Identification of the *Angel*-related elements in cyprinid fishes and their phylogenetic implications*

WANG Li[1,2], HE Shunping[1]** and CHEN Yiyu[3]

(1. Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China; 2. Graduate School of the Chinese Academy of Sciences, Beijing 100039, China; 3. National Natural Science Foundation of China, Beijing 100085, China)

Accepted August 10, 2006

Abstract *Angel*-related element belongs to the family of miniature inverted-repeat transposable elements (MITEs). In this paper we report the identification of an *Angel*-related element in the series Leuciscini of cyprinid fishes, which is located in the second intron of the growth hormone (GH) gene. We have also found that this element is absent in orthologous locus in the series Barbini of cyprinid fishes, that provides new evidence for the monophyly of the series Leuciscini. The insertion of *Angel*-related element into the GH gene might take place in the common ancestor of the series Leuciscini after its divergence from the series Barbini. The high sequence divergence and relatively broad species distribution of *Angel*-related elements implies that they might be ancient transposons which appeared about 26 million years ago.

Keywords: ***Angel*-related elements, the series Leuciscini, monophyly.**

Miniature inverted-repeat transposable elements (MITEs) are moderately repetitive DNA elements interspersed in eukaryotic genomes. They can be identified by some common characteristics: short in size (80—500 bp) and no-coding capacity, generating target site duplications (TSDs) in the host genome where the insertion event takes place, having short terminal inverted repeats (TIRs) at both ends and the potential to form hairpin-like structures. If some taxa have a MITE inserted at exactly the same position in their genome, this is one of synapomorphies or shared derived characters. Thus, MITEs with a known position in the genome may provide phylogenetic signal as long as they are homologous[1]. For instance, the MITE family *Heartbreaker* (*Hbr*) was successfully used as molecular marker in maize[2]. However, phylogenetic implications of MITEs in vertebrates have not been investigated so far. Izsvák et al.[3] originally identified *Angel* elements in the eIF-4E (eukaryotic initiation factor-4E) gene in zebrafish (*Danio rerio*). They reported the presence of *Angel*-related elements (members of *Angel* family) in the second intron of the growth hormone (GH) gene in both silver carp (*Hypophthalmichthys molitrix*) and grass carp (*Ctenopharyngodon idellus*). However, the existence of the *Angel*-related element in the GH gene among other cyprinid fishes remains unknown.

In this paper, we report the isolation and identification of the *Angel*-related element in the GH gene from cyprinids and discuss their phylogenetic implications.

1 Materials and methods

1.1 Genomic DNA extraction, PCR amplification and sequencing

The samples including 27 cyprinid species from 11 subfamilies[4] were deposited in the Institute of Hydrobiology, Chinese Academy of Sciences (Table 1). All the tissues used for extraction of DNA were preserved in 95% ethanol, and total genomic DNA was extracted from muscles by standard proteinase K, phenol/chloroform extraction. According to the five sequences of the GH gene downloaded from GenBank (Table 1), a pair of primers (GF: 5′- CCAGCGGC-TYTTCAAYAAYGCAGT-3′; GR: 5′- AGTCA-GARTTGCAGAAAGACAGAGG-3′) were designed, and used to amplify intron II of the GH gene. PCR reactions were performed in a 60 μL volume containing approximately 100 ng of template DNA, 1.5 μL of each primer (10 μmol/ L), 6 μL of 10× buffer, 0.75 μL dNTPs (10 mmol/L), and 3.0 units *Taq* DNA polymerase. Amplification procedure included an initial denaturing step at 94℃ for 3 min, followed

* Supported by National Natural Science Foundation of China (Grant Nos. 30225008 and 30530120)

** To whom correspondence should be addressed. E-mail: clad@ihb.ac.cn

by 32 cycles of denaturation at 94℃ for 30 s, annealing at 62.5℃ for 45 s, extension at 72℃ for 1 min and a final extension at 72℃ for 5 min. PCR products were fractionated by electrophoresis through 1.2% low melting point agarose gels stained with ethidium bromide, then recovered from the gels and purified using BioStar glassmilk DNA purification kit according to manufacturer's protocol. The sequencing of target fragment was performed in both directions with an automated sequencing apparatus (Applied Biosystems 377 Stretch) and the sequences obtained were deposited in the GenBank database (Table 1).

