ELSEVIER

Article

Characterization of Evolutionarily Conserved MicroRNAs in Amphioxus

Lei Wang[1,2#], Lan Jiang[1,2#], Songnian Hu[1*], and Yejun Wang[3*]

[1]*Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, China;*
[2]*Graduate University of Chinese Academy of Sciences, Beijing 100049, China;*
[3]*Department of Biology, the Chinese University of Hong Kong, Hongkong, China.*

Abstract

Amphioxus is an extant species closest to the ancestry of vertebrates. Observation of microRNA (miRNA) distribution of amphioxus would lend some hints for evolutionary research of vertebrates. In this study, using the publicly available scaffold data of the Florida amphioxus (*Branchiostoma floridae*) genome, we screened and characterized homologs of miRNAs that had been identified in other species. In total, 68 pieces of such homologs were obtained and classified into 33 families. Most of these miRNAs were distributed as clusters in genome. Inter-species comparison showed that many miRNAs, which had been thought as vertebrate- or mammal-specific before, were also present in amphioxus, while some miRNAs that had been considered as protostome-specific before also existed in amphioxus. Compared with ciona, amphioxus had an apparent miRNA gene expansion, but phylogenetic analysis showed that the duplicated miRNAs or clusters of amphioxus had a higher homology level than those duplicated ones in vertebrates.

Key words: amphioxus, microRNA, genome evolution

Introduction

MicroRNAs (miRNAs) are evolutionarily conserved small non-coding RNA gene products that regulate gene expression at the post-transcriptional level (*1, 2*). miRNAs can be transcribed as exons of non-coding transcripts, or located in introns of protein-coding or non-coding transcripts (*3, 4*). A substantial fraction of known miRNAs (about 50%) are clustered and found in close proximity to other miRNAs (*5-8*). The clus-

tered miRNAs usually have similar or related functions and share the same promoters and other transcription factors. There are some differences between animal and plant miRNAs. Animal miRNA genes are transcribed from genome as pri-miRNAs with about several kilobases (*4, 9, 10*). After processed by Drosha, an RNase III, hairpins about 70 nt named pre-miRNAs are generated (*2, 11*). Pre-miRNAs are then transported into cytoplasm and processed by another RNase, Dicer. The loops in hairpins are cut down and the double-string stems are incorporated into RNA-induced silencing complex. Then the double-string stems are separated, and for each stem, one string is degenerated, leaving the other one, mature miRNA, to function by complementarily combining

with 3'-UTR of target mRNAs (*2, 11*).

Amphioxus is an extant species closest to the ancestry of vertebrates (*12-17*). Observation of miRNA distribution of amphioxus would lend some hints for evolutionary research of vertebrates. A limited part of amphioxus miRNAs have been studied in literature, which particularly cared about their phylogenetic relationship with other species (*18, 19*). Up to now, however, no report has been available that systematically observed and analyzed whole genomic miRNAs in amphioxus. Apparently, this systematic miRNA study for amphioxus and other variety of species would be quite meaningful and informative.

In this study, we used a homology-miRNA prediction pipeline to find all homologs of known miRNAs in *Branchiostoma floridae*, an amphioxus species whose genome had been sequenced (*16*). Furthermore, a careful analysis of these homologous miRNAs was

conducted, including their distribution in amphioxus genome, cross-species conservation and phylogenetic comparison.

Results

miRNA homologs in amphioxus

With a strict and careful homology-searching pipeline (see Materials and Methods), we obtained 56 pieces of potential miRNAs and 12 possible miRNAs from the genome of *B. floridae* (**Table S1**). For further analysis, the potential and possible miRNAs were merged. According to the clustering strategy established in miRBase, the miRNAs were classified into 33 families based on sequence similarity. The family distribution is shown in **Table 1**.

Table 1   miRNA family distribution in amphioxus

No.	Family	miRNA	Copy No.	Strand	Possibility
1	MIPF0000002	let-7	4	4p	potential
2	MIPF0000009	mir-29	2	2p	potential
3	MIPF0000013	mir-25	7	7m	potential
4	MIPF0000014	mir-9	2	1p1m	potential
5	MIPF0000017	mir-125	2	2p	potential
6	MIPF0000019	mir-8	3	3m	potential
7	MIPF0000021	mir-124	1	1p	potential
8	MIPF0000022	mir-7	2	1p1m	potential
9	MIPF0000025	mir-99	2	2p	potential
10	MIPF0000028	mir-135	1	1m	possible
11	MIPF0000029	mir-133	1	1p	possible
12	MIPF0000033	mir-10	3	3p	potential
13	MIPF0000038	mir-1	2	2p	potential
14	MIPF0000044	mir-219	2	2m	potential
15	MIPF0000050	mir-153	1	1p	potential
16	MIPF0000054	mir-216	3	1p2m	potential
17	MIPF0000056	mir-148	1	1m	possible
18	MIPF0000059	mir-184	1	1p	potential
19	MIPF0000064	mir-31	1	1m	possible
20	MIPF0000066	mir-183	2	2m	potential
21	MIPF0000070	mir-33	4	4p	potential
22	MIPF0000072	mir-96	2	2m	potential
23	MIPF0000076	mir-190	2	2m	potential
24	MIPF0000077	mir-217	3	1p2m	potential
25	MIPF0000086	mir-210	2	1p1m	potential
26	MIPF00000114	mir-375	1	1p	potential
27	MIPF00000140	mir-3	1	1p	possible
28	MIPF00000173	mir-499	2	1p1m	possible
29	MIPF00000200	mir-330	2	2p	possible
30	MIPF00000269	mir-341	1	1m	possible
31	MIPF00000278	mir-71	1	1m	potential
32	MIPF00000285	mir-252	2	2m	possible
33	MIPF00000316	mir-467	2	1p1m	potential

