Chinese Science Bulletin

Cloning and sequence analysis of *Sox* genes in a tetraploid cyprinid fish, *Tor douronensis*

GUO BaoCheng[1,2], LI JunBing[1,2], TONG ChaoBo[1,2] & HE ShunPing[1†]

[1] Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China;
[2] Graduate University of the Chinese Academy Sciences, Beijing 100049, China

A PCR survey for *Sox* genes in a young tetraploid fish *Tor douronensis* (Teleostei: Cyprinidae) was performed to access the evolutionary fates of important functional genes after genome duplication caused by polyploidization event. Totally 13 *Sox* genes were obtained in *Tor douronensis*, which represent *Sox*B, *Sox*C and *Sox*E groups. Phylogenetic analysis of *Sox* genes in *Tor douronensis* provided evidence for fish-specific genome duplication, and suggested that *Sox19* might be a teleost specific *Sox* gene member. Sequence analysis revealed most of the nucleotide substitutions between duplicated copies of *Sox* genes caused by tetraploidization event or their orthologues in other species are silent substitutions. It would appear that the sequences are under purifying selective pressure, strongly suggesting that they represent functional genes and supporting selection against all null allele at either of two duplicated loci of *Sox4a*, *Sox9a* and *Sox9b*. Surprising variations of the intron length and similarities of two duplicated copies of *Sox9a* and *Sox9b*, suggest that *Tor douronensis* might be an allotetraploidy.

Tor douronensis, *Sox* gene, tetraploidy, genome duplication, gene diversity

Sox genes, a gene family of transcription factors involved in a variety of development processes and characterized by the presence of a DNA-binding HMG domain, are found throughout the animal kingdom, and were first discovered as a group of genes related to the mammalian testis-determining factor gene *Sry*[1]. Based on their sequence similarity, function, gene structure and chromosome location, Bowles et al.[2] systematically analyzed the evolutionary history of the *Sox* gene in metazoan, and suggested that *Sox* genes can be divided into groups A—J. Schepers et al.[3] proposed that the mouse and human genomes contain 20 orthologous pairs of *Sox* genes and represent the *Sox* groups A—H, respectively. Interestingly, studies of *Sox* genes in teleost fishes, such as *Danio rerio* and *Takifugu rubripes*, showed that most of *Sox* genes in mammalian lineage have two copies in teleost lineage, which was supposed to be caused by fish-specific genome duplication and have indicated some hints of important functional genes evolution in duplicated genome[4,5].

The whole genome duplication caused by polyploidization event is thought to be the most economical and prompt approach to producing abundant genetic materials and has played a pivotal role in the evolution of vertebrates[6−8]. Ohno[6] proposed that two rounds of genome duplications (the "2R" hypothesis) had occurred during the early evolution of the vertebrate lineage: one in the common ancestor of all vertebrates and the other after the divergence of Agnatha (jawless vertebrates) and Gnathostoma (jawed vertebrates)[9−11]. Recently, extensive comparative genomics studies have revealed that teleost fishes experienced another round of genome duplication, the so-called "fish-specific genome duplication" (FSGD), which has been proposed to contribute a lot to the successful radiation of teleost species[12−16]. Interestingly, to date, polyploidy still widely exists in

Received January 28, 2008; accepted April 12, 2008
doi: 10.1007/s11434-008-0277-6
†Corresponding author (email: clad@ihb.ac.cn)
Supported by the National Natural Science Foundation of China (Grant No. 30530120)

diverse fish groups, such as Acipenseriformes, Salmoniformes and Cypriniformes, and may be a significant phenomenon in the evolution of fishes[17]. Some primary species, such as paddlefish *Polyodon spathula*, lungfish *Protopterus dolloi* and spotted gar *Lepisosteus oculatus*, are tetraploidy ($N = 4$). Species in the family Salmonidae are believed to have evolved from an ancestor in which an autotetraploidization event occurred 25 to 100 million years ago (MYA)[6,18]. In cyprinids, polyploids are quite common, both at the whole subfamily level and genus level, mainly in subfamilies Cyprininae, Schizothoracinae, and Barbinae. Various types of polyploidy have been observed in the family Cyprinidae, for example, crucian (*Carassius auratus*) are tetraploidy ($N = 4$, $2n = 100$), and *Schizothorax prenanti* are hexaploidy ($N = 6$, $2n = 148$). Cytological studies suggest that all species of the genus *Sinocyclocheilus* are tetraploidy and the genus may have a tetraploid ancestor[19]. Thus, it has been suggested that polyploidization events repeatedly occurred in those different lineages of cyprinids[20,21].

