Gene 428 (2009) 41–46

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Computational prediction of amphioxus microRNA genes and their targets

Yan Luo *, Shicui Zhang

College of Marine life Sciences, Ocean University of China, Qingdao, Shandong 266003, PR China

ARTICLE INFO

Article history:
Received 5 August 2008
Received in revised form 19 September 2008
Accepted 19 September 2008
Available online 1 October 2008

Received by I. King Jordan

Keywords:
Amphioxus
Lancelet
Branchiostoma
MicroRNA

ABSTRACT

Using a homology search based on genomic survey sequence analysis and microRNA (miRNA) secondary structure, a total of 51 miRNA genes encoding 30 amphioxus miRNAs were identified. These amphioxus miRNAs belong to 27 miRNA families. Many miRNA genes were found in multiple locations and in more than one genomic context. Analysis of amphioxus miRNA genes revealed nearly 50% miRNA genes identified positioning within introns of protein-coding genes. MicroRNA genes were also positioned intergenically between known protein-coding genes. MicroRNA genes were also clustered throughout the genome, indicating the potential for the cotranscription and coordinate expression of miRNAs located in a given cluster. Cross-species comparison indicates that of the 27 miRNA families, 17 families are shared by both protostomia and deuterostomia, and 2, 3 and 5 families are protostome-specific, deuterostome invertebrate-specific and chordate-specific, respectively. Computational predictions of amphioxus miRNA targets, taking into account the relationship between miRNA target genes and their host genes/neighboring genes, show that 49 target sites for 34 potential target genes were complementary to 19 miRNAs. For the intronic miRNA genes co-expressed with the host gene, they may not be involved in the regulation of the host gene expression. However, some intergenic miRNAs could participate in the modulation of their neighboring gene expression.

1. Introduction

MicroRNAs (miRNAs) are small, endogenous noncoding RNA molecules that post-transcriptionally regulate expression of protein-coding genes (Bartel, 2004; Kloosterman1 and Plasterk, 2006), and play an important role in the control of diverse biological processes (Carrington and Ambros, 2003). MiRNAs are generated from long primary transcripts that are processed in multiple steps to cytoplasmic mature miRNAs consisting of about 22 nucleotides (nt). The mature miRNAs are incorporated into miRNA-induced silencing complex, which guides them to their target sequences. Most animal miRNAs recognize their target sites located in 3′ untranslated regions (UTRs) by incomplete base-pairing, which leads to mRNA degradation or translational repression of the target genes (He and Hannon, 2004; Bushati and Cohen, 2007).

A number of approaches have been used to define miRNAs in various organisms. Initially, miRNAs were identified using genetic or biochemical methods, like the first miRNAs, lin-4 and let-7, in *Caenorhabditis elegans* (Lee et al., 1993; Wightman et al., 1993). Later, direct cloning and sequencing of total small RNAs with appropriate size from isolated tissues or whole organisms enabled the identification of hundreds of miRNAs in plants and animals (Lagos-Quintana et al., 2001; Reinhart et al., 2002). Majority of recently identified miRNAs were first predicted by computational approaches and then validated by molecular techniques such as Northern blotting (Zhang et al., 2006). Apparently, computational approaches have played an increasingly important role in miRNA identification. A distinct advantage of computational approaches is that the miRNAs which are expressed in specific tissues, at certain stages of development or at low-copy number can be readily identified by computational searching, whereas they are difficult to identify and are often missed by approaches such as cloning and sequencing. The principles of computational approaches are based on the major characteristic features of miRNAs: hairpin-shaped stem-loop secondary structure with minimal folding free energy (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001) and high evolutionary conservation from species to species (Grad et al., 2003; Lai et al., 2003; Lim et al., 2003; Berezikov et al., 2005; Legendre et al., 2005). Accumulating evidence shows that many miRNAs are evolutionarily conserved in animals from worms to humans (Pasquinelli et al., 2000; Zhang et al., 2006), suggesting a powerful strategy to predict potential miRNAs by using homology search. In fact, homology search as a computational approach has been developed to identify miRNA genes in both plants and animals (Weber, 2005; Zhang et al., 2007). Using this approach, 35 new human and 45 new mouse miRNAs were identified by a BLAST search of the human and mouse genomes (Weber, 2005). Similarly, 154 new zebrafish miRNAs, 142 new *Xenopus* miRNAs, 58 new pig miRNAs, 91 new *Anopheles gambiae* miRNAs and 14 new *Ciona*

Abbreviations: ΔG, folding free energies; EST, expression sequence tag; nt, nucleotide(s); miRNA, microRNA; miRNA*, opposite miRNA sequence; pre-miRNA, microRNA precursor; pri-miRNAs, microRNA primary; UTR, untranslated region(s).

* Corresponding author.
E-mail address: luoyan@ouc.edu.cn (Y. Luo).

0378-1119/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2008.09.022

intestinalis miRNAs were identified (Chen et al., 2005; Chatterjee and Chaudhuri, 2006; Kim et al., 2006; Norden-Krichmar et al., 2007; Tang and Maxwell, 2008). Moreover, 338 potential plant miRNAs were identified from 60 different plant species by a BLAST search of the whole GenBank expressed sequenced tag (EST) database (Zhang et al., 2005).