Cluster distribution of amphioxus miRNA homologs

The miRNA homologs of amphioxus have clustering features, which are consistent with those in other species (**Figure 1**). Distribution information was characterized in detail as follows:
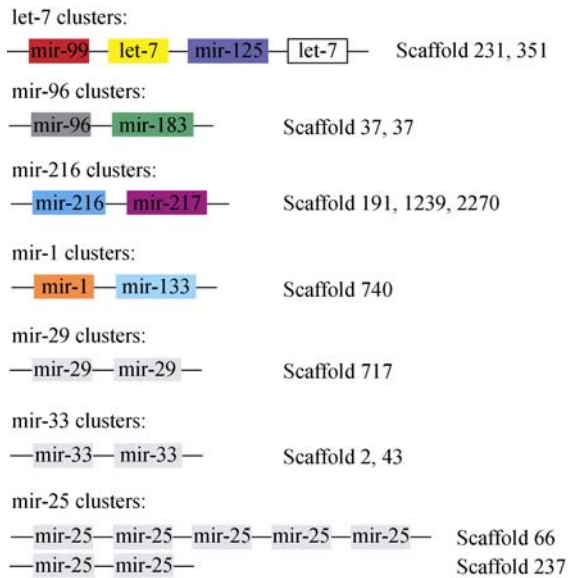

Figure 1 miRNA clusters in amphioxus. For each cluster, miRNA gene order is shown in the left; the scaffolds in which the clusters are located are shown in the right.

let-7 cluster

This cluster is conserved in most species with identified miRNAs. Classically, this cluster has one paralogue of each mir-99, let-7 and mir-125 family, and the miRNA order is constant among variety of species. This cluster is also conserved in amphioxus, and amphioxus seems to have another let-7 nearby the 3'-end of mir-125. Amphioxus has two copies of this let-7 cluster, each with an identical mir-99—let-7—mir-125—let-7 distribution within a span of 2,500 nt. One cluster is located in Scaffold 231, while the other one in Scaffold 351.

mir-96 cluster

The paralogs of mir-96 and mir-183 are often co-transcribed with mir-182 in a cluster. There are two mir-96—mir-183 clusters in amphioxus, both in the minus strand of Scaffold 37. Two clusters have a constant gene order and the range between mir-96 and

mir-183 is less than 1,100 nt. Mir-182 is absent in this cluster in amphioxus.

mir-216 cluster

Both mir-216 and mir-217 arise three times in amphioxus genome, and they are always found in the same genomic region within a range of less than 1,000 nt. The order of the miRNAs in each cluster is consistent.

mir-33 cluster

Though totally four copies of mir-33 genes are found in *B. floridae* genome, they could be divided into two clusters. Each cluster contains two mir-33 genes and is distributed in a single genomic location. In each cluster, the miRNA genes are in a constant order and within a span of less than 500 nt.

mir-25 cluster

Amphioxus has an unexpected seven copies of mir-25-family miRNAs. These seven copies could be divided into two clusters. One cluster containing five copies is located in the minus stand of Scaffold 66, with each copy arraying closely. The other two copies in another cluster are located in the minus strand of Scaffold 237.

mir-29 cluster

This cluster often contains two paralogs of mir-29 family. Amphioxus also contains one copy of mir-29 cluster located in the plus strand of Scaffold 717. Two mir-29 paralogs are distributed in a region less than 1,000 nt.

mir-1 cluster

Mir-1 or mir-206 is often transcribed with mir-133 to form a cluster of mir-1—mir-133 or mir-206—mir-133 in vertebrates. In amphioxus, there is one copy of mir-1 cluster composed of one mir-1 and one mir-133 located within a genomic region of less than 800 nt in the positive strand of Scaffold 740. Mir-1 has another copy in Scaffold 71, but there is not another mir-133 to form a second mir-1—mir-133 cluster. Moreover, no mir-206 homologs are found in amphioxus.

mir-8 cluster

In vertebrates, mir-8 has a lot of paralogs and these

paralogs often lie in the same cluster. In amphioxus, there are three paralogs of mir-8 family, and all these three paralogs are located in the minus strand of Scaffold 96 within a limited span.

Others

Mir-9, mir-7, mir-219, mir-210, mir-467 and mir-190 each has two separated copies. Mir-10 has three separated copies while each of the rest miRNAs only has one copy.

Comparative analysis of homologous miRNAs of amphioxus and ciona

It is interesting to observe conserved and non-conserved miRNA homologs between Cephalochordata and Urochordata, which would be informative for observation of miRNA evolution, and also helpful for analysis of species-shared or species-specific miRNAs. Conveniently, the genome of *Ciona intestinalis* has been sequenced and the data is available (*20*). Therefore, the miRNA homologs of ciona were screened using the same pipeline. In all, 19 miRNAs classified as 15 families were obtained (**Table 2**), including the 4 inter-species conserved ones that were also reported in literature (*18, 21*). Except that mir-1 and mir-133 were closely located in the plus strand of Scaffold 844 within a span of less than 300 nt, no other apparent cluster distribution feature was observed for the miRNAs of ciona.

