0.083$; $\omega_2 = 10.057$ $P_0 = 0.826$; $P_1 = 0.076$ $P_{2a} = 0.089$; $P_{2b} = 0.008$	<u>305P</u> , <u>340T</u> , <u>345V</u> , <u>386P</u> , <u>553S</u> , <u>556G</u> , <u>602T</u> , <u>647C</u> , <u>666V</u> , <u>672S</u> , <u>692E</u> , <u>713I</u> , <u>735S</u> , <u>737P</u> , <u>804T</u> , <u>827S</u> , <u>839S</u> , <u>983Y</u>
<i>Tdrd1</i>	28.065	1.173E-07	$\omega_0 = 0.183$; $\omega_2 = 86.504$ $P_0 = 0.588$; $P_1 = 0.199$ $P_{2a} = 0.157$; $P_{2b} = 0.053$	<u>5F</u> , <u>14L</u> , <u>31P</u> , <u>91N</u> , <u>161Y</u> , <u>163R</u> , <u>352Q</u> , <u>438N</u> , <u>453R</u> , <u>474K</u> , <u>496E</u> , <u>518P</u> , <u>662E</u> , <u>776S</u> , <u>787E</u> , <u>790G</u> , <u>791Q</u> , <u>793V</u> , <u>922P</u> , <u>1000P</u> , <u>1011F</u> , <u>1054L</u>
<i>Tdrkh</i>	20.404	6.269E-06	$\omega_0 = 0.116$; $\omega_2 = 60.256$ $P_0 = 0.598$; $P_1 = 0.146$ $P_{2a} = 0.204$; $P_{2b} = 0.050$	<u>69H</u> , <u>123Q</u> , <u>126V</u> , <u>131Q</u> , <u>138L</u> , <u>142C</u> , <u>152E</u> , <u>157V</u> , <u>159D</u> , <u>177K</u> , <u>190S</u> , <u>198E</u> , <u>213Q</u> , <u>241Q</u> , <u>251P</u> , <u>348S</u> , <u>373E</u> , <u>385Y</u> , <u>387D</u> , <u>402S</u> , <u>404L</u> , <u>418L</u> , <u>453T</u> , <u>480S</u> , <u>505K</u> , <u>515S</u> , <u>542G</u>
<i>Pld6</i>	4.124	0.042	$\omega_0 = 0.065$; $\omega_2 = 6.764$ $P_0 = 0.784$; $P_1 = 0.044$ $P_{2a} = 0.161$; $P_{2b} = 0.009$	<u>59H</u> , <u>68P</u> , <u>82S</u> , <u>94S</u> , <u>120V</u> , <u>146S</u> , <u>166G</u> , <u>178L</u> , <u>179T</u> , <u>196L</u>
<i>Ddx4</i>	9.197	0.002	$\omega_0 = 0.068$; $\omega_2 = 335.073$ $P_0 = 0.766$; $P_1 = 0.177$ $P_{2a} = 0.045$; $P_{2b} = 0.010$	<u>51S</u> , <u>104F</u> , <u>177R</u> , <u>268S</u> , <u>337Q</u> , <u>357A</u>
<i>Asz1</i>	10.732	0.001	$\omega_0 = 0.155$; $\omega_2 = 40.485$ $P_0 = 0.700$; $P_1 = 0.146$ $P_{2a} = 0.126$; $P_{2b} = 0.026$	<u>154C</u> , <u>190I</u> , <u>396K</u> , <u>414Q</u> , <u>419C</u>
<i>Mov10l1</i>	17.078	3.587E-05	$\omega_0 = 0.085$; $\omega_2 = 11.126$ $P_0 = 0.694$; $P_1 = 0.130$ $P_{2a} = 0.146$; $P_{2b} = 0.027$	<u>302K</u> , <u>311S</u> , <u>341P</u> , <u>366H</u> , <u>390P</u> , <u>409V</u> , <u>438A</u> , <u>560I</u> , <u>561G</u> , <u>737Q</u> , <u>762V</u> , <u>775H</u> , <u>778F</u> , <u>779H</u> , <u>787S</u> , <u>866N</u> , <u>875G</u> , <u>886D</u> , <u>892K</u> , <u>908C</u> , <u>912P</u> , <u>913R</u> , <u>920C</u> , <u>921Q</u>
<i>Fkbp6</i>	6.917	0.009	$\omega_0 = 0.109$; $\omega_2 = 46.815$ $P_0 = 0.721$; $P_1 = 0.049$ $P_{2a} = 0.213$; $P_{2b} = 0.014$	<u>79T</u> , <u>96A</u> , <u>118Q</u> , <u>158S</u> , <u>164T</u>
<i>Kif17</i>	8.569	0.003	$\omega_0 = 0.045$; $\omega_2 = 4.658$ $P_0 = 0.798$; $P_1 = 0.099$ $P_{2a} = 0.090$; $P_{2b} = 0.011$	<u>122F</u> , <u>222S</u> , <u>226A</u> , <u>366G</u> , <u>560L</u> , <u>575S</u> , <u>581S</u> , <u>595H</u> , <u>640G</u> , <u>641Q</u> , <u>645G</u> , <u>651V</u> , <u>678T</u> , <u>684V</u> , <u>730E</u>
<i>Prmt5</i>	19.085	1.250E-05	$\omega_0 = 0.036$; $\omega_2 = 97.992$ $P_0 = 0.874$; $P_1 = 0.030$ $P_{2a} = 0.091$; $P_{2b} = 0.003$	<u>5S</u> , <u>51E</u> , <u>95E</u> , <u>110C</u> , <u>124G</u> , <u>125P</u> , <u>135L</u> , <u>187S</u> , <u>199C</u> , <u>212T</u> , <u>216K</u> , <u>269S</u> , <u>407D</u> , <u>463S</u> , <u>488K</u> , <u>492C</u> , <u>523D</u> , <u>530Q</u> , <u>531C</u> , <u>538C</u> , <u>552T</u> , <u>562K</u> , <u>587S</u> , <u>590D</u> , <u>602G</u>
<i>Hen1</i>	4.658	0.031	$\omega_0 = 0.160$; $\omega_2 = 99.653$ $P_0 = 0.602$; $P_1 = 0.351$ $P_{2a} = 0.029$; $P_{2b} = 0.016$	<u>179C</u>
<i>Ago1</i>	6.078	0.013	$\omega_0 = 0.017$; $\omega_2 = 9.523$ $P_0 = 0.979$; $P_1 = 0.008$ $P_{2a} = 0.011$; $P_{2b} = 0.0001$	<u>131R</u> , <u>716S</u>
<i>Ago2</i>	2E-06	0.999	$\omega_0 = 0.012$; $\omega_2 = 1$ $P_0 = 0.964$; $P_1 = 0.004$ $P_{2a} = 0.030$; $P_{2b} = 0.0001$	
<i>Ago3a</i>	0	1	$\omega_0 = 0.014$; $\omega_2 = 6.090$ $P_0 = 0.999$; $P_1 = 0.0003$ $P_{2a} = 0$; $P_{2b} = 0$	
<i>Ago3b</i>	0	1	$\omega_0 = 0.030$; $\omega_2 = 12.715$ $P_0 = 0.992$; $P_1 = 0.007$ $P_{2a} = 0$; $P_{2b} = 0$	
<i>Ago4</i>	0	1	$\omega_0 = 0.007$; $\omega_2 = 10.684$ $P_0 = 1$; $P_1 = 0$ $P_{2a} = 0$; $P_{2b} = 0$	
<i>β-actin</i>	4.324	0.037	$\omega_0 = 0.012$; $\omega_2 = 998.998$ $P_0 = 0.982$; $P_1 = 0.009$ $P_{2a} = 0.008$; $P_{2b} = 0.00007$	

