

## ***Sox* genes in grass carp (*Ctenopharyngodon idella*) with their implications for genome duplication and evolution**

Lei ZHONG<sup>a,b</sup>, Xiaomu YU<sup>a</sup>, Jingou TONG<sup>a,\*</sup>

<sup>a</sup> State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, the Chinese Academy of Sciences, Wuhan 430072, P. R. China

<sup>b</sup> The Graduate School of the Chinese Academy of Sciences, Beijing 100039, P. R. China

(Received 7 February 2006; accepted 12 June 2006)

**Abstract** – The *Sox* gene family is found in a broad range of animal taxa and encodes important gene regulatory proteins involved in a variety of developmental processes. We have obtained clones representing the HMG boxes of twelve *Sox* genes from grass carp (*Ctenopharyngodon idella*), one of the four major domestic carps in China. The cloned *Sox* genes belong to group B1, B2 and C. Our analyses show that whereas the human genome contains a single copy of *Sox4*, *Sox11* and *Sox14*, each of these genes has two co-orthologs in grass carp, and the duplication of *Sox4* and *Sox11* occurred before the divergence of grass carp and zebrafish, which support the “fish-specific whole-genome duplication” theory. An estimation for the origin of grass carp based on the molecular clock using *Sox1*, *Sox3* and *Sox11* genes as markers indicates that grass carp (subfamily *Leuciscinae*) and zebrafish (subfamily *Danioninae*) diverged approximately 60 million years ago. The potential uses of *Sox* genes as markers in revealing the evolutionary history of grass carp are discussed.

**grass carp (*Ctenopharyngodon idella*) / *Sox* / genome duplication / co-ortholog / molecular clock**

### **1. INTRODUCTION**

The *Sox* (*SRY*-related genes containing an HMG box; HMG, high mobility group) gene family was first identified in 1990 as a group of genes related to the mammalian testis determining factor *Sry* based on conservation of the single HMG box, which encodes a 79-amino acid DNA-binding HMG domain [10]. The number of known *Sox* genes has been expanded through homology-based screening approaches recently [4, 7, 42]. The roles of some *Sox* proteins have

---

\* Corresponding author: jgtong@ihb.ac.cn

been revealed as important developmental regulators in a variety of developmental processes, *e.g.* during sex determination and the development of heart and CNS (central nervous system) [2, 35].

For all Sox proteins, the HMG domains, outside of which Sox sequences are highly variable, are highly conserved in primary structure, and all appear to be capable of binding to the same target DNA sequence [39]. Specificity in target selection is considered to be brought about *via* a combinatorial mechanism involving interaction with other tissue-specific transcription factors and spatio-temporal expression patterns [15, 16, 40]. A total of 20 Sox genes has now been identified in the mouse and human by whole-genome sequence analyses [35], and primary sequence comparison and other structural indicators such as intron-exon organization indicate that these genes fall into eight clear groups, A–H [2].

Sox genes have been identified in a broad range of animal taxa, including mammals, birds, reptiles, amphibians, fish, insects and nematodes. Among vertebrates, orthologs in different species are highly similar to each other. Most of these groups are represented by a single gene in the invertebrate model organisms *Drosophila melanogaster* and *Caenorhabditis elegans*, suggesting the occurrence of expansion of this single gene into multiple related genes during vertebrate evolution [2]. Since the divergence of regulatory genes is being considered necessary to bring about phenotypic variation and an increase in biological complexity, it is proposed that such duplication events have indeed been of major importance for evolution in vertebrates [25, 30, 31]. It is not yet clear whether this expansion is the result of numerous rounds of independent gene duplications or the two-round (three-round to fish) genome duplications during the evolution of vertebrates [38]. Analyses of additional vertebrate genomes are required to refine the understanding of molecular evolution of this gene family.

However, the organization and function of the Sox gene family are far less well understood in vertebrates than in mammals. Surprisingly, there have been limited studies on Sox genes in teleost fishes, which comprise half of the vertebrate species [3, 9, 17, 44] and whose development mechanisms may be related but markedly different to those employed by mammals [8]. In the present study, we cloned 12 members of the Sox gene family in grass carp (*Ctenopharyngodon idella*), one of the most important herbivorous fishes in the world and a species of particular significance in fisheries and aquaculture in China. The phylogenetic analyses and their implications for fish genome duplication and evolution were also exploited. The major aim of this study was to extend our understanding on the organization and evolution of such important regulatory

genes as the *Sox* transcription factor gene family in a teleost fish with economic potentials.

## 2. MATERIALS AND METHODS

### 2.1. Amplification, cloning and sequencing of *Sox* genes

*Sox* genes were amplified by PCR using a pair of degenerate primers designated as *SoxX* [9]. The *SoxX* primers (ATGAAYGCNTTYATGGTNTGG and GGNCGRTAYTTRTARTCNGG) correspond to the motifs MNAFMVW and PDYKYRP, which are found in the HMG boxes of almost all *Sox* proteins that belong to groups B and C. As template in PCR amplification, genomic DNA was extracted from fresh fin tissues of grass carp using the traditional phenol-chloroform method. The PCR reactions contained 50 pmol of each primer, 0.1  $\mu$ g genomic DNA, 200  $\mu$ M dNTPs, 1.5 mM MgCl<sub>2</sub> and 2.5 units of *Taq* DNA polymerase in a 25  $\mu$ L reaction mix. PCR amplifications were performed for 35 successive cycles of 95 °C for 1 min, 54 °C for 1 min, and 72 °C for 1 min. The PCR products were resolved on 1.5% agarose gels, and the expected bands were excised and gel-purified. Purified products were then sub-cloned into the vector pMD-18-T. The positive clones were sequenced on an ABI 3730 capillary sequencer.