For the important role of the genome duplication in the organism evolutionary process, the molecular diversity mechanism of duplicated genome caused by polyplodization has been a research hotspot in the molecular evolution filed. Nowadays, studies of duplicated genome evolution mainly focus on yeast[22], plant[23 – 25] and teleost[26] that experienced ancient polyploidization. However, little is known about the molecular evolution of duplicated genome in those species that experienced recent polyploidization events or the so-called evolutionary young polyploidy. As is well known, immediately after a genome duplication event, those duplicated genes have a redundant function. Has natural selection on genes in the duplicated genome completely relaxed? Do those duplicated genes evolve at the same rate in young polyploidy? In order to address these questions, a PCR survey for *Sox* genes in a young tetraploid cyprinid fish *Tor douronensis* was performed. *Tor douronensis* is an evolutionary tetraploidy ($N = 4$, $2n = 100$) according to previous cytological work[20]. The latest study in our lab shows that *Tor douronensis* is a very young tetraploiy, and the speciation event occurred just about 2 million years ago (MYA)[27], which make it an ideal model for investigating the gen(om)e evolution in young polyploidy. The aims of this study are: (i) to investigate the evolution of *Sox* genes in *Tor douronensis*; (ii) to test whether *Sox* genes in teleost fish support the fish-specific genome duplication; (iii) to infer the origin of *Tor douronensis*.

1 Materials and methods

1.1 Experimental materials

A female specimen used in this study was collected from Mengla, Yunan Province, and identified following the classification system of Chen et al.[28]. Muscle used for genomic DNA extraction was preserved in 95% ethanol. Specimen was deposited in the Fish Collection of the Institute of Hydrobiology of the Chinese Academy of Sciences and the specimen voucher number is IHBCY0405871.

1.2 DNA extraction, PCR amplification, cloning and sequencing

Total genomic DNA was isolated from muscle tissue using the standard phenol-chloroform method[29]. *Sox* genes were amplified by genomic polymerase chain reaction (PCR) using two pairs of degenerate primers designated as *Sox*N and *Sox*9[30]. The sequences of degenerate *Sox*N primers are 5′-ATG AAY GCN TTY ATG GTN TGG-3′ and 5′-GGN CGR TAY TTR TAR TCN GG-3′. The *Sox*9 primers 5′-ATG AAY GCS TTY ATG GTI TGG-3′ and 5′-GTC IGG GTG RTC YTT CTT RTG YTG-3′ were specific for *Sox*9-related genes. PCR reactions were conducted in a volume of 50 μL PCR cocktail consisting of 1 × buffer with 1.5 mmol/L $MgCl_2$ (Invitrogen), 0.25 mmol/L dNTPs (Promega), 2.5 U *Taq* DNA polymerase (Invitrogen), 0.2 pmol/L each oligonucleotide primer, and 40—50 ng genomic DNA. The PCR amplification profile included an initial 3 min denaturing period at 94℃. The following PCR conditions were used for the two pair degenerate primers *Sox*N and *Sox*9: 35 cycles at 94℃ for 30 s, 52℃ for 1 min and 72℃ for 1 min, followed by a final extension at 72℃ for 5 min. Resulting PCR products were resolved on 1% agarose gels, and purified with AxyPrep DNA gel extraction kit (Axygen). Then the purification products were ligated and cloned using the TA cloning kit (Invitrogen). Plasmid DNA of clones of each PCR product was extracted and sequenced using an ABI 3730 automatic sequencer. To ensure authenticity, all sequences were sequenced in both directions.

1.3 Data analysis

(i) Sequences identification. The sequences of PCR clones and their reverse complements were aligned using Clustal X[31] and verified manually according to the sequencing maps. Then, Basic Local Alignment (TBLASTX and BLASTN) searches of the GenBank database were first made to determine the identity of those PCR clones and exon-intron structure of *Sox* genes with intron. The DNA sequences of *Sox* genes in *Tor douronensis* were translated into their putative amino acid sequences in MEGA 4.0[32], and further identified by the key signature residues character of *Sox* genes[23].

(ii) Phylogenetic analysis. The evolutionary phylogenetic relationships of *Sox* genes in *Tor douronensis* were analyzed using minimum evolution (ME) and Bayesian methods. ME analysis was performed with Equal Input matrix using the putative amino acid sequences in MEGA 4.0. Bayesian analysis was executed using the DNA sequence in Mrbayes 3.1.2[33], with the TrN + I + G determined by Modeltest 3.7[34] as the best-fit model of the sequence evolution. Starting trees were random. Four simultaneous Markov chains were run for 4000000 generations. Trees were sampled every 100 generations and finally 40000 trees were produced. Chain stationarity was achieved after 2000000 generations and therefore 20000 trees were subsequently discarded (burn-in). In phylogenetic analysis, known *Sox* genes of human *Homo sapiens*, zebrafish *Danio rerio* and pufferfish *Takifugu rubripes* were used for comparison and trees were rooted with human TCF7 gene (NM_201632.1).

(iii) Evolutionary analysis. The sequence similarities including both the nucleic acid and amino acid were calculated in MatGET v2.02[35] with BLOSUM50 scoring matrix. The synonymous and nonsynonymous substitution rates (*Ks* and *Ka*) were calculated using the modified Nei-Gojobori[36] (assumed transition/transversion bias = 3.631) model with MEGA 4.0. The number of nucleotide substitutions per site (*d*) between two sequences was estimated using Kimura 2-parameter model with MEGA 4.0. The *Z* test was also performed using MEGA 4.0 to detect deviation from neutrality of those duplicated genes caused by polyploidization.