^aThe ancestral branch leading to the teleosts was assigned as the foreground branch. Twice the difference between the likelihood values of the null model and the alternative model is indicated as 2ΔL. The modified model A is used as the alternative model and the modified model A with ω_2 fixed at 1 is the null model.

^bP values significant at the 5% threshold after Bonferroni correction in bold.

^c P_0 is the proportion of codons that have ω_0 in all branches, P_1 is the proportion of codons that have $\omega_1 = 1$ in all branches, P_{2a} is the proportion of codons that have ω_0 in the background branches but ω_2 in the foreground branches, and P_{2b} is the proportion of codons that have ω_1 in the background branches but ω_2 in the foreground branches.

^dSites with the BEB PPs higher than 90% are shown. Sites higher than 95% are single underlined and sites higher than 99% are double underlined. The sites of Mov10l1 and Fkbp6 are indexed by the amino acids at the sites in the tetraodon, whereas the sites of the other genes are indexed by the amino acids at the sites in the zebrafish.

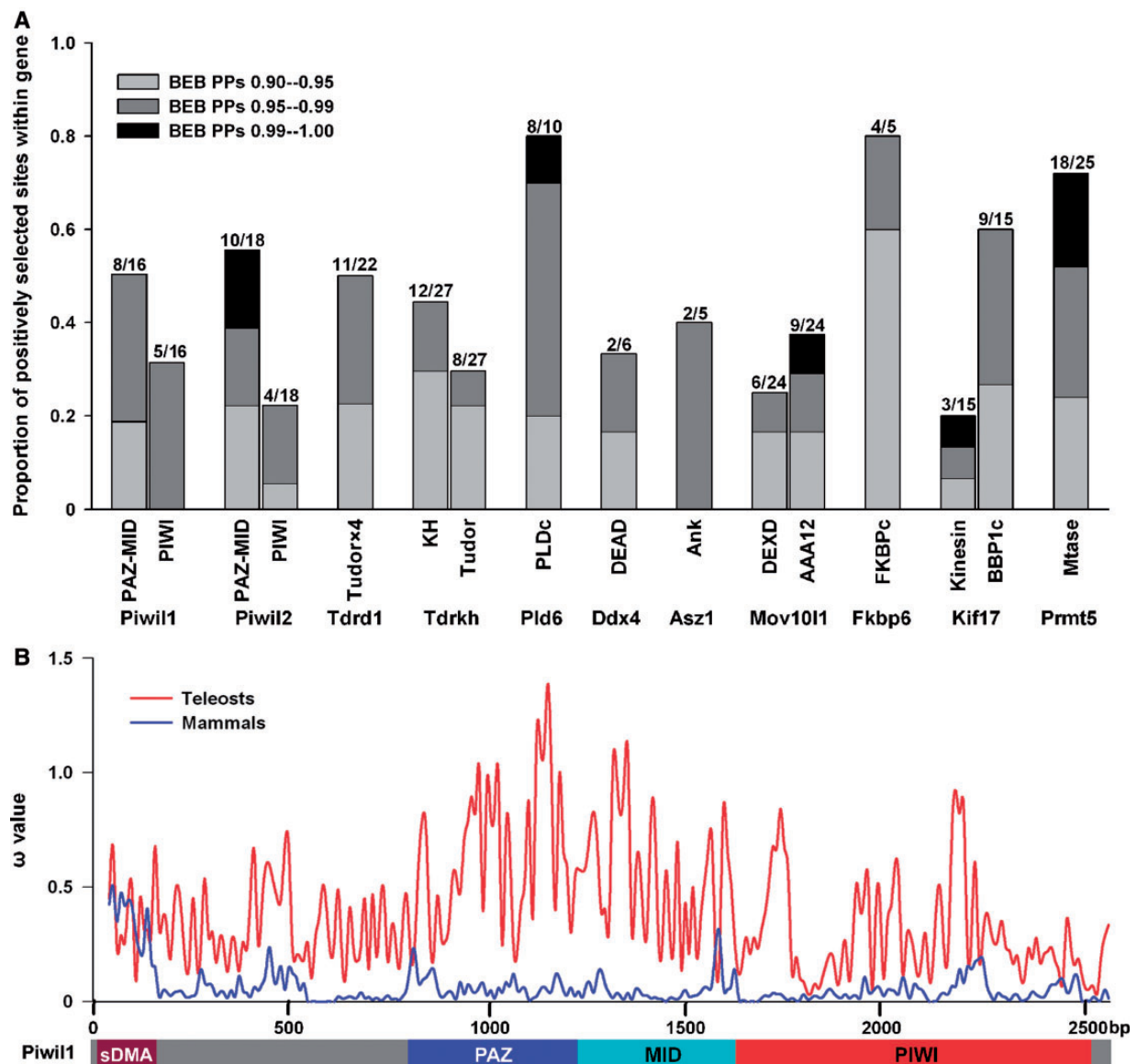


Fig. 3.—Distribution of the positively selected sites in functional domains. (A) Putatively positively selected sites with BEB PPs higher than 90% were shown. Functional domains in x axis: PAZ–MID (Piwi/Ago/Zwille–Middle), PIWI (P-element induced wimpy testis), Tudor, KH (K homology), PLDc (Phospholipase D family), DEAD (Asp–Glu–Ala–Asp box), Ank (Ankyrin repeat), DEXD (DEAD-like helicase superfamily ATP binding domain), AAA12 (ATPases Associated with a wide variety of cellular Activities superfamily), FKBPc (FKBP-type *peptidyl–prolyl cis–trans* isomerase), Kinesin (Kinesin motor domain), BBP1c (C-terminal domain of BBP1), Mtase (Arginine-*N*-methyltransferase). The numbers above stacked columns show the ratios of the sites distributed on functional domains/total positively selected sites for each gene. (B) Sliding window analysis shows ω value distribution (d_4/d_3) along the coding sequence of *Piwil1* between teleosts (red line) and mammals (blue line). Functional domains of *Piwil1*: PAZ, MID, and PIWI. sDMA represents the binding sites for Tudor domain of *Tdrd1* or *Tdrkh*.

We further analyzed locations of the positively selected sites in these genes. Using the protein topologies, we observed that their distributions were clearly concentrated on functional domains (fig. 3A). As for Piwi proteins, 50% (*Piwil1*) and 56% (*Piwil2*) positively selected sites were distributed in PAZ–MID (Piwi/Ago/Zwille–Middle) domain, which is

responsible for recognition of piRNA 3′-/5′-end (Tian et al. 2011). In addition, 31% (*Piwil1*) and 22% (*Piwil2*) positively selected sites were distributed in PIWI domain, which is similar to ribonuclease H (Song et al. 2004; Tolia and Joshua-Tor 2007). Branch-site results were further supported by sliding window analysis, which showed variation in ω values along