Many miRNAs have been identified in various animals, including some model species, by computational and experimental approaches (Griffiths-Jones et al., 2006, 2008). For example, 678 miRNAs are documented from humans, 337 are from zebrafish, 152 from fruit fly and 154 from nematode. Amphioxus, a protochordate, occupies an evolutionarily critical nodal point transient from invertebrate to vertebrate, and has long been regarded as a model organism for insights into the origin and evolution of vertebrates (Stokes and Holland, 1998; Zhang et al., 2001). However, little study has been conducted to investigate amphioxus miRNAs although a couple of miRNAs have been predicted (Sempere et al., 2006; Heimberg et al., 2008). Because of the critical position of amphioxus in zoological world and the high conservation of miRNAs, information regarding miRNAs and their targets of amphioxus may contribute to understanding the role of specific miRNAs in gene regulatory networks for other organisms.

The objectives of this study were therefore to identify amphioxus miRNA genes from the recently completed draft assembly and automated annotation of the *Branchiostoma floridae* genome and to determine the potential targets of miRNAs, especially for the genes adjacent to the miRNA genes. These initial investigations also serve to establish amphioxus as a model protochordate for further investigations of miRNA biogenesis and function during development.

2. Materials and methods

2.1. MicroRNA reference set

To search potential *B. floridae* miRNAs, the list of known mature miRNA sequences for several species, including human, frog, zebrafish, ascidians, fruit fly and nematode, was obtained and used as query sequences. The reason for using these miRNAs as reference miRNAs is that these species are across widely divergent lineages in which a large number of miRNAs have been identified and deposited in publicly available databases. In total 1539 mature miRNA sequences of *Homo sapiens*, *Xenopus tropicalis*, *Danio rerio*, *Ciona intestinalis*, *Drosophila melanogaster* and *Caenorhabditis elegans* miRNAs (678, 184, 337, 34, 152 and 154 sequences, respectively) were downloaded from the miRNA database (miRBase Sequence Database, http://microrna.sanger.ac.uk; release 11.0, April 2008) (Griffiths-Jones et al., 2008). Although some of these miRNAs were initially identified by computational approaches, a majority of them have been validated by experimental approaches (Griffiths-Jones et al., 2008).

2.2. Computational prediction of amphioxus microRNA genes

The *B. floridae* genome has been completely sequenced (Putnam et al., 2008). Computational prediction of *B. floridae* miRNAs was performed with the genome data available from the Department of Energy Joint Genome Institute (JGI). The *B. floridae* genome, Assembly v1.0 (December 5, 2006), were obtained from the JGI website (http://genome.jgi-psf.org/Brafl1/Brafl1.home.html). A miRNA prediction algorithm was implemented to search the *B. floridae* genome for conserved mature miRNAs (Fig. 1). Matches of 90% identity of the known miRNAs were retained for further processing. To prune the number of potential hairpin structures, 100 nts upstream and downstream from the match boundaries were extracted from the genomes and examined for a hairpin structure. The RNA folding software mfold (Zuker, 2003) was used to confirm the hairpin structure as the lowest energy folding form. The potential miRNA precursors conformed to the empirical criteria proposed by Ambros et al. (2003) in particular, a hairpin structure of the lowest free energy ($\Delta G < -25$ kcal/mol), as predicted by mfold, and a minimum of 16 nts of the mature miRNA engaged in Watson–Crick or G/U base pairings. The method used for searching the new miRNAs ensured their phylogenetic conservation.

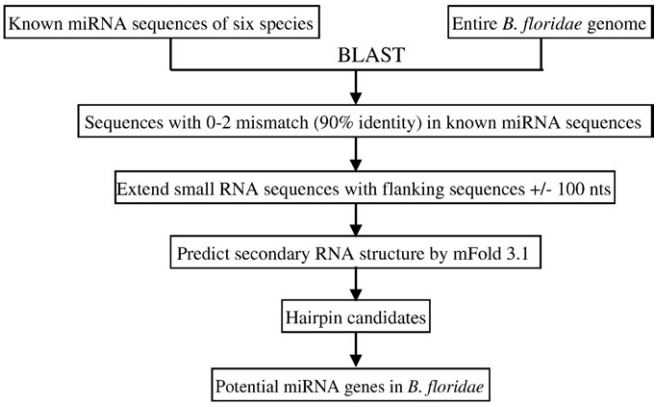

Fig. 1. MicroRNA genes prediction algorithm used to identify amphioxus homology to known miRNAs.

2.3. Computational prediction of amphioxus microRNA target genes

Because the *B. floridae* genome is incompletely annotated, many predicted transcripts across the genome did not contain the 5′ UTR and 3′ UTR regions. It is not easy to obtain the entire amphioxus 3′ UTR sequence data set. Therefore we chose those 49 genes, which are the neighboring genes (33) and host genes (16) for the putative *B. floridae* miRNA genes, to detect potential target sites for the amphioxus miRNA sequences. The purpose of choosing these genes is to determine whether the miRNAs regulate their host genes or neighboring genes. These 49 predicted gene transcripts for *B. floridae* were also obtained from the JGI website. These predicted mRNA sequences do not necessarily contain the 3′ UTR regions of mRNA. Thus, 1500 nts downstream from the transcript 3′ boundaries were extracted from *B. floridae* genome as the putative 3′ UTR regions of amphioxus mRNA.