2 Results

2.1 Identification and evolutionary phylogenetic analysis of *Sox* genes in *Tor douronensis*

Total 36 clones from the PCR products of two pairs degenerate *Sox*N and *Sox*9 primers were sequenced in our analysis. In order to assign names to *Sox* genes in *Tor douronensis* and determine orthology with counterparts in other organisms, we undertook a series of analyses based on sequence and organization of these genes. A summary of our analyses and proposed gene identification is presented in Table 1. First, those clone sequences were primarily identified and intron positions of *Sox* genes with intron in the HMG domain were also determined using the TBLASTX and BLASTN searches. Then, alignment of their DNA sequences (intron sequences were deleted.) were made, as shown in Figure 1(a). Figure 1(b) shows the putative amino acid se-

Table 1 Identified *Tor douronensis Sox* genes (and orthology with zebrafish *Sox* genes)

Tor douronensis Sox genes					Group	Zebrafish gene name
Original clone IDs	copies	size (b)	gene name	accession No.		
TDN15, TDN[a)]20	2	162	*Sox4ai*	EU399799	C	*Sox4a*
TDN16, TDN24	2	162	*Sox4aii*	EU399800		
TDN12, TDN14, TDN18	3	162	*Sox4b*	EU399801		*Sox4b*
TDN25, TD9[b)]112	2	162	*Sox11b*	EU399802		*Sox11b*
TD924—TD928, TD936	6	335(191) [c)]	*Sox9ai*	EU399808	E	*Sox9a*
TD941—TD944	4	627(483)	*Sox9aii*	EU399809		
TD9518—TD920	3	805(661)	*Sox9bi*	EU399810		*Sox9b*
TD953, TD958	2	823(679)	*Sox9bii*	EU399811		
TDN5, TDN10, TD9110,	3	144	*Sox14a*	EU399803		*Sox14a*[d)]
TD9113, TD9117	2	144	*Sox14b*	EU399804		*Sox14b*[d)]
TDN13, TDN19, TDN21	3	162	*Sox19b*	EU399807	B	*Sox19b*
TDN7, TDN9	2	162	*Sox21a*	EU399806		*Sox21a*
TDN4, TDN8	2	162	*Sox21b*	EU399805		*Sox21b*

a) TDN represents clones from the PCR products with primers *SoxN*; b) TD9 represents clones from the PCR products with primers *Sox9*; c) number in the brackets is the length of intron; d) *Sox* gene of *Takifugu rubripes*.

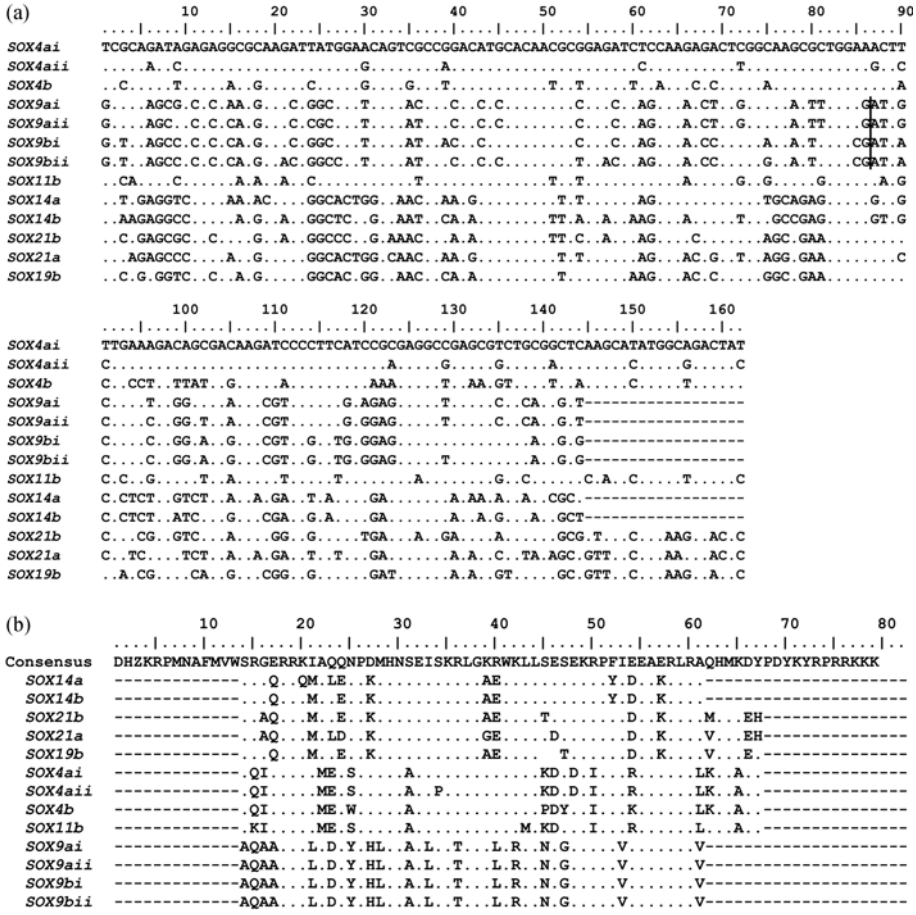

Figure 1 Alignment of the nucleic acid sequences (a) and putative amino acid sequences (b) encoded by the HMG domains of cloned *Tor douronensis* *Sox* genes. Identical residues are indicated by dots (·). Known intron positions are marked with (|) in (a).

quence of *Sox* genes in *Tor douronensis* they encode. According to BLAST searches and key signature residues character of *Sox* gene, 13 different *Sox* genes of *Tor douronensis* were obtained from 36 sequenced clones.