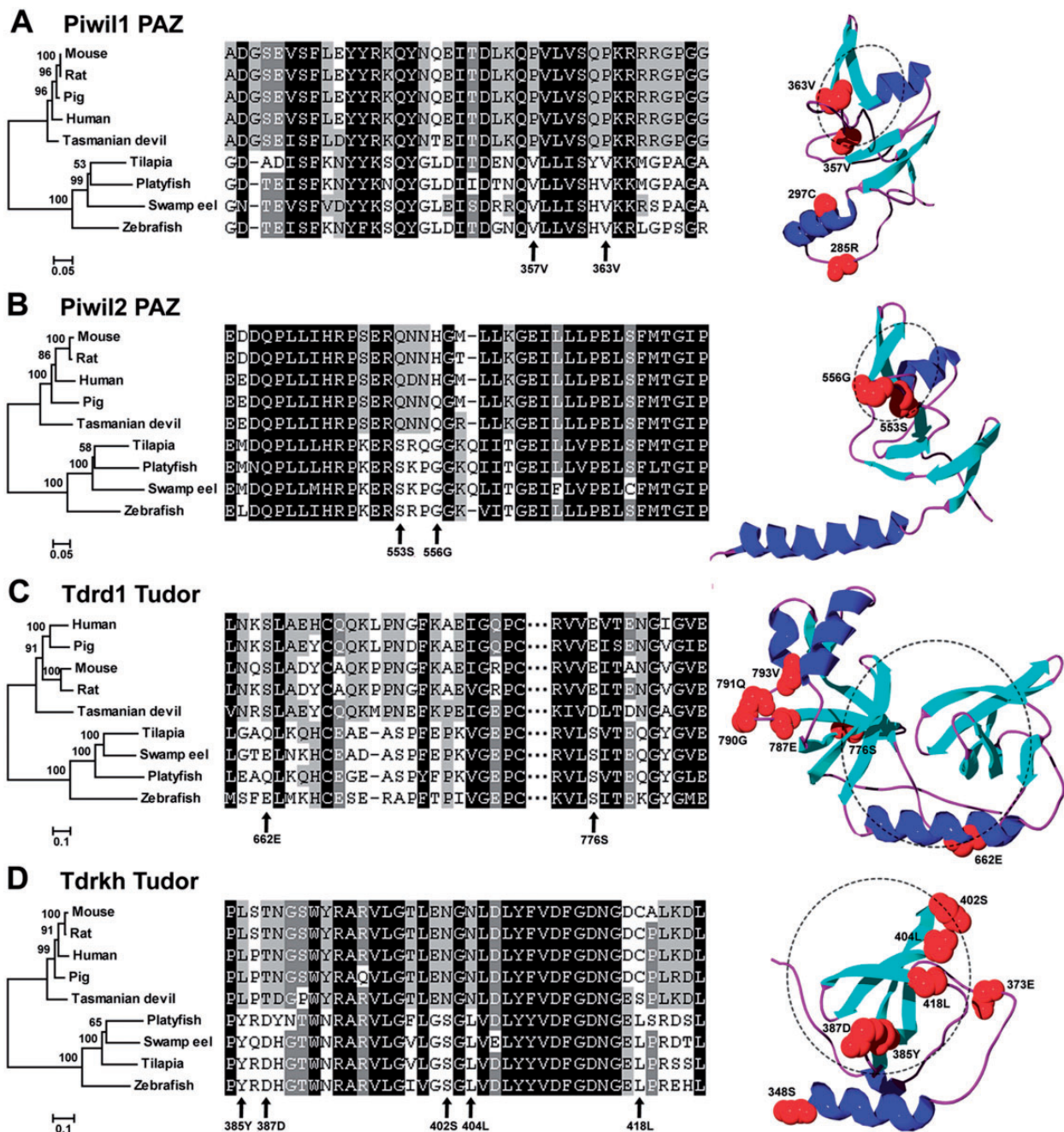


FIG. 4.—Structural modeling suggests functional importance of positively selected sites. NJ trees recapitulate vertebrate phylogeny. The amino acid alignments of nine representative species are shown in the middle panels. Homology modeling of zebrafish Piwil1 PAZ domain (from 275E to 387T) (A), zebrafish Piwil2 PAZ domain (from 459R to 574T) (B), the third Tudor domain (from 616L to 807G) of zebrafish Tdrd1 (C), and zebrafish Tdrkh Tudor domain (from 348S to 440C) (D). The binding pockets (piRNA 3'-end binds to PAZ; sDMA of Piwi binds to Tudor) are shown with dashed ellipse. Arrows (in amino acid alignments) or red spheres (in 3D models) indicate the positively selected sites. The modeling domains are colored by structural motifs: α helix, blue; β sheet, cyan; coil, purple. The modeling structures in Piwil1, Piwil2, and Tdrkh are not a whole domain because of incompleteness of corresponding template.

the *Piwil1* gene (fig. 3B). These data suggested that positive selection was strongest at the interface between Piwi proteins and target piRNA molecules, implicating their possible roles in modulating interaction of Piwi proteins with diverse piRNAs in teleosts.