Table 2   miRNA family distribution in ciona

No.	Family	miRNA	Copy No.	Strand
1	MIPF0000002	let-7	1	1m
2	MIPF0000013	mir-25	2	2p
3	MIPF0000019	mir-8	2	2p
4	MIPF0000021	mir-124	2	2m
5	MIPF0000022	mir-7	1	1p
6	MIPF0000029	mir-133	1	1p
7	MIPF0000038	mir-1	1	1p
8	MIPF0000044	mir-219	1	1m
9	MIPF0000050	mir-153	1	1p
10	MIPF0000059	mir-184	1	1m
11	MIPF0000064	mir-31	1	1m
12	MIPF0000066	mir-183	1	1m
13	MIPF0000070	mir-33	1	1m
14	MIPF0000204	mir-297	2	2p
15	MIPF0000277	mir-72	1	1m

After comparison of the miRNA homologs in amphioxus and ciona, we found that most of ciona miRNAs were conserved in both species, whereas many miRNAs only present in amphioxus but not in ciona. Only two families, MIPF0000204 and MIPF0000277, were found in ciona but absent in amphioxus.

Cross-species distribution of homologous miRNAs

According to the inter-species distribution, known homologous miRNAs are categorized as inter-group common or group/species-specific ones (**Table S2**). After introduction of homologs of amphioxus and ciona, some miRNAs that had been recognized as vertebrate-, protostome- or subgroup-specific before should be re-examined and may be classified into other categories precisely (**Table 3**). The inter-species miRNA distribution could be summarized as follows:

Bilateria-common miRNAs

Table 3 shows that among the homologous miRNAs, there are 17 miRNA families conserved in multiple species and all the 17 families exist in the latest common ancestor of protostomes and deuterostomes. Only one family (MIPF0000039) has not been found in amphioxus or ciona, nor in *Fugu rubripes*. Existence of miRNAs belonging to the other families in amphioxus and/or ciona corroborates that these families have been existing at least in the latest bilateria species before diverging into protostomes and deuterostomes (Table 3). Another family, MIPF0000070, was lost or has not been identified in nematodes and pisces, but both amphioxus and ciona contain this family, which further supports that this family of miRNAs are bilateria-common ones.

There are also several families that could have been recognized as subgroup-specific but recognized as bilateria-common ones due to existence in amphioxus and/or ciona (Table 3). MIPF0000140 was thought as drosophila- or arthropoda-specific, because no other species up to now has been identified with this family of miRNAs. There is one such family, however, found in amphioxus, and therefore, this family should have been classified as bilateria-common, and it maybe has been lost or could not be recognized unambiguously

in many latest extant species due to mutations. Same cases are MIPF0000277 and MIPF0000285, which should not have been thought as nematode-specific, but bilateria-common. Also, MIPF0000278 should have existed in the latest common ancestor of protostomes and deuterostomes rather than only arose in ancient protostome species.

Table 3 Group-specificity of homologous miRNA families re-examined including amphioxus and/or ciona miRNAs

Family	Bilateria-common	Vertebrate/mammal-specific	Protosome-specific
MIPF0000002	√		
MIPF0000009	√		
MIPF0000013	√		
MIPF0000014	√		
MIPF0000017	√		
MIPF0000019	√		
MIPF0000021	√		
MIPF0000022	√		
MIPF0000025	√		
MIPF0000028		√	
MIPF0000029	√		
MIPF0000033	√		
MIPF0000038	√		
MIPF0000044	√		
MIPF0000050		√	
MIPF0000054		√	
MIPF0000056		√	
MIPF0000059	√		
MIPF0000064	√		
MIPF0000066		√	
MIPF0000070	√		
MIPF0000072		√	
MIPF0000076		√	
MIPF0000077		√	
MIPF0000086	√		
MIPF0000114		√	
MIPF0000140			√
MIPF0000173		√	
MIPF0000200		√	
MIPF0000204		√	
MIPF0000269		√	
MIPF0000277			√
MIPF0000278			√
MIPF0000285			√
MIPF0000316		√	

Chordate-specific miRNAs

Knowledge about miRNAs in cephalochordates and urochordates is helpful for discriminating whether one miRNA family is vertebrate-specific or chordate-specific. Before this report, several miRNA families could only be recognized as vertebrate-specific because no conserved ones were identified outside vertebrates (Table 3 and Table S2). However, we have found that there are homologs in amphioxus and/or ciona, so they are further classified as chordate-specific. In other words, they were existent in the chordate ancestor.

Other group-specific miRNAs

No other families of miRNAs were observed to be conserved in amphioxus or ciona, and therefore for these miRNAs, the phylogenetic relationship was not re-examined (Table S2). When more species, especially those of hemichordates, echinodermates, molluscs and others, are analyzed for their miRNAs, the relationship would be finally determined.

Expansion of miRNA genes

miRNAs represent a group of conserved genes that date back to ancient species, thus they are preferable for phylogenetic analysis of proposed ancient genome duplication. By comparison of miRNA family type and copy number between amphioxus and ciona, two species whose genomes do not experience hypothesized 2R whole genome duplication, we found that amphioxus has an apparent miRNA gene expansion from miRNA types to miRNA copies of each family (**Table 4**). There are 19 miRNAs belonging to 15 families in ciona while 68 miRNAs belonging to 33 families in amphioxus. The average gene copies per family approximate to 1.3 in ciona and in amphioxus it is over 2.0. In fact, this gene expansion in amphioxus is somewhat beyond our expectation, because as in ciona, most genes examined in amphioxus contain only single copy.