In order to exactly assign orthologues and names to *Sox* genes of *Tor douronensis*, evolutionary phylogenetic relationships of *Sox* genes of *Tor douronensis* and their corresponding genes from human or other teleost fishes were analyzed using ME and Bayesian methods. The result of ME analysis using putative amino acid sequences is shown in Figure 2(a). As is evident in the phylogenetic tree, *Sox* genes of *Tor douronensis* fall into three *Sox* gene groups, *Sox*B, *Sox*C and *Sox*E. The genes amplified with the *Sox*N primers represent the *Sox*B and *Sox*C groups and lack of introns in the HMG domain[30]. Their human orthologues are *Sox4*, *Sox11* and *Sox21*, and one *Sox* gene of *Tor douronensis* has no corresponding gene in human beings. The lengths of PCR products amplified with primers *Sox*9 range from 200 to 1300 bp. *Sox* genes obtained from those products belong to the *Sox*B and *Sox*E groups, and their human orthologues are *Sox9* and *Sox14*. Interestingly, in most cases, *Sox* genes in *Tor douronensis* have several copies, but in human there is only one copy. The phylogenetic tree of Bayesian analysis using the DNA sequences (Figure 2(b)) also supports the conclusion that the obtained *Sox* genes of *Tor douronensis* represent the *Sox*B, *Sox*C and *Sox*E groups. Their orthologues in teleost fishes zebrafish *Danio rerio* or pufferfish *Takifugu rubripes* are *Sox4b*, *Sox11b*, *Sox14a*, *Sox14b*, *Sox19b*, *Sox21a*, *Sox21b* and two isoforms *Sox4a*, *Sox9a* and *Sox9b*. The ortholog of one *Sox* gene in *Tor douronensis* without corresponding gene in human is *Sox19b* in zebrafish.