We used the miRanda method (Enright et al., 2003; software version 2.0 as available at http://www.microrna.org/miranda) to detect potential target sites for the amphioxus miRNA sequences. The parameters and conditions employed are described as follows: match score $S \geq 90$ and target duplex free energy $\Delta G \leq -20$ kcal/mol; scaling parameter=2 for complementary nucleotide match score in positions 2–9, counting from the miRNA 5′ end; no non-Watson–Crick base pair at positions 2–9 and less than four G:U base pairs at positions 9–21. These parameters were also used to predicted targets for human and zebrafish miRNAs (John et al., 2004; Chen et al., 2005).

3. Results

3.1. Characterization of amphioxus microRNAs

Using the homology searches, a total of 51 miRNA genes are identified in the genome of amphioxus. Many miRNA genes were found in multiple locations and can be found at different genome positions. One possible reason for this was that little relevant work has been done on the amphioxus genome. Some repeated sequences would probably turn out to be identical in more advanced assemblies of the genome. Removing repeated sequences, we predicted 30 miRNA genes for *B. floridae* (Table S5). These miRNAs were found belonging to 27 miRNA families. Since miRNAs are believed to occur at a frequency of approximately 0.5–1.5% of the total genes in the genome (Carthew,

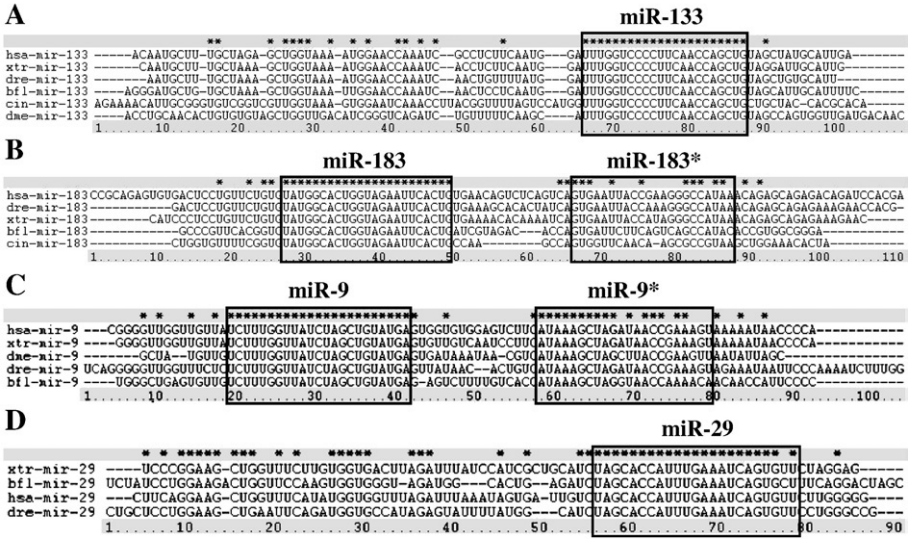

Fig. 2. ClustalX multiple sequence alignment of microRNAs sequences from various species. (A–D) Alignment of the sequences from *B. floridae* and reference species mir-133 (A), mir-183 (B), mir-9 (C) and mir-29 (D). Asterisks indicate conserved positions in sequence alignment. The sequences of mature miRNA and minor miR* are boxed. Species abbreviations: dme—*Drosophila melanogaster* (fruit fly); cin—*Ciona intestinalis* (ascidians); bfl—*Branchiostoma floridae* (amphioxus); dre—*Danio rerio* (zebrafish); xtr—*Xenopus tropicalis* (pipid frog); hsa—*Homo sapiens*.

2006), the 21,900 genes of *B. floridae* (Putnam et al., 2008) should have generated 109 to 328 miRNAs. Thus, the number of miRNAs from this computational prediction is apparently much lower than the actual number of miRNA genes in amphioxus.

Two members were identified in the let-7, miR-33 and miR-200 families while only a single member was found in the majority of miRNA families. The length of miRNA precursors varied little in amphioxus, ranging from 79 to 93 with an average of 86 nts. Mature miRNAs were usually located at each arm of the stem-loop hairpin structure; some were at the 5′ end of the miRNA precursor sequences, others at the 3′ end. In the 30 identified amphioxus miRNAs, 12 were located at the 3′ end while 18 were located at the 5′ end.