Besides ciona, the miRNA homologs of amphioxus were further compared with other species. **Table S3** summarizes the copies of each miRNA family presenting in amphioxus or ciona compared with those in representative species. Complete miRNA copy distribution among species is demonstrated in **Table S4**.

Table 4   miRNA expansion among species

Expansion of miRNAs in amphioxus compared with ciona		
Copy number per family	Family number in ciona	Family number in amphioxus
1	11	10
2	4	17
3	0	3
4	0	2
>4	0	1
family sum	15	33
miRNA sum	19	68
average copy	1.3	2.1

Expansion of miRNAs in fish		
Copy number per family	Family number in amphioxus	Family number in zebrafish
1	8	6
2	13	4
3	3	2
4	1	6
>4	1	8
family sum	26	26
miRNA sum	54	102
average copy	2.1	3.9

Expansion of miRNAs in mammals			
Copy number per family	Family number in amphioxus	Family number in mouse	Family number in human
1	8	11	9
2	13	4	6
3	3	9	9
4	2	2	2
>4	1	2	2
family sum	27	28	28
miRNA sum	58	71	73
average copy	2.1	2.5	2.6

miRNAs in zebrafish are observed to be expanded when compared to those of amphioxus (Table 4). For the 26 conserved families between amphioxus and zebrafish, zebrafish has more multiple (four or more than four)-copy families (zebrafish vs. amphioxus: 14 vs. 2) and fewer two-copy families (zebrafish vs. amphioxus: 4 vs. 13) and similar single-copy families (zebrafish vs. amphioxus: 6 vs. 8). Generally, the average-copy (copy number per family) ratio of zebrafish to amphioxus is about 4:2. We further observed the difference of miRNA copies among amphioxus, mouse and human (Table 4). For human and mouse, however, they do not have a higher or similar average-copy ratio to amphioxus than zebrafish. To be contrast, the average-copy ratio of human or mouse is close to that of amphioxus. The copy number distribution is also similar among the three species (Table 4). The results demonstrate that amphioxus may have an miRNA gene expansion, but this might be due to a local and species-specific gene/cluster duplication rather than whole genome duplication happening at the divergence of vertebrates. Species-specific and local gene duplication may also have contributed to the abnormal increase of average copy in fish in terms of the miRNAs examined here.

To confirm this species-specific miRNA gene expansion, we took several individual miRNAs or miRNA clusters for phylogenetic analysis. Amphioxus has two copies of let-7 clusters composed of mir-99, let-7 and mir-125 family. In each cluster, there

are two let-7 homologs close to each other and separated by a mir-125 homolog. We collected corresponding genes of all conserved let-7 clusters containing consistent three miRNAs from other species and constructed a phylogenetic tree for each gene (**Figures 2, 3 and 4A**), and then used the combination of three miRNA hairpin sequences for tree construction (**Figure 4B**). Though it could not reflect all the phylogenetic relationship among species and some differences exist among trees based on different miRNA genes, the overall relationship of clusters is similar. Taking let-7 gene tree for example, as shown in Figure 2, the clusters of one species in different loci can be distinguished at different branch. For example, human chromosome-21 cluster is closer to that of rat chromosome 11 and chicken chromosome 1, human chromosome-11 cluster is closer to that of rat chromosome 8 and ze-

brafish chromosome 5, and human chromosome-19 let-7 is in the same branch with that of rat chromosome 1. In all the individual-gene trees or combined-cluster tree, the two let-7 clusters of amphioxus are always in the same branch, with nearly the smallest branching distances. That is to say, the divergence of the two clusters happened very late after the divergence of clusters among vertebrates, and therefore it is potentially species-specific. Furthermore, we added the rest two let-7 hairpin sequences and reconstructed the phylogenetic relationship among them and other let-7 cluster-related let-7 hairpins (Figure 3). It appeared that let-7 cluster had a local let-7 gene duplication and then a whole-cluster duplication, since the phylogenetic distance between the two let-7 genes in one cluster is larger than that between the two genes in the same order but different genomic regions.
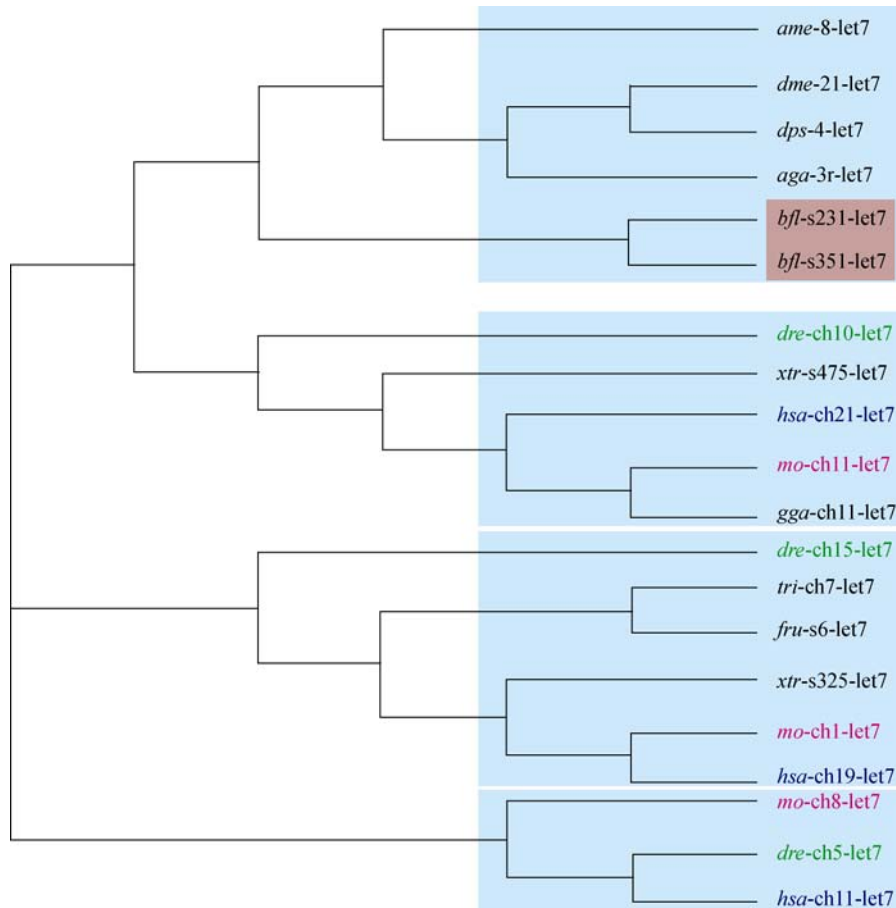