2.2 Sequence character and diversity analysis of *Sox* genes in *Tor douronensis*

Similarities including both nucleic acid and amino acid

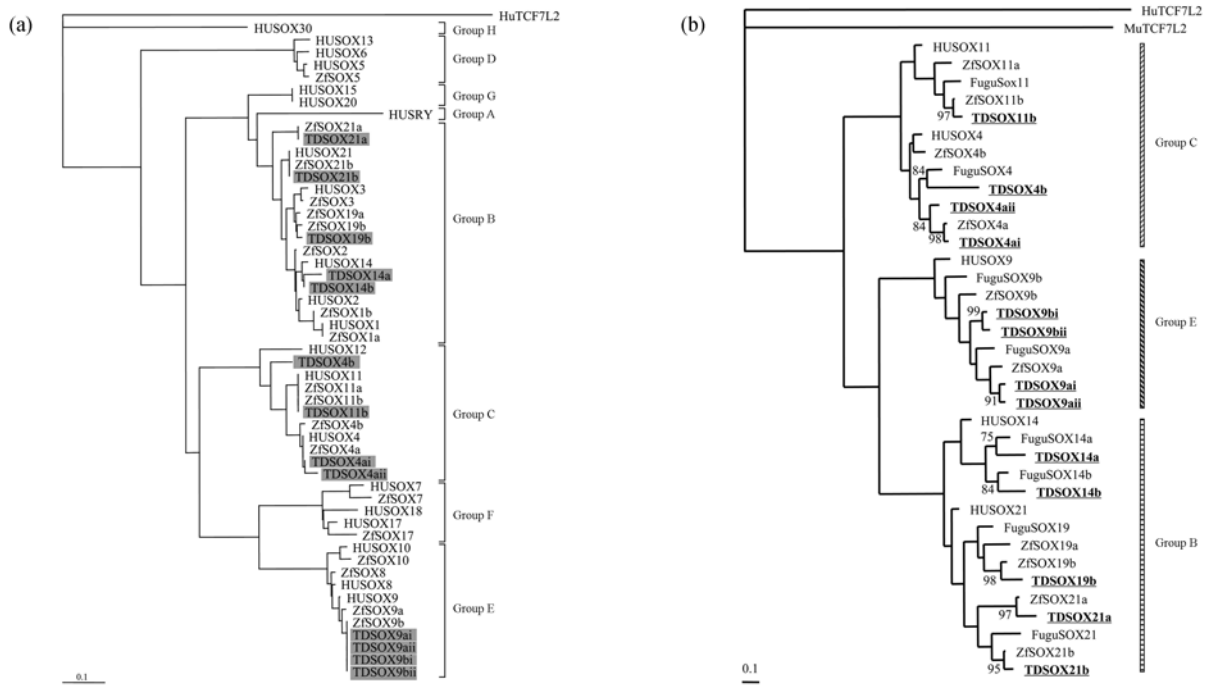

Figure 2 Rooted evolutionary phylogenetic analysis of *Sox* HMG domain in *Tor douronensis*. (a) The minimum evolution tree (Model: equal put) inferred from the aligment of putative amino acid sequences of *Sox* genes and (b) the Bayesian analysis inferred from the aligment of nucleid acid sequence of *Sox* genes, using Hu-TCF7L2 as an outgroup. *Sox* genes of *Tor douronensis* are shaded in (a) and underlined in (b). *Sox* groups are indicated on the right. Branch lengths are representatives of the extent of divergence. The numbers on the nodes of (b) represent Bayesian posterior probabilities (BPP). HU, Human *Homo sapiens*; TD, *Tor douronensis*; Zf, Zebrafish, *Danio rerio*; Fugu, *Takifugu rubripes*.

sequences between the homeologous/paralogous *Sox* genes of *Tor douronensis* and the reported orthologues from human *Homo sapiens*, zebrafish *Danio rerio* or pufferfish *Takifugu rubripes* were calculated, and listed in Table 2. Similarities of nucleic acid sequences between *Sox* genes of *Tor douronensis* and zebrafish or pufferfish are higher than orthologues from human, and there is no difference among amino acid sequences similarities, which suggests that most substitutions of *Sox* genes are synonymous substitutions. For the two copies *Sox4a*, *Sox4ai* and *Sox4aii*, their nucleotide similarity is 90.1%, while their similarities to zebrafish *Sox4a* are 96.9% and 90.7% respectively. The amino acid sequences of *Sox4ai* and zebrafish *Sox4a* are completely identical, and proteins similarity between *Sox4aii* and zebrafish *Sox4a* is 98.1%. Surprisingly, variation of intron length and similarity between different *Sox9a* and *Sox9b* copies are extremely distinct. The intron length of *Sox9ai* is 191 bp, while the intron length of *Sox9aii* is 483 bp, and their sequences similarity is just 35%. The intron length of two isoforms *Sox9b* are 661 and 679 bp, and sequences similarity is only 68.8%.

To determine whether natural selection on duplicated genes caused by polyploidization events is completely relaxed, we calculated the substitution rates of several *Sox* gene pairs of *Tor douronensis*. The results are summarized in Table 3. Analyses indicate that the ratio of non-synonymous to synonymous substitution rates (*Ka*/*Ks*) for those *Sox* gene pairs near to zero or equal to zero, and significantly reject the hypothesis that evolution of those duplicated genes caused by polyploidization is neutral.

3 Discussion

3.1 *Sox* gene complement of *Tor douronensis* and nomenclature

In total, clones of 13 *Tor douronensis Sox* genes have been obtained, including members of the *Sox*B, *Sox*C and *Sox*E group[2]. Each of these genes was represented by at least two independent clones, which make it very unlikely that any of the sequences presented in our analysis contain PCR artifact. According to the phylogenetic analysis and identification using key amino acid residues (especially for *Sox4b*), *Sox* genes obtained from *Tor douronensis* are *Sox4b*, *Sox11b*, *Sox14a*, *Sox14b*, *Sox19b*, *Sox21a*, *Sox21b*. For two copies of *Sox4a*, *Sox9a* and *Sox9b*, we follow the nomenclature of

Table 2 Percentage nucleic acid/amino acid similarity between the orthologous and paralogous

No.	*Tor douronensis* gene	Duplicated sequences	Zebrafish		Human
			a	b	
1	*Sox4ai*	90.1[a)]/98.1[b)]	96.9/100.0	84.6/98.1	85.8/100.0
2	*Sox4aii*		90.7/98.1	87.7/96.3	90.1/98.1
3	*Sox4b*	Vs. *Sox4ai*: 76.5/94.4, Vs. *Sox4aii*: 76.5/92.6	77.8/94.4	74.1/94.4	76.5/94.4
4	*Sox9ai*	84.6/100.0	78.4/87.0	75.3/88.9	74.1/87.0
5	*Sox9aii*		78.4/87.0	74.7/88.9	74.7/87.0
6	*Sox9bi*	84.0/100.0	74.1/87.0	78.4/88.9	72.8/87.0
7	*Sox9bii*		75.3/87.0	77.8/88.9	70.4/87.0
8	*Sox11b*	NC[c)]	85.2/100.0	95.1/100.0	84.6/100
9	*Sox14a*[d)]	Vs. *Sox14b*: 76.4/97.9	72.8/87.0	72.2/87.0	67.9/87
10	*Sox14b*[d)]		73.5/88.9	75.3/88.9	66.0/88.9
11	*Sox19b*	NC	79.6/98.1	88.9/98.1	NC
12	*Sox21a*	Vs. *Sox21b*: 75.3/96.3	88.9/100.0	76.5/96.3	75.9/96.3
13	*Sox21b*		78.0/96.3	95.1/100.0	80.2/100.0

a) Similarity of nucleic acid; b) similarity of amino acid; c) NC represents the similarity cannot be calculated; d) *Sox* gene of *Takifugu rubripes*.