Using the clustalX program, sequences similarities between mature miRNAs from different species were identified, and the precursor miRNAs were also analyzed. Table S1 showed the pairwise sequence identity of the mature miRNAs and pre-miRNAs between amphioxus and other reference species. Supplement figure showed the alignments of pre-miRNA from various species. Mature microRNAs are better conserved than the precursor sequences. Of the miRNAs identified in amphioxus, eight mature sequences, including miR-7, miR-133, miR-9, miR-183, miR-10, miR-125, miR-96 and miR-71, were 100% identical with the known reference species' miRNAs, others were found to be at least 77% similar to the miRNAs of various other organisms. However, the corresponding pre-miRNA sequences were not conserved to that extent (Table S1). For example, the mature sequence of bfl-miR-71 was 100% identical with dme-miR-71, while their precursor miRNAs only were 24% identical. In these miRNAs identified in amphioxus, some miRNA families were evolutionarily highly conserved. For example, the mature sequences of miR-133 of the five species, including amphioxus, human, frog, zebrafish, ascidians and fruit fly, were 100% identical, and their pre-miRNA sequences still had 76.4% sequence identity (Table S1, Fig. 2A). Besides, miR-183, miR-9, miR-29, miR-125, miR-100 and miR-33 were also the well-conserved miRNAs (Fig. 2B–D, Fig. S1). *B. floridae* miRNAs had few nucleotide differences to the corresponding miRNAs of the vertebrate species including human, frog and zebrafish. The mean identity score of total miRNAs between amphioxus and other reference species were estimated and presented in Table S1. The mean identity of total mature miRNAs of *B. floridae* is 97.5%, 97.8% and 97.9% identical with those of human, frog and zebrafish, respectively. However, *B. floridae* miRNAs possessed more nucleotide differences compared to ascidian miRNAs, for the mature miRNAs sequence identity between amphioxus and ascidian is 93.8%, and the pre-miRNA sequence identity is 40.7%. These suggest that amphioxus miRNAs are less divergent than ascidian miRNAs from vertebrate miRNAs.

The available *B. floridae* genome databases enabled a determination of miRNA genomic context. Recent analyses of miRNA gene locations showed that the majority of mammalian miRNA genes are located in defined transcription units (Rodriguez et al., 2004). We found that the amphioxus miRNA precursor sequences were distributed over 51 unique genomic locations (Table S2). Based on their genomic locations, amphioxus miRNA genes were categorized into three types (designated in Table 1): intronic miRNAs (9 miRNAs) located within introns; intergenic miRNAs (12 miRNAs) located within intergenic regions; and intronic and intergenic miRNAs (6 miRNAs) present within either an intron or intergeneic regions. Approximately 50% of the miRNAs precursors (25 sequences) were located within introns, 14 in the sense orientation, and 11 in the antisense orientation. A majority of intronic miRNAs were encoded in the putative introns, predicted to be within the hypothetical protein-coding sequences; and a few miRNAs were located within the introns of known protein-coding genes. Other miRNAs (26 sequences) were located within intergenic regions. For example, bfl-miR-7 was located between fgenesh2_pg.scaffold_113000077 (encoding hypothetical protein LOC495333) and estExt_fgenesh2_pg.C_1130076 (encoding heterogeneous nuclear ribonucleoprotein K, like); bfl-miR-10 was located 2410 nts upstream of the *AmphiHox4* gene. No miRNA was found to be located in the exons.

Table 1
Genomic organization of identified amphioxus microRNAs

Intronic miRNAs (9)	bfl-miR-124, bfl-miR-33, bfl-miR-183, bfl-miR-29, bfl-miR-96, bfl-miR-22, bfl-miR-190, bfl-miR-71 and bfl-miR-281
Intergenic miRNAs (12)	bfl-miR-1, bfl-miR-7, bfl-miR-31, bfl-miR-92, bfl-miR-133, bfl-miR-184, bfl-miR-9, bfl-miR-10, bfl-miR-210, bfl-miR-375, bfl-miR-200 and bfl-miR-141
Intronic and intergenic miRNAs (6)	bfl-let-7, bfl-miR-219, bfl-miR-100, bfl-miR-125, bfl-miR-216 and bfl-miR-217

MiRNA clustering is a common phenomenon in animals (Seitz et al., 2004; Tanzer and Stadler, 2004). We found that amphioxus miRNA genes were also clustered in the genome. We considered the pre-miRNAs as clusters if they mapped <20 kb apart and had the same direction of transcription as suggested by Chen et al. (2005). Almost 50% of the 51 identified *B. floridae* miRNA genes were observed to occur at least once in a miRNA cluster. Table 2 lists the amphioxus microRNA clusters. A given cluster could be a simple repeat of a single miRNA species (e.g., bfl-miR-92 cluster [scaffold 66] and bfl-miR-219 [scaffold 190]). However, the vast majority of clusters possessed different miRNAs. Interestingly, several clusters were found to be repeated in the amphioxus genome. Repeated clusters could be identical in miRNA composition, such as the bfl-let-7 cluster found on two different DNA scaffolds (scaffold 231 and 351), and bfl-miR-216 cluster found on scaffold 191, scaffold 1239 and scaffold 2270. This may be a result of gene duplication. The length of DNA scaffold containing miRNA clusters of *B. floridae* varied from several hundred to more than 10,000 bp. The close proximity of some clustered miRNAs suggests cotranscription in a single pri-miRNA transcript, whereas distantly positioned miRNAs may suggest otherwise (Tang and Maxwell, 2008). The bfl-miR-33 clusters found on both DNA scaffold 2 and scaffold 43 had a length of 497 bp and 491 bp respectively. The precursors of bfl-miR-216 and bfl-miR-217 were both located within the same intron of a protein-coding gene (Fig. 3). Therefore the bfl-miR-216 and bfl-miR-217 may co-transcripted in a pri-miRNA transcript. Five clusters, including the cluster let-7, miR-1, miR-96, miR-141, and miR-216, were also observed in vertebrates and/or invertebrates. The miR-96 cluster consists of miR-96, miR-183 and miR-182 in vertebrates, whereas, in amphioxus, miR-182 seems to be absent.