Figure 2 Neighbor-joining tree of let-7 hairpins within conserved let-7 clusters among species. The let-7 miRNAs of *B. floridae* are highlighted in red. miRNAs of human, rat and fish are noted in dark blue, purple and green, respectively. Abbreviations: ame, *Apis mellifera*; dme, *Drosophila melanogaster*; dps, *Drosophila pseudoobscura*; aga, *Anopheles gambiae*; bfl, *Branchiostoma floridae*; dre, *Danio rerio*; xtr, *Xenopus tropicalis*; hsa, *Homo sapiens*; rno, *Rattus norvegicus*; gga, *Gallus gallus*; tni, *Tetraodon nigroviridis*; fru, *Fugu rubripes*.
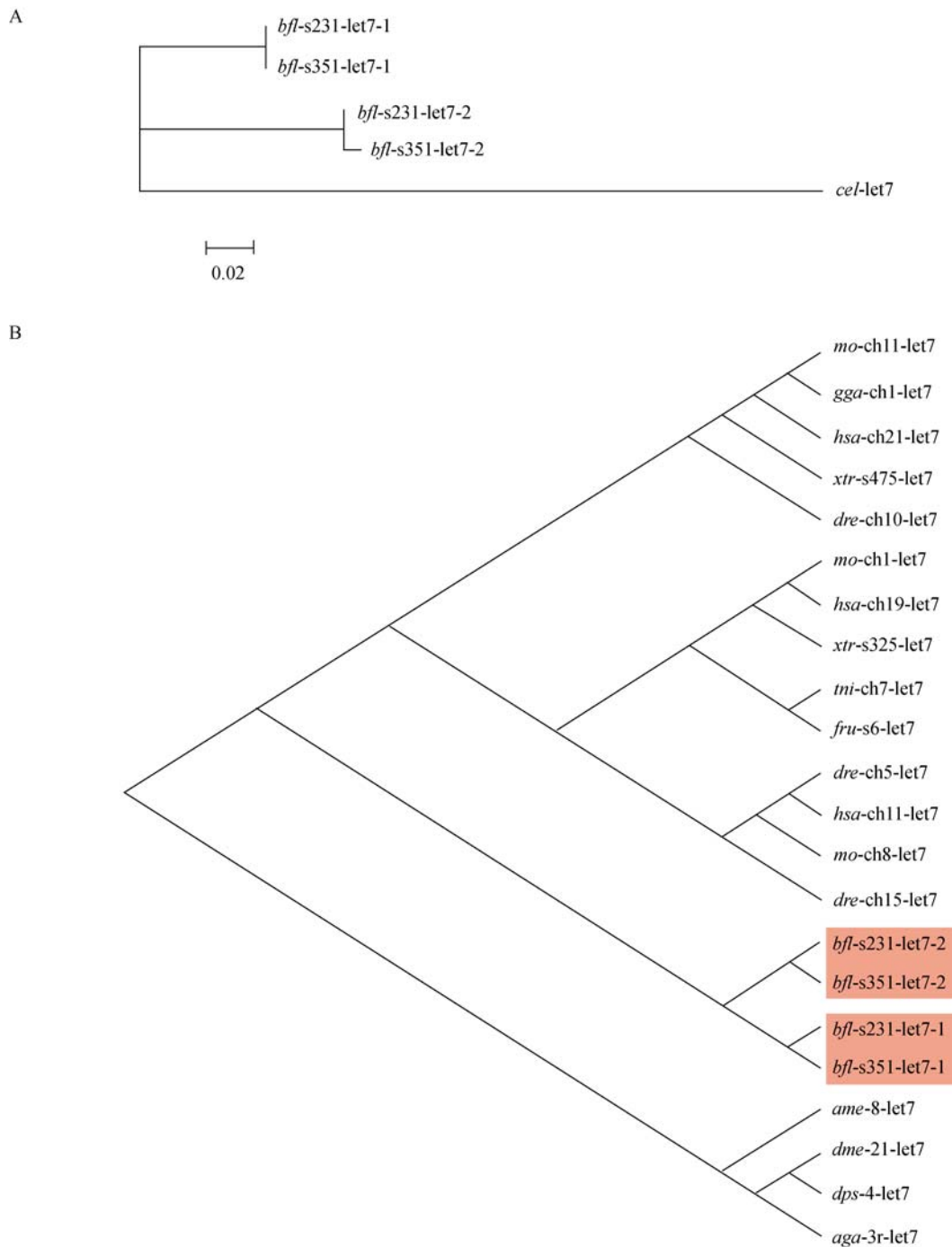
Figure 3 Phylogenetic relationship of let-7 hairpins of amphioxus. **A**. Neighbor-joining tree of four let-7 miRNAs in *B. floridae*. The let-7 gene of *Caenorhabditis elegans* (cel) was used as outgroup. **B**. Neighbor-joining tree of let-7 hairpins among species. The let-7 miRNAs of *B. floridae* are highlighted in red. Abbreviations are the same as Figure 2.