Table 3 Comparison of substitutions rates of young duplicated *Sox* genes of *Tor douronensis*

Duplicated gene pairs	Calculated length (bp)	Number of synonymous substitution	Number of nonsynonymous substitution	*Ka*	*Ks*	*Ka/Ks*	*P* value of *Z* test
Sox4ai vs. *Sox4aii*	162	15	1	0.0087	0.4276	0.0203	0.0009[a)]
Sox9ai vs. *Sox9aii*	144	7	0	0.0000	0.1748	0.0000	0.0118
Sox9bi vs. *Sox9bi*	144	8	0	0.0000	0.1984	0.0000	0.0075

a) $P < 0.05$.

*H*ox genes in rainbow trout *Oncorhynchus mykiss*[37], arbitrarily designated different copies of these genes with a lowercase *i* or *ii*.

Considering that *Tor douronensis* is an evolutionary tetraploidy ($N = 4$, $2n = 100$)[20] and sequences of *Sox* genes HMG domain providing sufficient information to identify *Sox* gene[2], different isoforms of *Sox4a*, *Sox9a* and *Sox9b* should be resulted from tetraploidization, rather than allelic heterogeneity or genome segmental duplication[37,41]. To resolve this perplexity, one way is to map the chromosome location of these *Sox* genes on genetic map of *Tor douronensis*. For most of *Sox* genes obtained in this study, such as *Sox4b*, *Sox11b*, *Sox14a*, *Sox14b*, etc., only one copy is available. One explanation is that both copies of those *Sox* genes are sequence identical own to the short time since tetraploidization event[27]. An alternative possibility might be attribute to different copies of those *Sox* genes unbalanced amplification in PCR process due to primers, temperature, etc., which give rise to clones that contain different copies of those *Sox* genes are quantity bias. To obtain another *Sox* gene member in *Tor douronensis*, more clones should be sequenced. Besides, because some *Sox* genes are related to testis-determining in mammalian, we should make a survey of *Sox* genes in male *Tor douronensis* to get *Sox* genes, which may be significant for studying the role of *Sox* gene in the sex determination in teleost.

3.2 *Sox* gene diversity in *Tor douronensis*

The most obvious consequence of genome duplication is that it enriches important functional genes. In this study, we found two isoforms of *Sox4a*, *Sox9a* and *Sox9b* in *Tor douronensis*. Theoretically, Lynch et al.[38] summarized that fates of duplicated genes are as follows: (i) one copy may simply get silenced by degenerative mutations (nonfunctionalization); (ii) one copy may acquire a novel, beneficial function and be preserved by natural selection, with the other copy retaining the original function (neofunctionalization); (iii) both copies may be partially compromised by mutation accumulation to the point at which their total capacity is reduced to the level of the single-copy ancestral gene (subfunctionalization). By analyzing the genomes of several model organisms, Lynch et al.[38] found that most duplicated genes experienced a brief period of relaxed selection early in their history, with a moderate fraction of them evolving in an effective neutral manner during that period. Brunet et al.[26] held that in teleost fishes following duplication, there is an asymmetric acceleration of evolutionary rate in one of two paralogs. Result of PCR survey *Hox* genes indicated that pseudogene formation caused by deletions in tetraploid goldfish is more prevalent, which may be

the consequence of dosage effects as suggested by Luo et al.[39]. In our analysis, *Sox* genes in *Tor douronensis* don't show the evolutionary pattern as *Hox* genes in goldfish, although we just obtained partial of *Sox* gene sequenses in the HMG domain. Since most of the nucleotide differences between duplication (such as *Sox4a*) or their orthologues in other species are silent substitutions, it would appear that the sequences are under selective pressure, strongly suggesting that they represent functional genes. Our results seem not to support the neutral evolution hypothesis in the early stage after gene duplication, but are in agreement with previous works in frog *Xenopus laevis*[40] and *Zea perennis*[41]. It is interesting to note the extremely high variation of the intron length and similarities of two copies of *Sox9a* and *Sox9b*. Sequences analysis indicated that there was no mobile elements insertion and the sequence alignment scores were very low. Given the evolutionary rate of nucleotide and the speciation event of *Tor douronensis*, the possible explanation for this interesting phenomenon is that *Tor douronensis* may be an allopolyploidy, and the variation of intron length and lower sequence similarities might be existed in its ancestors. Of course, to get an unambiguous answer, more genes in *Tor douronensis* and its closely related tetraploid species in genus *Tor* should be investigated.

3.3 The evolution of *Sox* genes in vertebrates and fish-specific genome duplication

Interestingly in our analysis, one member of *Sox*B group, *Sox19b* has no corresponding gene in human, but its othologous are present in teleost fishes zebrafish[42] and pufferfish[5]. One possible explanation is that *Sox19* may represent an ancient vertebrate gene that was subsequently lost in the mammalian lineage after divergence with teleost fishes. However, we cannot rule out the alternative possibility that *Sox19* evolved through an ancient fish-specific random duplication of one *Sox* gene. To reveal the origin of this gene, some primary mammalian species should be investigated to see whether it exists in their genomes or not.