3.2. Potential microRNA targets

Current computational target prediction methods differ in their emphasis on qualitative aspects. The miRNA target prediction methods used in this study require a perfect match between 9 bp region of the 5′ end of the miRNA and its complementary target site, such as perfect complementary for positions 1–7, 1–8, 2–7, 2–8 and 2–9, of the 5′ end of the miRNA; we excluded G:U wobble pairs.

Table 2
Amphioxus microRNA clusters

miRNA gene cluster	Members	Locations	Families	Strand	Length (nt)
bfl-let-7	bfl-let-7-1,bfl-let-7-2, bfl-miR-100 and bfl-miR-125	Scaffold_351: 8381–10348	3	+	1968
bfl-let-7	bfl-let-7-1,bfl-let-7-2, bfl-miR-100 and bfl-miR-125	Scaffold_231: 1232058–1234264	3	+	2207
bfl-miR-1	bfl-miR-1 and bfl-miR-133	scaffold_740: 118343–125413	2	+	7071
bfl-miR-92	3 copies of bfl-miR-92	scaffold_66: 774431–870930	1	−	96500
bfl-miR-219	2 copies of bfl-miR-219	scaffold_190: 187573–353369	1	−	165797
bfl-miR-33	bfl-miR-33-1 and bfl-miR-33-2	scaffold_43: 73280–73776	1	+	497
bfl-miR-33	bfl-miR-33-1 and bfl-miR-33-2	scaffold_2: 6380998–6381488	1	+	491
bfl-miR-96	2 copies of bfl-miR-183 and 2 copies of bfl-miR-96	scaffold_37: 857541–1032232	2	−	174692
bfl-miR-141	bfl-miR-200b, bfl-miR-200c and bfl-miR-141	scaffold_96: 355970–361594	2	−	5625
bfl-miR-216	bfl-miR-216 and bfl-miR-217	scaffold_1239: 34899–35929	2	−	1031
bfl-miR-216	bfl-miR-216 and bfl-miR-217	scaffold_2270: 3464–4442	2	−	979
bfl-miR-216	bfl-miR-216 and bfl-miR-217	scaffold_191: 687375–688370	2	+	996

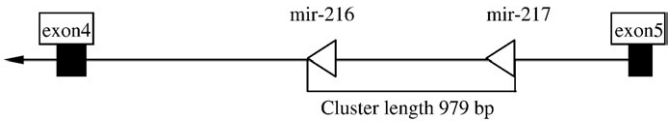

Fig. 3. Amphioxus mir-216 cluster. The white triangle indicates the location of a miRNA gene, and the exons are shown in black.

The 49 putative 3′ UTR sequences were input to the prediction pipeline as potential targets for amphioxus miRNA. Of the 27 miRNAs we identified, 19 had at least one target with those chosen genes (Table S3), 8 had no targets found yet. Totally, 49 target sites for 34 potential target genes were identified (Table S4). Of these predicted target genes in *B. floridae*, 9 had more than one predicted target sites for a given miRNA (Table S4). For example, fgenesh2_pg.scaffold_740000010 (PREDICTED: hypothetical protein) was hit by the miR-210 four sites. Twenty-two target genes were hit by only one miRNA; 12 target genes appeared to have multiple hits by several distinct miRNAs. Specific examples were that the ornithine aminotransferase-like 1 gene (gene ID: fgenesh2_pg.scaffold_37000049) was hit by the miR-133 (two sites), miR-210 (three sites), miR-375 (two sites) and miR-9 (one site); estExt_fgenesh2_pg.C_60113 (encoding protein phosphatase 4) was hit by the miR-124, miR-217 and miR-22. Multiple target sites on a 3' UTR for one or more miRNAs are not uncommon and reflect cooperative regulation of transcription (Chen et al., 2005).

4. Discussion

4.1. Phylogenetic distribution of microRNAs in amphioxus

In this study, a computational approach using all mature miRNA query sequences of six species ranging from protostomes to deuterostomes and from protochodate to vertebrates, 27 miRNA sequences are predicted in *B. floridae* genome. Relying on the currently available sequence information for human, frog, zebrafish, fruit fly, ascidians, sea urchin and nematode (miRNAs information of sea urchin from Hertel et al., 2006, Sempere et al., 2006 and Prochnik et al., 2007, not from the Sanger miRBase), the miRNA families can be subdivided by the pattern of evolutionary conservation into four subfamilies as follows (Fig. 4).

4.1.1. MicroRNAs conserved in both protostomes and deuterostomes

Of the 20 miRNA families that are shared by both protostomes and deuterostomes (Sempere et al., 2006) or the 30 miRNAs conserved across bilaterians (Prochnik et al., 2007), we predict 17 of these families exist in *B. floridae* (Fig. 4). Two phylogenetically conserved miRNAs, miR-34 and miR-153 conserved in chordate and sea urchin (Sempere et al., 2006; Prochnik et al., 2007), are not found in the genome of *B. floridae*. Thus, these two conserved miRNAs have probably been lost in the cephalochordate lineage.

4.1.2. MicroRNAs conserved in deuterostomes

Three amphioxus miRNAs, including miR-141, miR-183 and miR-200 are found to be present in both vertebrates and ascidian as well as sea urchin. They must therefore have been present in their last common ancestor. MiR-141 and miR-200, the two members of the miR-141 cluster, evolutionarily appear before the origin of the chordates (Hertel et al., 2006), and are thought to be the markers for sensory epithelia (Wienholds et al., 2005). Interestingly, in *B. floridae*, there are two tandem copies of miR-200 and a single copy of miR-141.