To exclude the possibility that the observation of two let-7 clusters is due to the incompleteness of assembly of genome sequences and in fact these two clusters are the same cluster in corresponding chromosome, mir-96 and mir-216 clusters in different scaffolds were also selected for phylogenetic analysis. The results are similar to that of let-7 cluster (data not shown). Taking all together, it seems that amphioxus did experience local, species-specific gene duplication during its evolutionary history.
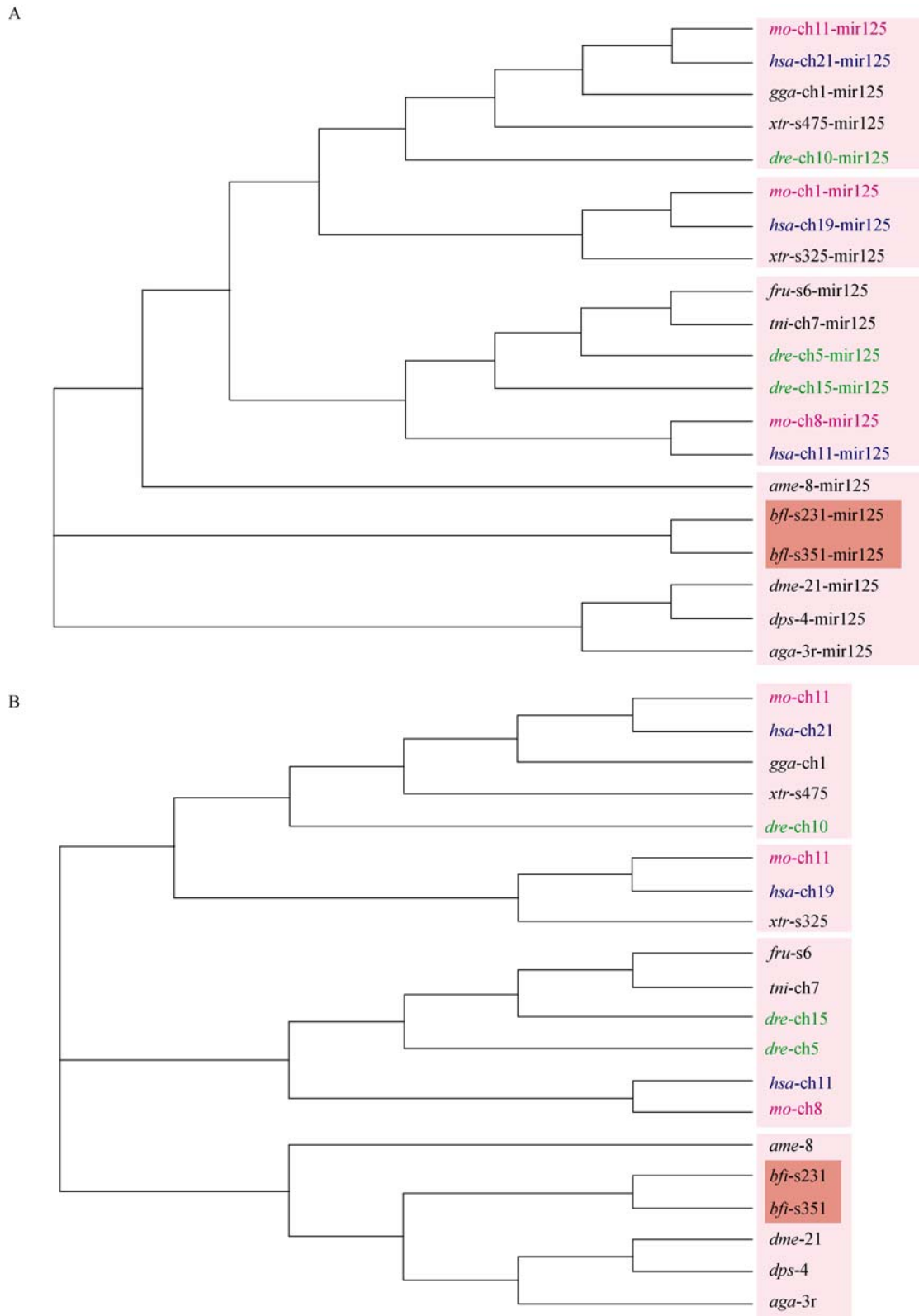
Figure 4 Neighbor-joining trees of mir-125 hairpins and let-7 cluster-combination among species. **A**. Neighbor-joining tree of mir-125 hairpins within conserved let-7 clusters among species. **B**. Neighbor-joining tree of combination sequence of let-7 clusters among species. The miRNAs of *B. floridae* are highlighted in red. miRNAs of human, rat and fish are noted in dark blue, purple and green, respectively. Abbreviations are the same as Figure 2.