Our analysis shows that *Sox* genes of *Tor douronensis* (*Sox4, Sox9, Sox14* and *Sox21*) have 2 to 4 copies compared to mammals, and the pattern was also observed in other teleost fishes[4,5,43]. A large number of duplicated copies of single copy mammalian genes have been identified in teleost fishes. For example, in contrast to four clusters of *Hox* genes in mammals, several teleost fishes contain at least seven *Hox* clusters, respectively[12,13,44,45]. Comparative genomics studies have revealed that teleost fish experienced the so-called "fish-specific genome duplication" (FSGD), which caused duplicated copies of single copy mammalian genes in teleost fishes. In this study, the number of *Sox* genes in *Tor douronensis* and results of phylogenetic analysis together with other teleost *Sox* genes also support the fish-specific genome duplication hypothesis (Figure 3). Meyer et al.[46] proposed the one-to-four (-to-eight in teleost fishes) rule in vertebrates gene and genome duplications. For the polyploid fishes, the one-to-four-to-eight-to-sixteen-to-more rule may be more suitable, which makes polyploid fishes are ideal models for investigating the gene and genome evolution in the duplicated genomes. Compared to the *Hox* genes clusters, *Sox* genes are not present in clusters in any of the genomes characterized so far. Thus, *Sox* genes together with *Hox* genes clusters may be the excellent genetic markers for testing the fish-specific genome duplication hypothesis.

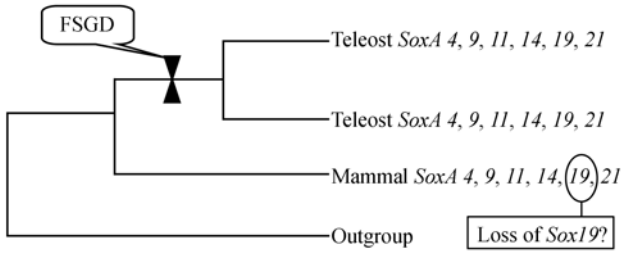

Figure 3 Evidence of fish-specific genome duplication (FSGD) inferred from *Sox* genes in *Tor douronensis* and other teleost fish.

We thank Ku Xiying for reading an early version of this manuscript.

1 Gubbay J, Collignon J, Koopman P, et al. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. Nature, 1990, 346(6281): 245－250
2 Bowles J, Schepers G, Koopman P. Phylogeny of the Sox family of developmental transcription factors based on sequence and structural indicators. Dev Biol, 2000, 227(2): 239－255
3 Schepers G E, Teasdale R D, Koopman P. Twenty pairs of *Sox*: Extent, homology, and nomenclature of the mouse and human *Sox* transcription factor gene families. Dev Cell, 2002, 3(2): 167－170
4 de Martino S, Yan Y L, Jowett T, et al. Expression of *Sox11* gene duplicates in zebrafish suggests the reciprocal loss of ancestral gene expression patterns in development. Dev Dyn, 2000, 217(3): 279－292
5 Koopman P, Schepers G, Brenner S, et al. Origin and diversity of the *Sox* transcription factor gene family: genome-wide analysis in *Fugu*