As for two TDRDs proteins, 50% and 30% positively selected sites were distributed in Tudor domain of Tdrd1 and Tdrkh, respectively. Tudor domain can interact with sDMA in the N-terminal region of Piwi proteins (Chen et al. 2009; Mathioudakis et al. 2012). In addition, 44% positively selected

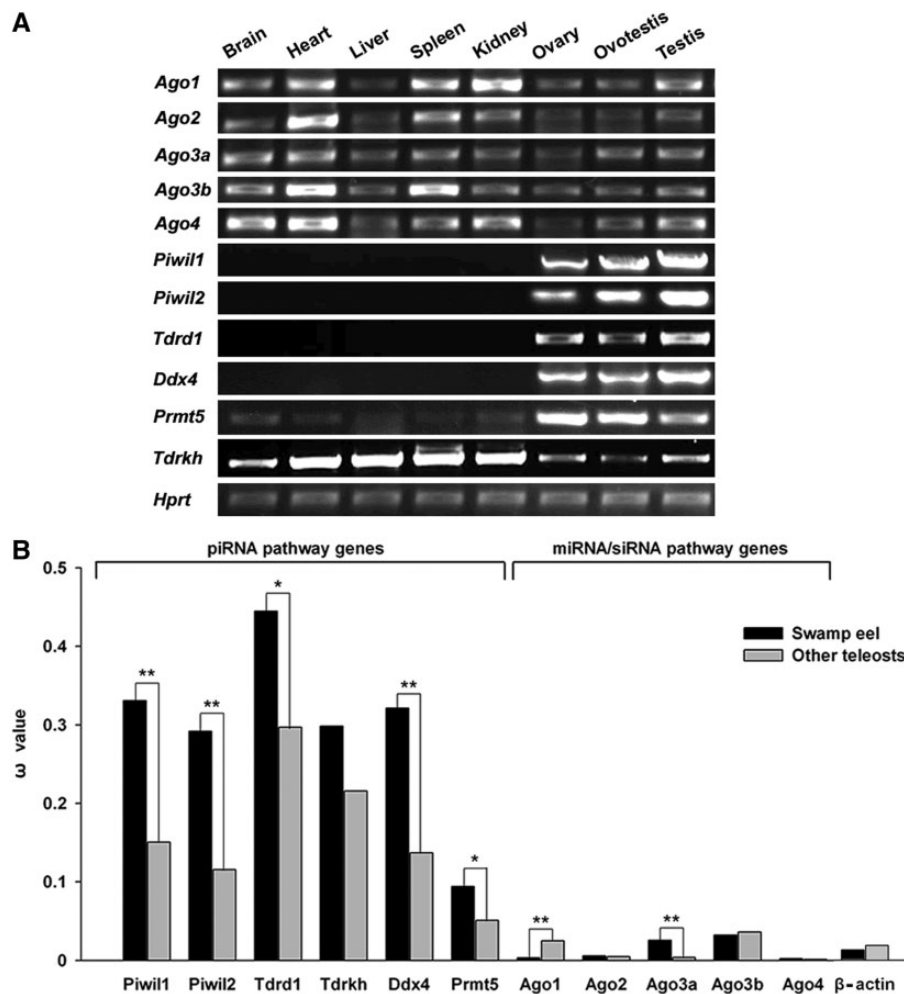


Fig. 5.—Branch model analysis of the ω values for 12 genes in the teleost lineage. (A) Cloning and expression levels of five *Ago* genes and six piRNA pathway genes in the swamp eel detected by RT-PCR. *Hprt* was used as an inner control. *Ago* genes are expressed ubiquitously, whereas the piRNA pathway genes are mainly expressed in three kinds of gonads except *Tdrkh*. (B) The branch model analysis (one-ratio model against two-ratio model). Black columns indicate ω values of the genes in the swamp eel (foreground branch), and gray columns indicate ω values of the other fish species (background branch). *LRT, $P < 0.05$; ** $P < 0.01$. *Ago* genes and β -actin were used as controls.

sites were distributed in Tdrkh-KH domain, which is an RNA-binding domain (Siomi et al. 1994; Valverde et al. 2008). The adaptive substitutions in the Tudor domain suggested their possible roles in modulating protein–protein interaction in the piRNA pathway. In contrast, LRTs were not significant and no sites have been detected in *Ago2*, *Ago3a*, *Ago3b*, and *Ago4*, whereas only two positively selected sites have been detected in *Ago1*.