Discussion

miRNAs have a high conservation characteristic, making prediction by homology-searching feasible and reliable. According to a series of strict filtering procedures, we obtained a list of miRNAs in amphioxus. These miRNAs have an apparent cluster-distribution feature, which is also a common feature of miRNA genes. In addition, we have sequenced a part of genome sequences of *Branchiostoma belcheri tsingtausus* (data unpublished), a species close to *B. floridae*, and also found some of the homologous miRNAs (data not shown). These restults make the predicted miRNAs more reliable.

In this study, we have found that most of miRNA families that are known to be conserved among protostomes and deuterostomes, at least in some representative species of each group, also have their close homologs in amphioxus, except MIPF0000039, which has not been found in *Fugu rubripes*, too. This may be due to the mutation that makes distinguishability impossible, or because of local gene loss. In protostomes, especially nematodes, there are extremely limited conserved miRNA families with vertebrates (nematodes: 4 and fruit fly: 9, both including MIPF0000039). Even in urochodates, there are only a small portion of miRNA families conserved with vertebrates or protostomes. This imbalanced loss and gain of miRNAs among individual species make miRNA phylogenetic research more complicated. Therefore, the wide conservation of miRNAs in amphioxus lends more convincing evidence for the view on miRNA conservation between invertebrates and vertebrates.

Amphioxus has more conserved miRNA families with vertebrates than protostomes or other chordates. This may demonstrate the closer relationship between cephalochordates and vertebrates, though the small number of homologous miRNA families in ciona may be partially due to its gene loss or wide mutation, and also there might be a lot of unknown miRNAs conserved between ciona and amphioxus. Because of introduction of amphioxus miRNAs, many families that had been thought as vertebrate-specific or mammal-specific before should be classified as chordate-specific or deuterostome-specific more precisely. Though there are very few families existing in amphioxus (or ciona) that had been thought as protostome-specific or nematode/fly-specific before, the presence of these families does change the view of such species-specificity of miRNAs and provides support that more miRNAs are in fact conserved evolutionarily rather than species-specific.

When compared with those of ciona, we found homologous miRNAs are much richer in amphioxus, in terms of families or copies per family. Average copy number per family in ciona is about 1 while in amphioxus approximating to 2. This is somewhat different with the phenomenon that most transcribed genes in amphioxus only contain single copy (*13*). We assume that such miRNA gene expansion is due to local, species-specific gene duplication. It has been suggested that part of the genes in amphioxus have experienced local duplication, and the same phenomenon existed in other species such as fishes (*13*). Phylogenetic analysis of repeated clusters and genes further confirmed this assumption. The same clusters or repeated miRNAs located in different region of genome have a very close relationship and the divergence is much later than the divergence of amphioxus species (Figure 2).

Though the miRNAs obtained here are close homologs to the known ones identified in other species, and the miRNAs share a lot of distribution features, such as clustering, with that of other species, further experimental confirmation is required to observe their expression and examine their regulation function. There are also many other miRNAs whose homologs are not present or identified in other species. In fact, we have tried some computational tools to predict miRNAs in amphioxus. One software called miRscan (*22*), taking the conservative genomic regions between amphioxus and ciona, found hundreds of possible miRNAs that maybe uro-cephalo-chordate specific. Another tool considering Drosha-processing sites (*23*) also obtained 361 miRNA candidates, of which 7 were also homologous ones. We are now performing biological experiments to confirm their expression.

Recently, Luo and Zhang (*24*) published a manuscript on amphioxus miRNA prediction using the methods similar to ours. Comparing with their results, we found 13 additional miRNAs that they have not found. Moreover, we used the same strategy to predict

miRNAs in ciona, making the comparison between amphioxus and ciona more reasonable. Besides observation of individual miRNA conservation in different species, which was also performed by Luo and Zhang (*24*), we further compared the total numbers of miRNA paralogues in species of phylogenetically different stages, in order to give some clues about the possible amplification of miRNA genes in evolution history and the importance of miRNAs in regulation of life processes. Finally and most importantly, through a carefully phylogenetic analysis, we found that the miRNAs are duplicated in amphioxus through a species-specific local duplication, rather than segmental-duplication or whole-genome-duplication manners. These features make our work unique and interesting. Because of the difference of parameter values, there are also 11 miRNAs in their results that we have not found. So it would be desired to perform a combination of our results and theirs, which would give a more comprehensive analysis and comparison of amphioxus miRNA genes.

Conclusion

Amphioxus has a group of conserved miRNA genes and families. These miRNAs have similar sequence and distribution characteristics with those in other species. Compared to more primary species, the miRNA genes in amphioxus have experienced an expansion, and this expansion potentially has nothing to do with genome duplication, but just local gene/cluster duplication.

Materials and Methods

Genome and miRNA sequences

The *B. floridae* masked genomic sequences and the *C. intestinalis* genome were downloaded from JGI website at http://genome.jgi-psf.org (*16*). Known metazoa miRNAs were downloaded from miRBase (miRBase version 10.0; http://microrna.sanger.ac.uk/sequences/).