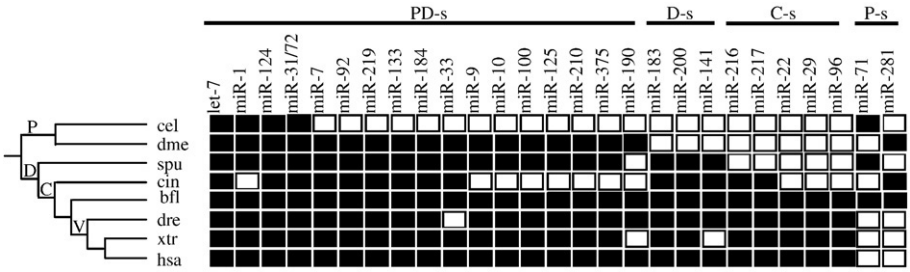

Fig. 4. Phylogenetically conserved microRNAs in amphioxus. The presence of miRNA is indicated by black; the absence of miRNA is indicated by white. Abbreviations: P—Protostomia; D—Deutostomia; C—Chordata; V—Vertebrata; PD-s—Protostomia and Deuterostomia specific; D-s—Deutostome-specific; C-s—Chordata specific; P-s—Protostome-specific; cel—*Caenorhabdits elegans* (nematode); dme—*Drosophila melanogaster* (fruitfly); spu—*Strongylocentrotus purpuratus* (sea urchin); cin—*Ciona intestinalis* (ascidians); bfl—*Branchiostoma floridae* (amphioxus); dre—*Danio rerio* (zebrafish); xtr—*Xenopus tropicalis* (pipid frog); hsa—*Homo sapiens*.

4.1.3. MicroRNAs in only chordate

Five miRNAs predicted in *B. floridae*, including miR-22, miR-29, miR-96, miR-216 and miR-217, are shared by both vertebrates and ascidian *C. intestinalis*, and thus could be chordate-specific miRNAs. These include two members of the miR-216 cluster miR-216 and miR-217 (Hertel et al., 2006), which have been presumed to be chordate-specific miRNAs (Heimberg et al., 2008) and are expressed in human pancreas (Lu et al. 2005), and miR-22 and miR-29, which are expressed abundantly in human muscle (Lu et al., 2005). In these five amphioxus miRNAs, three are found in vertebrates, including miR-22, miR-29 and miR-96, but not in ascidian. In addition, two ascidian miRNAs, miR-126 and miR-155 are found in vertebrates (Nordenkrichmar et al., 2007), but not in amphioxus. Therefore, the five miRNAs, including miR-22, miR-29, miR-96, miR-126 and miR-155, appear to be novel additions to chordate lineage.

4.1.4. MicroRNAs shared with invertebrates

Of the remaining two miRNAs in amphioxus miR-71 and miR-218 are protostomia-specific miRNAs (Sempere et al., 2006). MiR-71 present in most of protostomes is also shared by sea urchin, but not by ascidian and vertebrates; miR-281 present in most of protostomes is found in ascidian, but not in sea urchin and vertebrates. The presence of these two miRNAs in *B. floridae* genome therefore may be a reflection of its basal position in the phylum Chordata.

4.2. MicroRNA targets in microRNA neighboring genes

Of the 51 miRNAs identified, 25 miRNAs were located in the introns of protein-coding genes. Where clusters of miRNAs overlap with a single host transcript, the vast majority of miRNAs are located in the same intron. This phenomenon is also reported in human and other mammalian species, and indicates that miRNAs are commonly associated with complex transcriptional loci (Rodriguez et al., 2004). Other 26 miRNAs are located intergenically, their neighboring genes including upstream and downstream genes. These microRNA host genes and neighboring genes encode the proteins involved in metabolism, regulation of transcription, and cell signaling. Intronic miRNAs are linked to the transcription of the host genes. This finding raises the question whether the miRNAs regulate their host genes or neighboring genes.

In this study, a total of 34 genes including 9 microRNA host genes and 25 microRNA neighboring genes are regulated by 19 miRNAs. No intronic miRNA has target sites within its host gene. Five intergenic miRNAs, including miR-7, miR-10, miR-31/72, miR-141 and miR-210 have target sites in their neighboring genes (Table S4). For example, gene estExt_fgenesh2_pg.C_2530029 (encoding heterogeneous nuclear ribonucleoprotein K) is the nearest upstream gene for bfl-miR-7. There is only one target site by miR-7 on this gene. Therefore, the amphioxus miR-7 can regulate its neighboring gene.