Table 1. Samples used in this study and their GenBank accession numbers

Subfamily	Species	Collection locality	Voucher No.	Accession No.
Cultrinae	*Megalobrama pellegrini*	Hejiang, Sichuan	IHB0405133	DQ219449
	Megalobrama amblycephala			AF463498[a]
	Parabramis pekinensis	Yueyang, Hunan	IHB0411120	DQ219441
	Cultrichthys erythropterus	Lingshan, Guangxi	IHB0301134	DQ219450
	Pseudolaubuca sinensis	Taoyuan, Hunan	IHB0305194	DQ219440
	Pseudohemiculter dispar	Chongzuo, Guangxi	IHB1303127	DQ219439
Leuciscinae	*Elopichthys bambusa*	Taoyuan, Hunan	IHB040511	DQ219444
	Squaliobarbus curriculus	Jinkou, Hubei	IHB0301132	DQ219437
	Ctenopharyngodon idellus			X60419[a]
	Ochetobius elongatus	Tengxian, Guangxi	IHB0817003	DQ219447
	Mylopharyngodon piceus	Wuhan, Hubei	IHB0401127	DQ219438
Hypophthalmichthyinae	*Aristichthys nobilis*	Wuhan, Hubei	IHB040923	DQ219435
	Hypophthalmichthys molitrix			M94348[a]
Danioninae	*Nicholsicypris normalis*	Diaoluoshan, Hainan	IHB0301017	DQ219451
	Aphyocypris chinensis	Pengxian, Sichuan	IHB0305216	DQ219452
	Gobiocypris rarus	Pengxian, Sichuan	IHB0305124	DQ219448
	Opsariichthys bidens	Hengxian, Guangxi	IHB040687	DQ219445
	Zacco platypus	Xilin, Guangxi	IHB0303120	DQ219446
	Danio rerio			BX005440[b]
Acheilognathinae	*Rhodeus ocellatuas*	Leshan, Sichuan	IHB0207101	DQ219442
Gobioninae	*Saurogobio dabryi*	Hechuan, Chongqing	IHB0405356	DQ219443
Xenocyprinae	*Distoechodon tumirostris*	Jinkou, Hubei	IHB020636	DQ219436
Cyprininae	*Cyprinus carpio*			X51969[a]
	Carassius auratus	Wuhan, Hubei	IHB040925	DQ219453
Barbinae	*Spinibarbus sinensis*	Nanchong, Sichuan	IHB031259	DQ219454
Labeoninae	*Garra orientalis*	Tian'e, Guangxi	IHB0403442	DQ219456
Schizothoracinae	*Schizothorax meridionalis*	Tengchong, Yunnan	IHB03468	DQ219455

a) Sequences downloaded from GenBank; b) the sequence in zebrafish genome shows the strongest similarity to the GH gene of grass carp (e value of 2×10^{-70}).

1.2 Data analyses

Sequence alignment was performed using Clustal X[5], SEAVIEW alignment editor[6] and verified by eye, which included the five sequences of intron II of the GH gene retrieved from GenBank and our obtained sequences. Base compositional frequencies, nucleotide substitutions and sequence divergence between pairwise distances were estimated using MEGA 3.1[7]. The secondary structures of the *Angel*-related elements based on minimum-energy were folded using RNA structure 3.2[8].

Based on the second intron sequence of GH gene, maximum parsimony analysis was performed using PAUP* 4.0b10[9]. In this analysis, all sites were equally weighted, and heuristic search option was conducted with the tree bisection-reconnection (TBR) branch swapping algorithm. Non-parametric bootstrap analyses[10] were conducted with 1000 pseudoreplicates and 10 random sequences additions. When several equally parsimonious trees were obtained, a consensus tree was produced to summarize the data.

2 Results and discussion

2.1 Characteristics and identification of *Angel*-related elements

We obtained the sequences of intron II of the GH gene from 27 cyprinid species (Table 1). The alignment of these sequences revealed an insertion existed in 21 species with the sizes ranging from 67 to 306 bp. In four species, the insert fragment was obviously shorter, 67 bp in *Parabramis pekinensis*, 148 bp in *Aphyocypris chinensis*, 177 bp in *Aristichthys nobilis* and 224 bp in *Saurogobio dabryi*. The average base compositions of the insert fragments in 21 species were A = 34.3%, T = 32.2%, G = 16.4% and C = 17.2%, with the A + T contents (66.5%) much higher than that of G + C (33.6%). Overall transition/transversion ratio was 1.073. The sequence divergence was approximately 0.4%—22.4%.