Rubripes. Gene, 2004, 328: 177－186
6 Ohno S. Evolution by Gene Duplication. New York: Springer-Verlag, 1970
7 Wolfe K H. Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet, 2001, 2(5): 333－341
8 Comai L. The advantages and disadvantages of being polyploidy. Nat Rev Genet, 2005, 6(11): 836－846
9 Holland P W, Garcia-Fernandez J, Williams N A, et al. Gene duplications and the origins of vertebrate development. Development, 1994, Suppl: 125－133
10 Skrabanek L, Wolfe K H. Eukaryote genome duplication－Where's the evidence? Curr Opin Genet Dev, 1998, 8(6): 694－700
11 Hughes A L, Friedman R. 2r or Not 2r: Testing hypotheses of genome duplication in early vertebrates. J Struct Funct Genom, 2003, 3(1-4): 85－93
12 Amores A, Force A, Yan Y L, et al. Zebrafish Hox clusters and vertebrate genome evolution. Science, 1998, 282(5394): 1711－1714
13 Christoffels A, Koh E G, Chia J M, et al. Fugu genome analysis provides evidence for a whole-Genome duplication early during the evolution of ray-finned fishes. Mol Biol Evol, 2004, 21(6): 1146－1151
14 Hoegg S, Brinkmann H, Taylor J S, et al. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost Fish. J Mol Evol, 2004, 59(2): 190－203
15 Vandepoele K, De Vos W, Taylor J S, et al. Major events in the genome evolution of vertebrates: Paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc Natl Acad Sci USA, 2004, 101(6): 1638－1643
16 Crow K D, Stadler P F, Lynch V J, et al. The "fish-specific" Hox cluster duplication is coincident with the origin of teleosts. Mol Biol Evol, 2006, 23(1): 121－136
17 Le Comber S C, Smith C. Polyploidy in fishes: Patterns and processes. Biol J Linnean Soc, 2004, 82(4): 431－442
18 Allendorf F W, Thorgaard G H. Tetraploidy and the evolution of salmonid fish. In: Turner J B, ed. Evolutionary Genetics of Fish. New York: Plenum Press, 1984. 1－53
19 Xiao H, Zhang R D, Feng J G, et al., Nuclear DNA content and ploidy of seventeen species of fishes in *Sinocyclocheilus*. Zoolog Res (in Chinese), 2002, 23(3): 195－199
20 Zan R, Song Z, Liu W. Studies on karyotypes and DNA contents of some Cyprinoid fishes, with notes on fish polyploids in China. In: Uyeno T, Arai R, Taniuchi T, et al, eds. Indo Pacific Fish Biology. Tokyo:Ichthyol Soc Japan, 1986. 877－885
21 Yu X J, Zhou T, Li Y C, et al. Chromosomes of Chinese Fresh-Water Fishes (in Chinese). Beijing: Science Press, 1989
22 Kellis M, Birren B W, Lander E S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature, 2004, 428(6983): 617－624
23 Maere S, de Bodt S, Raes J, et al. Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci USA, 2005, 102(15): 5454－5459
24 Moore R C, Purugganan M D. The early stages of duplicate gene evolution. Proc Natl Acad Sci USA, 2003, 100(26): 15682－15687
25 Akhunov E D, Akhunova A R, Dvorak J. Mechanisms and rates of birth and death of dispersed duplicated genes during the evolution of a multigene family in diploid and tetraploid wheats. Mol Biol Evol, 2007, 24(2): 539－550
26 Brunet F G, Crollius H R, Paris M, et al. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. Mol Biol Evol, 2006, 23(9): 1808－1816
27 Wang X Z, Li J B, 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, 2007, 42(1): 157－170
28 Chen Y Y. Fauna Sinica, Class Teleostei, Cypriniformes II (in Chinese). Beijing: Science Press, 1998
29 Sambrook J, Fritsch E F, Maniatis T. Molecular Cloning: A Laboratory Manual, 2nd ed. New York: Cold Spring Harbor Laboratory Press, 1989
30 Galay-Burgos M, Llewellyn L, Mylonas C C, et al. Analysis of the *Sox* Gene family in the european sea bass(*Dicentrarchus Labrax*). Comp Biochem Physiol B Biochem Mol Biol, 2004, 137(2): 279－284
31 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
32 Tamura K, Dudley J, Nei M, et al. Mega 4: molecular evolutionary genetics analysis (Mega) software version 4.0. Mol Biol Evol, 2007, 24(8): 1596－1599
33 Ronquist F, Huelsenbeck J P. Mrbayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 2003, 19(12): 1572－1574
34 Posada D, Crandall K A. Modeltest: Testing the model of DNA substitution. Bioinformatics, 1998, 14(9): 817－818
35 Campanella J, Bitincka L, Smalley J. Matgat: An application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics, 2003, 4(1): 29
36 Zhang J, Rosenberg H F, Nei M. Positive darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci USA, 1998, 95(7): 3708－3713
37 Moghadam H K, Ferguson M M, Danzmann R G. Evidence for *Hox* gene duplication in rainbow trout (*Oncorhynchus Mykiss*): A tetraploid model species. J Mol Evol, 2005, 61(6): 804－818
38 Lynch M, Conery J S. The evolutionary fate and consequences of duplicate genes. Science, 2000, 290(5494): 1151－1155
39 Luo J, Stadler P F, He S P, et al. PCR survey of *Hox* genes in the goldfish *Carassius auratus auratus*. J Exp Zoolog B Mol Dev Evol, 2007, 308(3): 250－258
40 Hughes M K, Hughes A L. Evolution of duplicate genes in a tetraploid animal, *Xenopus Laevis*. Mol Biol Evol, 1993, 10(6): 1360－1369
41 Tiffin P, Gaut B S. Sequence diversity in the tetraploid *Zea Perennis* and the closely related diploid *Z. diploperennis*: insights from four nuclear loci. Genetics, 2001, 158(1): 401－412
42 Vriz S, Lovell-Badge R. The zebrafish Zf-Sox 19 protein: a novel member of the Sox family which reveals highly conserved motifs outside of the DNA-binding domain. Gene, 1995, 153(2): 275－276
43 Chiang E F, Pai C I, Wyatt M, et al. Two *Sox9* genes on duplicated zebrafish chromosomes: expression of similar transcription activators in distinct sites. Dev Biol, 2001, 231(1): 149－63
44 Kurosawa G, Yamada K, Ishiguro H, et al. *Hox* gene complexity in medaka fish may be similar to that in pufferfish rather than zebrafish. Biochem Biophys Res Commun, 1999, 260(1): 66－70
45 Kurosawa G, Takamatsu N, Takahashi M, et al. Organization and structure of *Hox* gene loci in medaka genome and comparison with those of pufferfish and zebrafish genomes. Gene, 2006, 370: 75－82
46 Meyer A, and Schartl M. Gene and genome duplications in vertebrates: the one-to-four(-to-eight in fish)rule and the evolution of novel gene functions. Curr Opin Cell Biol, 1999, 11(6): 699－704