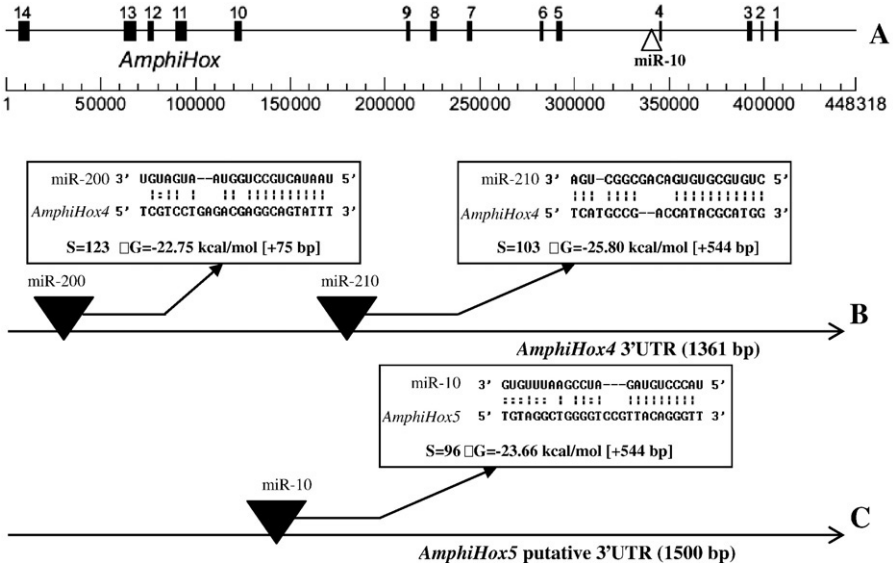

Fig. 5. Amphioxus Hox genes and microRNA targets. (A) Hox cluster organization of *B. floridae* (Amemiya et al., 2008). The location of the miR-10 gene is shown by a white triangle. (B) Representation of 3′ UTR of amphioxus *Amphihox4* gene for potential miRNA target gene. (C) Representation of 3′ UTR of amphioxus *Amphihox5* gene for potential miRNA target gene. Each individual hit between a miRNA and a target gene is marked by a black triangle on the UTR. Arrows indicate target site locations that are illustrated in more detail [alignment, score (S), duplex energy (ΔG)].

Another example is miR-10, which is the only phylogenetically conserved miRNA embedded in the Hox cluster, and appears to have a role in body plan patterning as opposed to tissue and organ specification (Pearson et al., 2005). Like vertebrates, bfl-miR-10 gene is located between Hox5 and Hox4 in amphioxus (Tanzer et al., 2005; Amemiya et al., 2008) (Fig. 5), 2410 nts upstream of the *AmphiHox4* gene. Thus, *AmphiHox5* and *AmphiHox4* are neighboring genes for miR-10, and are included in our miRNA targets study. *AmphiHox5* is regulated by miR-10 at position 544–569 of putative 3′ UTR (The *AmphiHox5* transcripts are not known in detail. Thus, 1500 nucleotides downstream from the transcript 3′ boundary were extracted from *B. floridae* genome as putative 3′ UTR of *AmphiHox5* mRNA.) (Fig. 5). *AmphiHox4* gene is regulated by miR-200 at position 75 and by miR-210 at position 544 of 3′ UTR (Fig. 5). Interestingly, these two Hox genes have the same complementary location (544 of 3′ UTR) for miRNAs. Therefore, miRNA can regulate the expression of its neighboring genes. Intronic gene is co-expressed with its host gene but does not regulate it.

Because of the limit in the prediction of miRNA targets for amphioxus in this study, we could not estimate the phylogenetic conservation with respect to other organisms. We expect further improvements in amphioxus genome annotation to provide an overview of its miRNAs and targets.

Acknowledgments

We thank Dr. I. King Jorden and an anonymous referee for their valuable comments on the manuscript. We are very grateful to Dr. Chun-lei Gao of SUNY-Upstate Medical University for her critical reading of the manuscript. This work was supported by a grant of Ministry of Science and Technology (MOST) of China (2008AA09Z411).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2008.09.022.