The 21 insert fragments possessed all characteristics of the MITE family of eukaryotic genomes (Table 2), including the potential to form stem-loop secondary structures (figures not shown). While MITEs are classified into families according to TSD and TIR sequence[1]. Most of the 21 insert sequences had the same TSD sequences as that of typical *Angel* elements, their TIRs were about 26 bp in length and they showed strong similarity to that of *Angel* elements. All of these suggest that they belong to the *Angel* family. Furthermore, among the 21 sequences, the two in grass and silver carps were corresponding to the reported *Angel*-related elements[3]. For the other 19 sequences, they also showed relatively strong similarity to the above two ones. Thus, these 19 insert fragments were identified as *Angel*-related elements.

Table 2. The comparison of characteristics of the *Angel*-related element (in the GH gene of *Rhodeus ocellatuas*) and *Angel* element

Characters	*Angel*-related element	*Angel* element
Length (bp)	282	312
Target site duplication (TSD) sequence	TTAA	TTAA
Terminal inverted repeat (TIR) sequence	5′-TTAAAGGG-ATAGTTCACT-CAAAAGTA-3′	5′-TTAAAGGA-TAGTTCACCC-AAAAATGA-3′
Folding into stem-loop secondary structures	Yes	Yes

2.2 Phylogenetic implications of *Angel*-related elements in cyprinid fishes

When we integrated the *Angel*-related element in the GH gene into the maximum parsimonious tree (Fig. 1), it showed that the *Angel*-related element is present in all species of the series Leuciscini but is absent in the series Barbini and zebrafish. Morphological and molecular phylogenetic studies demonstrated two major lineages in Cyprinidae: the series Leuciscini and the series Barbini, though there was incongruence about what the two series included[11—14]. Therefore, our results indicated that the *Angel*-related element insertion could be taken as one of the shared characteristics of the series Leuciscini and further confirmed the monophyly of the series Leuciscini.

Because MITEs did not encode any transposase (TPase) or TPase remnant, they were considered as non-autonomous DNA elements[15]. It was proposed that they were capable of both insertion and excision under the catalysis of TPase from their related autonomous elements[16]. Insertion is always common, while excision is relatively rare[17]. Our data indicated that *Angel*-related elements might have inserted into the second intron of the GH gene in the ancestor of the series Leuciscini after its divergence from the series Barbini. It has been known that, after insertion, MITEs are subjected to random mutation, and the sequence and size homogeneity of a particular MITE family will decrease with time[18]. For the 21 *Angel*-related elements we obtained, they have variable sequence lengths, their TSDs and TIRs contain many mutations, and even a few of them missed one of TSDs or TIRs. This high degree of sequence divergence together with their relatively broad species distribution suggests that they should belong to a comparatively ancient transposon family. According to the reported divergence time between the series Leuciscin and the series Barbini[11], we estimated that this element might have appeared within the series Leuciscini 26 million years ago.

Comparing to other phylogenetic molecular markers such as mitochondrial or nuclear gene markers, MITEs insertions are advantageous to be used in phylogenetic studies: firstly, the insertion is usually irreversible because there is no known mechanism that specifically removes MITEs from the genome; secondly, the influence of convergent evolution and parallel evolution can be avoided; thirdly, phylogene-

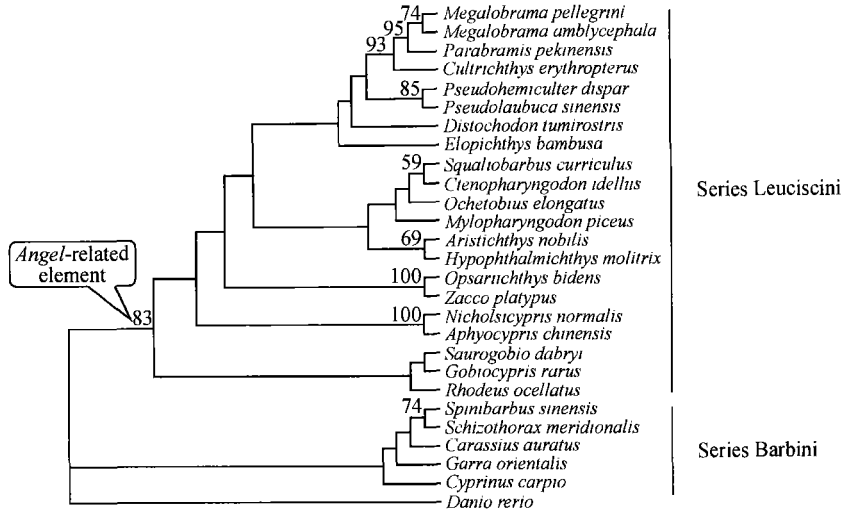

Fig. 1. The 50% majority consensus unrooted tree inferred from maximum parsimony method based on the second intron sequence of GH gene. Tree length = 461 and consistency index (CI) = 0.8091, retention index (RI) = 0.6879. Numbers at the nodes refer to the bootstrap; only values above 50% are reported. An *Angel*-related element is present in the second intron of GH gene in the series Leuciscini but absent in the series Barbini and zebrafish.