3D structure homology modeling further showed that some positively selected sites in Piwil1-PAZ and Piwil2-PAZ were distributed on pocket of Piwi-PAZ for binding piRNA 3'-end, which may facilitate piRNA recognition (fig. 4A and B). Some positively selected sites in Tdrd1-third Tudor domain and Tdrkh-Tudor domain were localized in the pocket of Tudor for binding Piwi-sDMA (fig. 4C and D) to modulate

Piwi–Tudor protein interaction in the piRNA pathway. These analysis highlighted importance of amino acid changes in PAZ domain and Tudor domain, which are involved in protein–protein/RNA interaction. Overall, these results suggested that evolution of functional domains in piRNA pathway genes is likely to be adaptive to their functions in piRNA pathway in the teleost fish species.

Higher Substitution Rate in piRNA Pathway Genes in the Swamp Eel

The longest branch in the swamp eel (hermaphroditic fish) *Piwil1* as observed in the teleost lineages (fig. 1) indicated a higher substitution in swamp eel than in other fish species. To test possible accelerated evolution in this species, we first cloned five core piRNA pathway genes (*Piwil1*, *Piwil2*, *Tdrd1*,

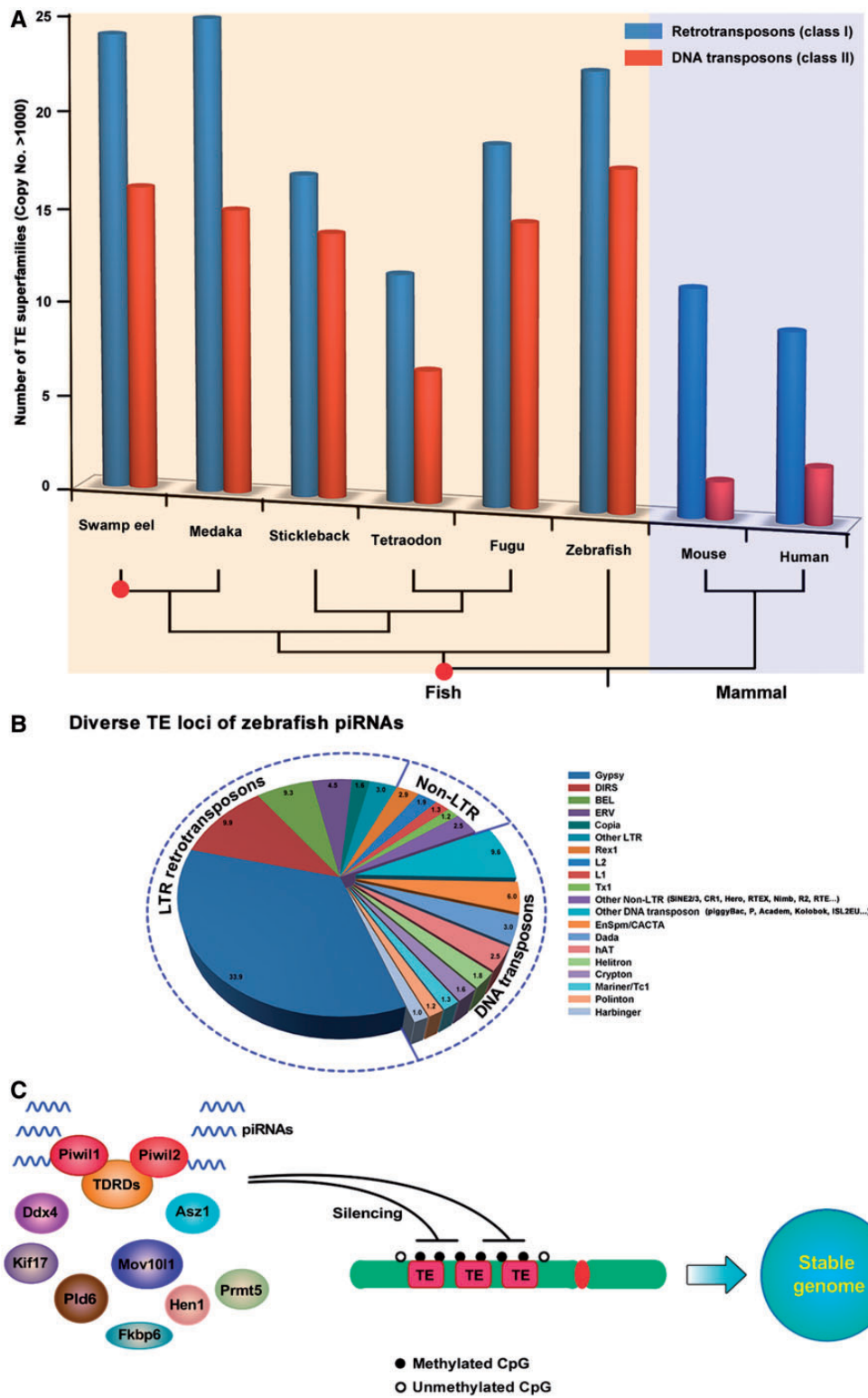


FIG. 6.—Association between rapid evolution of piRNA pathway and TEs diversity in the teleost fish species. (A) Comparison of TE superfamilies with copy number > 1,000 between teleost fish species and mammals. Blue columns, retrotransposon superfamily numbers (Class I); red columns, DNA transposon superfamily numbers (Class II). The data of human and mouse are from a previous study results (Lander et al. 2001; Mouse Genome Sequencing Consortium et al. 2002). As many as 12–25 superfamilies of retrotransposons have been observed in the teleost fish species, whereas only 10–12 superfamilies are detected in the genomes of both human and mouse. For Class II, 7–18 superfamilies are detected in the fish species, whereas there are 2–3