Homologous miRNA searching

All mature and mature-star miRNAs were blastered against *B. floridae* genomes with the non-strict limits "-e 1.8 -W 7". Then Perl scripts were written to retrieve all the closely similar sequences (non-identities <6 nt) and corresponding flanking sequences. For mature miRNA homologs, the 3'-side 70 nt flanking sequence and 5'-side 20 nt sequence were retrieved, and for mature-star miRNA homologs, the 5'-side 70 nt sequence and 3'-side 20 nt sequence were retrieved. Because some mature sequences reside in the 3'-arm and some mature-star sequences in the 5'-arm of the hairpins, another round of searching was performed to retrieve 5'-side 70 nt flanking sequences of mature miRNA homologues and 3'-side 70 nt flanking sequences of mature miRNA homologs. The homologs retrieved with flanking sequences were prepared for structure analysis with RNAfold package (*25*). For the obtained sequences with structural and MFE information, some filtering procedures were performed, including (1) hairpin forming, (2) mature miRNA homolog lying in stem region and (3) no loops in homolog region. The resulting homologous hairpins were further filtered to exclude stuttering sequences. For precise evaluation of the candidates, manually piece-by-piece analysis was performed, and the MFE and non-symmetry of stems were taken into account. The candidates conformed to all the stringencies stated about were recorded as potential ones, while other ones were also recorded but as possible ones for those that had been well-reported conserved miRNAs and only few small conditions could not be satisfied. The miRNA homologs of *C. intestinalis* were analyzed with similar procedures.

Classification, clustering and inter-species comparison of miRNA

According to the known miRNA family summarized in miRBase, each amphioxus miRNA candidate was classified into a family that its closest homolog belongs to. The site of each miRNA candidate was recorded and its site relationship with other candidates was calculated and compared. Taking 2,500 nt as inter-genic range of a single cluster, the candidates were recognized as a cluster or not. The known miRNAs of one species were classified into families as that of amphioxus, and inter-species conserved families were retrieved by Perl scripts.

Phylogenetic analysis

The hairpin sequence of each miRNA paralog in one cluster was retrieved from miRBase, and together with that of amphioxus, cluster-gene sequences were compared and the distance between each other was calculated. The distances were used for constructing phylogenetic trees, based on neighbor-joining method (*26*). The trees were drawn with MEGA3 software (*27*).

Authors' contributions

LW and LJ collected the datasets, conducted data analyses, and co-wrote the manuscript. SH and YW supervised the project. All authors read and approved the final manuscript.

Competing interests

The authors have declared that no competing interest exist.

References

1 Ambros, V. 2004. The functions of animal microRNAs. *Nature* 431: 350-355.

2 Kim, V.N. and Nam, J.W. 2006. Genomics of microRNA. *Trends Genet.* 22: 165-173.

3 Rodriguez, A., *et al.* 2004. Identification of mammalian microRNA host genes and transcription units. *Genome Res.* 14: 1902-1910.

4 Kim, V.N. 2005. MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* 6: 376-385.

5 Lagos-Quintana, M., *et al.* 2001. Identification of novel genes coding for small expressed RNAs. *Science* 294: 853-858.

6 Lau, N.C., *et al.* 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294: 858-862.

7 Lee, R.C. and Ambros, V. 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 294: 862-864.

8 Altuvia, Y., *et al.* 2005. Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res.* 33: 2697-2706.

9 Cai, X., *et al.* 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* 10: 1957-1966.

10 Lee, Y., *et al.* 2004. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* 23: 4051-4060.

11 Kong, Y. and Han, J.H. 2005. MicroRNA: biological and computational perspective. *Genomics Proteomics Bioinformatics* 3: 62-72.

12 Gans, C. 1989. Stages in the origin of vertebrates: analysis by means of scenarios. *Biol. Rev. Camb. Philos. Soc.* 64: 221-268.

13 Minguillón, C., *et al.* 2003. The amphioxus Hairy family: differential fate after duplication. *Development* 130: 5903-5914.

14 Blair, J.E. and Hedges, S.B. 2005. Molecular phylogeny and divergence times of deuterostome animals. *Mol. Biol. Evol.* 22: 2275-2284.

15 Delsuc, F., *et al.* 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439: 965-968.

16 Putnam, N.H., *et al.* 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453: 1064-1071.

17 Holland, L.Z., *et al.* 2008. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res.* 18: 1100-1111.

18 Hertel, J., *et al.* 2006. The expansion of the metazoan microRNA repertoire. *BMC Genomics* 7: 25.

19 Sempere, L.F., *et al.* 2006. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J. Exp. Zool. B Mol. Dev. Evol.* 306: 575-588.

20 Dehal, P., *et al.* 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298: 2157-2167.

21 Missal, K., *et al.* 2005. Non-coding RNAs in *Ciona intestinalis*. *Bioinformatics* 21: ii77-78.

22 Lim, L.P., *et al.* 2003. The microRNAs of *Caenorhabditis elegans*. *Genes Dev.* 17: 991-1008.

23 Helvik, S.A., *et al.* 2007. Reliable prediction of Drosha processing sites improves microRNA gene prediction. *Bioinformatics* 23: 142-149.

24 Luo, Y. and Zhang, S. 2009. Computational prediction of amphioxus microRNA genes and their targets. *Gene* 428: 41-46.

25 Hofacker, I.L., *et al.* 1994. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.* 125: 167-188.

26 Saitou, N. and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.

27 Kumar, S., *et al.* 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* 5: 150-163.