tic inference using MITEs does not rely on DNA sequence data but the presence or absence of the MITEs insertion, which can avoid the mistakes caused by mutational saturation and biased base composition or sequencing. However, resolving the phylogenetic relationships of a group usually needs using many MITEs insertion sites. Thus, the usage of multiple MITEs insertion sites, similar to SINEs[19], can be used as powerful tools for molecular systematics.

References

1 Petersen G. and Seberg O. Phylogenetic evidence for excision of *Stowaway* miniature inverted-repeat transposable elements in Triticeae (Poaceae). Mol. Biol. Evol., 2000, 17(11): 1589—1596.

2 Alexandra M. C., Cory B., Alexander N. et al. The MITE family *Heartbreaker* (*Hbr*): Molecular markers in maize. Proc. Natl. Acad. Sci. USA, 2000, 97(18): 10083—10089.

3 Izsvák Z., Ivics Z., Shimoda N. et al. Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. J. Mol. Evol., 1999, 48(1): 13—21.

4 Chen Y. Y. Fauna Sinica, Class Teleostei, Cypriniformes, Part II (in Chinese), Beijing: Science Press, 1998.

5 Thompson J. D., Gibson T. J., Plewniak F. et al. The Clustal_X windows interface: Flexible strategies for multiple sequences alignment aided by quality analysis tools. Nucleic. Acids. Res., 1997, 25(24): 4876—4882.

6 Galtier N., Gouy M. and Gautier C. SEAVIEW and PHYLO_WIN: Two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci., 1996, 12(6): 543—548.

7 Kumar S., Tamura K. and Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinformatics, 2004, 5: 150—163.

8 Mathews D. H., Disney M. D., Childs J. L. et al. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc. Natl. Acad. Sci. USA, 2004, 101: 7287—7292.

9 Swofford, D. L. PAUP*. Phylogenetic Analysis Using Parsimony (* and other methods). Version 4. Sunderland, Massachusetts, Sinauer Associates, 2002.

10 Felsenstein J. P. Confidence limits on phylogenies: An approach using the bootstrap, Evolution, 1985, 39: 783—791.

11 Wang X. Z., Li J. B. and He S. P. Molecular evidence for the monophyly of East Asian groups of Cyprinidae (Teleostei: Cypriniformes) derived from the nuclear recombination activating gene 2 sequences. Mol. Phylogenet. Evol., 2006 (Accepted).

12 Chen X. L., Yue P. Q. and Lin R. D. Major groups within the family Cyprinidae and their phylogenetic relationships. Acta Zootaxonomica Sinica (in Chinese), 1984, 9(4): 424—440.

13 Liu H. Z. and Chen Y. Y. Phylogeny of the East Asian Cyprinids inferred from sequences of the mitochondrial DNA control region. Can. J. Zool., 2003, 81(12): 1938—1946.

14 He S. P., Liu H. Z., Chen Y. Y. et al. Molecular phylogenetic relationships of Eastern Asian Cyprinidae (Pisces: Cypriniformes) inferred from cytochrome b sequences. Science in China (Series C), 2004, 47(2): 130—138.

15 Feschotte C., Jiang N. and Wessler S. R. Plant transposable elements: where genetics meets genomics. Nat. Rev. Genet., 2002a, 3(5): 329—341.

16 Feschotte C., Zhang X. and Wessler S. R. Miniature Inverted-Repeat Transposable Elements (MITEs) and Their Relationship with Established DNA Transposons, Mobile DNA Ⅱ. Washington, DC: American Society of Microbiology Press, 2002, 1147—1158.

17 Wessler S. R. Transposable elements associated with normal plant genes. Physiol. Plant., 1998, 103(4): 581—586.

18 Santiago N., Herráiz C., Goñi J. R. et al. Genome-wide analysis of the Emigrant family of MITEs of *Arabidopsis thaliana*. Mol. Biol. Evol., 2002, 19(12): 2285—2293.

19 Shedlock A. M. Okada N. SINE insertions: powerful tools for molecular systematics. BioEssays, 2000, 22(2): 148—160.