(continued)

Tdrkh, and *Prmt5*) and five *Ago* genes (*Ago1*, *Ago2*, *Ago3a*, *Ago3b*, and *Ago4*) from swamp eel using degenerate PCR and RACE technique. RT-PCR showed that piRNA pathway genes were mainly expressed in three kinds of gonads (ovary, ovotestis, and testis), whereas *Ago* genes were expressed in all adult tissues detected, indicating a role of piRNA pathway genes in the germline (fig. 5A). We then examined evolutionary rate of piRNA pathway genes in the species. The branch model analysis (one-ratio model vs. two-ratio model) showed that five piRNA pathway genes (*Piwil1*, *Piwil2*, *Tdrd1*, *Ddx4*, and *Prmt5*) in the swamp eel evolved more rapidly than that in other fish species significantly (fig. 5B). These results suggested that piRNA pathway genes in the swamp eel evolved at an elevated rate compared with the other gonochoristic fishes.

Diversity of TEs and TE-Derived piRNAs in Teleosts

To connect rapid evolution of piRNA pathway genes with TEs diversity in the teleost fish species, using newly released whole genome sequence data, together with new genome (swamp eel), TEs in six teleost fish species were analyzed. There were 12–25 superfamilies of retrotransposons (Class I TEs) detected in different fish species (fig. 6A), whereas only 10–12 superfamilies in human and mouse genome. For Class II, 7–18 superfamilies were detected in the fish species, whereas there were 2–3 superfamilies in both human and mouse (Lander et al. 2001; Mouse Genome Sequencing Consortium et al. 2002). Thus, the teleost fish genomes have much higher TE diversity compared with mammalian genomes, suggesting a correlation between TEs diversity and the rapid evolution of piRNA pathway. Further analysis of loci distribution of piRNAs in zebrafish genome showed that TE-derived piRNAs were corresponding to 16 superfamilies of retrotransposons and 13 superfamilies of DNA transposons (fig. 6B). However, most TE-derived piRNAs were originated from about three superfamilies of retrotransposons in mouse (Aravin et al. 2007). In addition, DNA transposons (Class II) are inactive in mammals (Huang et al. 2012). These data suggested an association between rapid evolution of piRNA pathway and TEs diversity in the teleost fish species.

Discussion

Advantageous mutation accumulation through positive selection may result in adaptive evolution. Adaptive evolution often drives rapid evolution (Swanson et al. 2001). Genes involved in some biological processes often evolve rapidly. For example, immune-related genes are generally thought as a class of rapidly evolving genes, because they are often involved in host–parasite coevolutionary arms race. In fact, many reproductive genes, such as those involved in sex determination (Whitfield et al. 1993; Zhang 2004) and gamete recognition (Swanson et al. 2001), also evolve rapidly as a result of adaptive evolution. In this report, we present that adaptive evolution may have shaped piRNA pathway in the teleost fish lineage, and the rapid evolution of piRNA pathway is consistent with the diversity of TEs, which support a correlation between piRNA pathway evolution and transposon diversity in the teleost fish. Based on these studies, a model of piRNA pathway in silencing diverse TEs in teleost fish genome is proposed (fig. 6C). Diverse piRNAs transcribed from corresponding TE loci are loaded onto Piwi proteins, which have undergone an adaptive evolution in the lineage of the teleost fish. Besides Piwi proteins, other factors, such as *Tdrd1* and *Tdrkh*, are also involved in piRNA pathway. Finally, piRNA pathway represses TE activation through DNA methylation to protect genome in germline cells.

Studies in *Drosophila* have indicated that RNAi genes evolve rapidly (Obbard et al. 2006; Kolaczkowski et al. 2011). Because RNAi processing is associated with TEs silencing, the synergetic effects between RNAi and TEs should be important in maintenance of genome stability through adaptive evolution between small RNA pathway and TEs. Our observations in the teleost fish lineages further suggest that rapid evolution of fish piRNA pathway is likely to be involved in the adaption to transposon diversity. Rapid evolution of piRNA pathway in fish may silence TEs more efficiently, which is consistent with observations in the *Drosophila* (Fablet et al. 2014). We find that positive selection has occurred in the teleost lineages with strongest at the interface between silencing machinery and target piRNAs. These results extend our understanding of piRNA pathway adaptation to ensure genome stability.

piRNA pathway genes play crucial roles in the transposon repression and the germline functions. Rapid evolution of

FIG. 6.—Continued

superfamilies in human and mouse, which are inactive in mammals. The red dots indicate the branches with rapid evolution. (B) Diverse TE loci of zebrafish piRNAs. Nonredundant piRNA sequences of zebrafish from Ketting's study (Houwing et al. 2007) together with our data (Zhou et al. 2010) were used to search against TE superfamilies using Repeat-Masking with default parameters in RepeatMasker. The number indicates percentage for each superfamily with respect to total TE-associated piRNAs. Different colors in pie diagram represent different kinds of the TE superfamily. TE superfamilies with percentages less than 1% are included in other TEs. DIRS, named for the founding member from *Dictyostelium discoideum*; ERV, endogenous retrovirus; LTR, long terminal repeat; L1, long interspersed element 1; L2, long interspersed element 2; non-LTR, nonlong terminal repeat; SINE, short interspersed elements. (C) A model of adaptive evolution of piRNA pathway to silence diverse TEs in fish genome. Diverse piRNAs transcribed from corresponding TE loci are loaded onto Piwi proteins, which have undergone an adaptive evolution in the lineage of the teleost fish. Besides Piwi proteins, other factors, such as *Tdrd1* and *Tdrkh*, are also involved in piRNA pathway. Finally, piRNA pathway represses TE activation through DNA methylation to protect genome in germline cells.

piRNA pathway genes in the teleost fish lineages is consistent with observations in the *Drosophila* species (Simkin et al. 2013), supporting an evolutionary “arms race” between TEs and the piRNA pathway. Moreover, key functional domains of Piwi proteins we detected may serve as hotspots of adaptive evolution. The adaptation at MID and PAZ domains of Pivwil1 and Pivwil2 provides access to their diverse target piRNA molecules, which is consistent with much more diversified TEs in the teleost genomes. As an epigenetic regulation, the piRNA/TE system is important for transposon silencing to ensure integrity of animal genomes. In addition, other epigenetic modes are also implicated in TE regulation, such as chromatin structure and DNA methylation (Puszyk et al. 2013). All these epigenetic regulations play crucial roles in genome defense.