References

Ambros, V., et al., 2003. A uniform system for micro-RNA annotation. RNA 9, 277–279.
Amemiya, C.T., et al., 2008. The amphioxus Hox cluster: characterization, comparative genomics, and evolution. J. Exp. Zool. 310B, 465–477.
Bartel, D.P., 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281–297.
Berezikov, E., Guryev, V., van de Belt, J., Wienholds, E., Plasterk, R.H.A., Cuppen, E., 2005. Phylogenetic shadowing and computational identification of human microRNA genes. Cell 120, 21–24.
Bushati, N., Cohen, S.M., 2007. MicroRNA functions. Ann. Rev. Cell. Dev. Bio. 23, 175–205.
Carthew, R.V., 2006. Gene regulation by microRNAs. Curr. Opin. Genet. Dev. 16, 203–208.
Chatterjee, R., Chaudhuri, K., 2006. An approach for the identification of microRNA with an application to *Anopheles gambiae*. Acta Biochim. Pol. 53, 303–309.
Carrington, J.C., Ambros, V., 2003. Role of microRNAs in plant and animal development. Science 301, 336–338.
Chen, P.Y., et al., 2005. The developmental miRNA profiles of zebrafish as determined by small RNA cloning. Genes Dev. 19, 1288–1293.
Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., Marks, D., 2003. MicroRNA targets in *Drosophila*. Genome Biol. 5:RI, 1–14.
Grad, Y., et al., 2003. Computational and experimental identification of *C. elegans* microRNAs. Mol. Cell 11, 1253–1263.
Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., Enright, A.J., 2006. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 34, D140–D144.
Griffiths-Jones, S., Saini, H.K., van Dongen, S., Enright, A.J., 2008. miRBase: tools for microRNA genomics. Nucleic Acids Res. 36, D154–D158.
He, L., Hannon, G.J., 2004. MicroRNAs: small RNAs with a big role in gene regulation. Nat. Rev. Genet. 5, 522–531.
Heimberg, A.M., Semphere, L.F., Moy, V.N., Donoghue, P.C.J., Peterson, K.J., 2008. microRNAs and the advent of vertebrate morphological complexity. PNAS 105, 2946–2950.
Hertel, J., et al., 2006. The expansion of the metazoan microRNA repertoire. BMC Genomics 7, 25–39.
John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., Marks, D., 2004. Human miRNA targets. PLos Biol. 2, 1862–1879.
Kim, H.J., Cui, X.S., Kim, E.J., Kim, W.J., Kim, N.H., 2006. New porcine micro-RNA genes found by homology search. Genome 49, 1283–1286.
Kloosterman1, W.P., Plasterk, R.H.A., 2006. The diverse functions of microRNAs review in animal development and disease. Dev. Cell 11, 441–450.
Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T., 2001. Identification of novel genes coding for small expressed RNAs. Science 294, 853–858.
Lai, E.C., Tomancak, P., Williams, R.W., Rubin, G.M., 2003. Computational identification of *Drosophila* microRNA genes. Genome Biol. 4, R42.
Lau, N.C., Lim, L.P., Weinstein, E.G., Bartel, D.P., 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science 294, 858–862.
Lee, R.C., Feinbaum, R.L., Ambros, V., 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell 75, 843–854.
Lee, R.C., Ambros, V., 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. Science 294, 862–864.
Legendre, M., Lambert, A., Gautheret, D., 2005. Profile-based detection of microRNA precursors in animal genomes. Bioinformatics 21, 841–845.
Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., Bartel, D.P., 2003. Vertebrate microRNA genes. Science 299, 1540.
Lu, J., et al., 2005. microRNA expression profiles classify human cancers. Nature 435, 834–838.
Norden-Krichmar, T.M., Holtz, J., Pasquinelli, A.E., Gaasterland, T., 2007. Computational prediction and experimental validation of *Ciona intestinalis* microRNA genes. BMC Genomics 8, 445–457.
Pasquinelli, A.E., et al., 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature 408, 86–89.
Pearson, J.C., Lemons, D., McGinnis, W., 2005. Modulating Hox gene functions during animal body patterning. Nat. Rev. Genet. 6, 893–904.
Putnam, N.H., et al., 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064–1071.
Prochnik, S.E., Rokhsar, D.S., Aboobaker, A.A., 2007. Evidence for a microRNA expansion in the bilaterian ancestor. Dev. Genes Evol. 217, 73–77.
Weber, M.J., 2005. New human and mouse microRNA genes found by homology search. FEBS J. 272, 59–73.
Wienholds, E., et al., 2005. microRNA expression in zebrafish embryonic development. Science 309, 310–311.
Wightman, B., Ha, I., Ruvkun, G., 1993. Posttranscriptional regulation of the heterochronic gene *Lin-14* by *Lin-4* mediates temporal pattern-formation in C. elegans. Cell 75, 855–862.
Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., Bartel, D.P., 2002. microRNAs in plants. Genes Dev. 16, 1616–1626.
Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L., Bradley, A., 2004. Identification of mammalian microRNA host genes and transcription units. Genome Res. 14, 1902–1910.
Sempere, L.F., Cole, C.N., Mcpeek, M.A., Peterson, K.J., 2006. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. J. Exp. Zool. 306, 575–588.
Seitz, H., Royo, H., Bortolin, M.L., Lin, S.P., Ferguson-Smith, A.C., Cavaille, J., 2004. A large imprinted microRNA gene cluster at the mouse Dlkl-Gtl2 domain. Genome Res. 14, 1741–1748.
Stokes, M.D., Holland, N.D., 1998. The lancelet: also known as 'amphioxus', this curious creature has returned to the limelight as a player in the phylogenetic history of the vertebrates. Am. Sci. 86, 552–560.
Tang, G.Q., Maxwell, E.S., 2008. *Xenopus* microRNA genes are predominantly located within introns and are differentially expressed in adult frog tissues via post-transcriptional regulation. Genome Res. 18, 104–112.
Tanzer, A., Stadler, P.F., 2004. Molecular evolution of a microRNA cluster. J. Mol. Biol. 339, 327–335.
Tanzer, A., Amemiya, C.T., Kim, C., Stadler, P.F., 2005. Evolution of microRNAs located within Hox gene clusters. J. Exp. Zool. 304B, 75–85.
Zhang, B.H., Pan, X.P., Wang, Q.L., Cobb, G.P., Anderson, T.A., 2005. Identification and characterization of new plant microRNAs using EST analysis. Cell Res. 15, 336–360.
Zhang, B.H., Pan, X.P., Wang, Q.L., Cobb, G.P., Anderson, T.A., 2006. Computational identification of microRNAs and their targets. Comput. Biol. Chem. 30, 395–407.
Zhang, B.H., et al., 2007. Identification of cotton microRNAs and their targets. Gene 397, 26–37.
Zhang, S.C., Yuan, J.D., Li, H.Y., 2001. Amphioxus-model animal for insights into the origin and evolution of the vertebrates. Chin. Bull. Life Sci., 13, 214–218.
Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.