A distinct biological feature that may shape susceptibility to TEs is fertilization mode (Huang et al. 2012). The mammals reproduce by internal fertilization, an adaptation to live on land rather than in water as teleosts do. In contrast, most teleosts usually reproduce by external fertilization. As both sperm and eggs are mobile in a water environment, the fertilization in water is subjected to pressure from harsh aquatic environment, climate change, and water pollution. In order to ensure sufficient fertilized eggs for species survival, external fertilization animals definitely have to shed much more sperms and eggs into water than internal fertilization animals do. Thus, gametogenesis is very active so as to maintain sustained reproductive capacity, giving rise to higher risk of transposon jumping in genomes at the same time. The piRNA pathway genes, as guardians of the genome playing crucial role in silencing transposons, are supposed to be subjected to strong and persistent selective pressure in the teleosts. A typical example of the rapid evolution in the piRNA pathway is observed in the hermaphroditic fish, swamp eel; piRNA pathway genes evolved at an elevated rate compared with the other gonochoristic fish species. In response to selective pressures from diversified TEs and reproductive mode of external fertilization, piRNA pathway genes in the teleosts are supposed to evolve faster than their orthologous genes in the mammals.

Conclusions

The piRNA pathway is responsible for germline specification, gametogenesis, transposon silencing, and genome integrity. Mobile elements can disrupt genome and its functions. piRNA pathway evolution and its adaptation to transposon diversity in the teleost fish remain unknown. In this report, we present that adaptive evolution is likely to shape piRNA pathway in the teleost fish lineages, and the rapid evolution of piRNA pathway is associated with the diversity of TEs, which suggested a link of adaptation evolution of piRNA pathway to transposon diversity in the teleost fish.

Supplementary Material

Supplementary tables S1–S4 are available at Genome Biology and Evolution online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

This work was supported by the National Natural Science Foundation of China, the National Key Basic Research project, National Key Technologies R&D Program, and Hubei Province Science and Technology project. R.Z. and H.C. designed research; M.Y., F.C., M.L., Y.C., and H.Z. performed research; M.Y. and R.Z. wrote the paper. All authors read and approved the final manuscript. The authors declare that they have no conflict of interest.

Literature Cited

- Aparicio S, et al. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297:1301–1310.
- Aravin A, et al. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442:203–207.
- Aravin AA, et al. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell*. 31: 785–799.
- 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.
- Bohne A, et al. 2012. Zisupton—a novel superfamily of DNA transposable elements recently active in fish. *Mol Biol Evol*. 29:631–645.
- Carmell MA, et al. 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell*. 12: 503–514.
- Castillo DM, Mell JC, Box KS, Blumenstiel JP. 2011. Molecular evolution under increasing transposable element burden in *Drosophila*: a speed limit on the evolutionary arms race. *BMC Evol Biol*. 11:258.
- Chen C, et al. 2009. Mouse Piwi interactome identifies binding mechanism of Tdrkh Tudor domain to arginine methylated Miwi. *Proc Natl Acad Sci U S A*. 106:20336–20341.
- Chen WJ, Mayden RL. 2010. A phylogenomic perspective on the new era of ichthyology. *Bioscience* 60:421–432.
- Cheng H, Guo Y, Yu Q, Zhou R. 2003. The rice field eel as a model system for vertebrate sexual development. *Cytogenet Genome Res*. 101: 274–277.
- Chenna R, et al. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res*. 31:3497–3500.
- Chuma S, et al. 2006. Tdrd1/Mtr-1, a tudor-related gene, is essential for male germ-cell differentiation and nuage/germinal granule formation in mice. *Proc Natl Acad Sci U S A*. 103: 15894–15899.
- Coggill P, Finn RD, Bateman A. 2008. Identifying protein domains with the Pfam database. *Curr Protoc Bioinformatics*. Chapter 2:Unit 2.5.
- Duvernell DD, Pryor SR, Adams SM. 2004. Teleost fish genomes contain a diverse array of L1 retrotransposon lineages that exhibit a low copy number and high rate of turnover. *J Mol Evol*. 59:298–308.
- Eickbush TH, Furano AV. 2002. Fruit flies and humans respond differently to retrotransposons. *Curr Opin Genet Dev*. 12:669–674.
- Fablet M, Akkouche A, Braman V, Vieira C. 2014. Variable expression levels detected in the *Drosophila* effectors of piRNA biogenesis. *Gene* 537:149–153.
- Frost RJ, et al. 2010. MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs. *Proc Natl Acad Sci U S A*. 107:11847–11852.

- Furano AV, Duvernell DD, Boissinot S. 2004. L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends Genet.* 20:9–14.
- Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442:199–202.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:696–704.
- Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59:307–321.
- Houwing S, Berezikov E, Ketting RF. 2008. Zili is required for germ cell differentiation and meiosis in zebrafish. *EMBO J.* 27:2702–2711.
- Houwing S, et al. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* 129:69–82.
- Huang CR, Burns KH, Boeke JD. 2012. Active transposition in genomes. *Annu Rev Genet.* 46:651–675.
- Huang H, Gao Q, et al. 2011. piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. *Dev Cell.* 20:376–387.
- Huang HY, Houwing S, et al. 2011. Tdrd1 acts as a molecular scaffold for Piwi proteins and piRNA targets in zebrafish. *EMBO J.* 30:3298–3308.
- Ishizu H, Siomi H, Siomi MC. 2012. Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. *Genes Dev.* 26:2361–2373.
- Jurka J. 2000. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.* 16:418–420.
- Kalmykova AI, Klenov MS, Gvozdev VA. 2005. Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. *Nucleic Acids Res.* 33:2052–2059.
- Kamminga LM, et al. 2010. Hen1 is required for oocyte development and piRNA stability in zebrafish. *EMBO J.* 29:3688–3700.
- Kirino Y, Mourelatos Z. 2007. The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *RNA* 13:1397–1401.
- Kirino Y, et al. 2009. Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. *Nat Cell Biol.* 11:652–658.
- Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 7:474.
- Kolaczowski B, Hupalo DN, Kern AD. 2011. Recurrent adaptation in RNA interference genes across the *Drosophila* phylogeny. *Mol Biol Evol.* 28:1033–1042.
- Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. 2002. Selection in the evolution of gene duplications. *Genome Biol.* 3:RESEARCH0008.
- Kotaja N, Lin H, Parvinen M, Sassone-Corsi P. 2006. Interplay of PIWI/Argonaute protein MIWI and kinesin KIF17b in chromatoid bodies of male germ cells. *J Cell Sci.* 119:2819–2825.
- Kuramochi-Miyagawa S, et al. 2010. MVH in piRNA processing and gene silencing of retrotransposons. *Genes Dev.* 24:887–892.
- Lander ES, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
- Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol.* 25:1307–1320.
- Ma L, et al. 2009. GASZ is essential for male meiosis and suppression of retrotransposon expression in the male germline. *PLoS Genet.* 5:e1000635.
- Malone CD, Hannon GJ. 2009. Molecular evolution of piRNA and transposon control pathways in *Drosophila*. *Cold Spring Harb Symp Quant Biol.* 74:225–234.
- Mathioudakis N, et al. 2012. The multiple Tudor domain-containing protein TDRD1 is a molecular scaffold for mouse Piwi proteins and piRNA biogenesis factors. *RNA* 18:2056–2072.
- McFarlane L, et al. 2011. Expansion of the Ago gene family in the teleost clade. *Dev Genes Evol.* 221:95–104.
- Mouse Genome Sequencing Consortium, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562.
- Obbard DJ, Jiggins FM, Halligan DL, Little TJ. 2006. Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr Biol.* 16:580–585.
- Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. *Bioinformatics* 21(Suppl. 1):i351–i358.
- Pride DT. 2000. SWAAP: a tool for analyzing substitutions and similarity in multiple alignments. Available at: <http://bioinf.man.ac.uk/robertson/recombination/programs.shtml>.
- Punta M, et al. 2012. The Pfam protein families database. *Nucleic Acids Res.* 40:D290–D301.
- Puszyk W, et al. 2013. The epigenetic regulator PLZF represses L1 retrotransposition in germ and progenitor cells. *EMBO J.* 32:1941–1952.
- Reuter M, et al. 2009. Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nat Struct Mol Biol.* 16:639–646.
- Schartl M, et al. 2013. The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat Genet.* 45:567–72.
- Schwede T, Kopp J, Guex N, Peitsch MC. 2003. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* 31:3381–3385.
- Simkin A, Wong A, Poh YP, Theurkauf WE, Jensen JD. 2013. Recurrent and recent selective sweeps in the piRNA pathway. *Evolution* 67:1081–1090.
- Siomi H, Choi M, Siomi MC, Nussbaum RL, Dreyfuss G. 1994. Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. *Cell* 77:33–39.
- Song JJ, Smith SK, Hannon GJ, Joshua-Tor L. 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305:1434–1437.
- Swanson WJ, Yang Z, Wolfner MF, Aquadro CF. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc Natl Acad Sci U S A.* 98:2509–2514.
- Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.
- Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics.* Chapter 4:Unit 4.10.
- Tian Y, Simanshu DK, Ma JB, Patel DJ. 2011. Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. *Proc Natl Acad Sci U S A.* 108:903–910.
- Tolia NH, Joshua-Tor L. 2007. Slicer and the argonautes. *Nat Chem Biol.* 3:36–43.
- Vagin VV, et al. 2009. Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev.* 23:1749–1762.
- Valverde R, Edwards L, Regan L. 2008. Structure and function of KH domains. *FEBS J.* 275:2712–2726.
- Vermaak D, Henikoff S, Malik HS. 2005. Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in *Drosophila*. *PLoS Genet.* 1:96–108.
- Vollf JN. 2005. Genome evolution and biodiversity in teleost fish. *Heredity (Edinb)* 94:280–294.
- Vollf JN, Bouneau L, Ozouf-Costaz C, Fischer C. 2003. Diversity of retrotransposable elements in compact pufferfish genomes. *Trends Genet.* 19:674–678.
- Watanabe T, et al. 2011. MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. *Dev Cell.* 20:364–375.

- Whitfield LS, Lovell-Badge R, Goodfellow PN. 1993. Rapid sequence evolution of the mammalian sex-determining gene SRY. *Nature* 364: 713–715.
- Xiol J, et al. 2012. A role for Fkbp6 and the chaperone machinery in piRNA amplification and transposon silencing. *Mol Cell*. 47:970–979.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.
- Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol*. 22: 1107–1118.
- Zhang J. 2004. Evolution of DMY, a newly emergent male sex-determination gene of medaka fish. *Genetics* 166:1887–1895.
- Zhang J, Rosenberg HF, Nei M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A*. 95:3708–3713.
- Zheng K, et al. 2010. Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. *Proc Natl Acad Sci U S A*. 107:11841–11846.
- Zhou RJ, Cheng HH, Tiersch TR. 2001. Differential genome duplication and fish diversity. *Rev Fish Biol Fish*. 11:331–337.
- Zhou X, et al. 2010. Profiling sex-specific piRNAs in zebrafish. *Genetics* 186:1175–1185.

Associate editor: Ellen Pritham