### 2.2. Sequence analysis and phylogenetic construction

Sequence identities were assigned following combined analyses including BlastX searches of GenBank and the human genome sequences, searches of signature residues from putative amino acid sequences [17] and phylogenetic clustering analyses using software MEGA3.1 [18]. The relationships between the proteins encoded by the grass carp *Sox* genes and the corresponding proteins in humans (and zebrafish in the case of *Sox21*) were analyzed using the Minimum-Evolution method [34] with the Dayhoff Matrix Model [5]. In order to estimate the relative age of the paralogs of duplicate *Sox* genes in grass carp, the number of nucleotide substitutions at third codon positions of *Sox4* and *Sox11* and also *Sox1*, *Sox2* and *Sox3* were used to construct a linearized tree by the Minimum-Evolution method with the p-distance model using third-codon position substitutions of the corresponding nucleotide sequences and 1000 bootstrap replicates were performed. For the estimation of the divergence time between grass carp and zebrafish, a linearized tree of *Sox1*, *Sox3* and *Sox11* was constructed by using third-codon position substitutions of the

corresponding nucleotide sequences and Minimum-Evolution method with the p-distance model, and 1000 bootstrap replicates were performed. Another linearized tree of concatenated dataset of *Sox1a*, *Sox3* and *Sox11a* was also built using synonymous substitutions by the UPGMA method [36] with the modified Nei-Gojobori (p-distance) model, transition/transversion ratio of 2.

### 3. RESULTS

#### 3.1. Assignment of orthology and nomenclature of the grass carp *Sox* genes

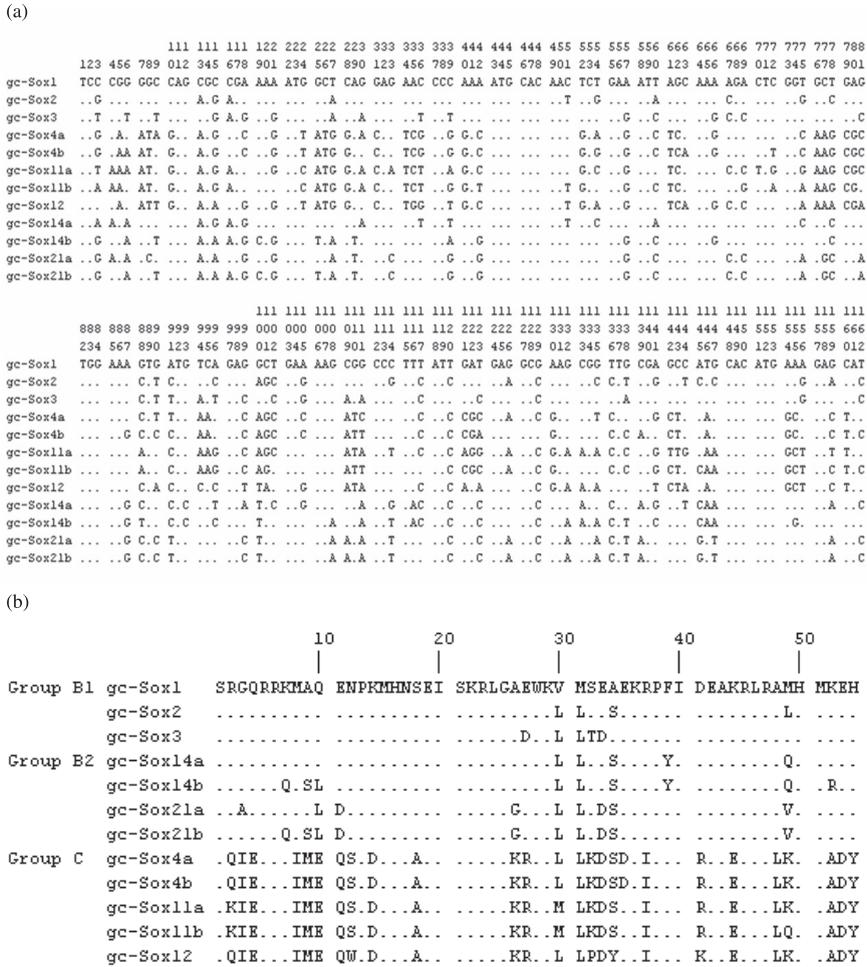
Electrophoresis of PCR products generated from genomic DNA using the *SoxX* primers showed a single band at a size of about 200 bp as expected for intron-free *Sox* genes. This band was gel-purified and subcloned, and 36 individual clones were sequenced. Twelve different *Sox* genes (*Sox1*, *Sox2*, *Sox3*, *Sox4a*, *Sox4b*, *Sox11a*, *Sox11b*, *Sox12*, *Sox14a*, *Sox14b*, *Sox21a*, *Sox21b*) were obtained in grass carp (GenBank accession numbers DQ642604-DQ642615), and the alignment results of these genes are shown in Figure 1.

Each of these genes was represented in at least two independent clones except for *Sox11b* and *Sox21b*, and the number of the variable sites among these *Sox* sequences is too large to result from PCR errors, making it unlikely that our data would be affected by PCR artifacts. Since our combined analyses led to unambiguous gene assignment, we refer to the grass carp *Sox* genes by their nomenclature proposed in this paper.

The relationships between the proteins encoded by grass carp *Sox* genes and the corresponding proteins from human or zebrafish were analyzed, and the phylogenetic tree is shown in Figure 2. The clustering of grass carp and human sequences provides strong support for the gene assignments given to the grass carp sequences. All cloned *Sox* genes have mammalian orthologs. The human genome contains a single copy of *Sox4*, *Sox11*, *Sox14* and *Sox21*, whereas the grass carp holds two copies of the corresponding genes.

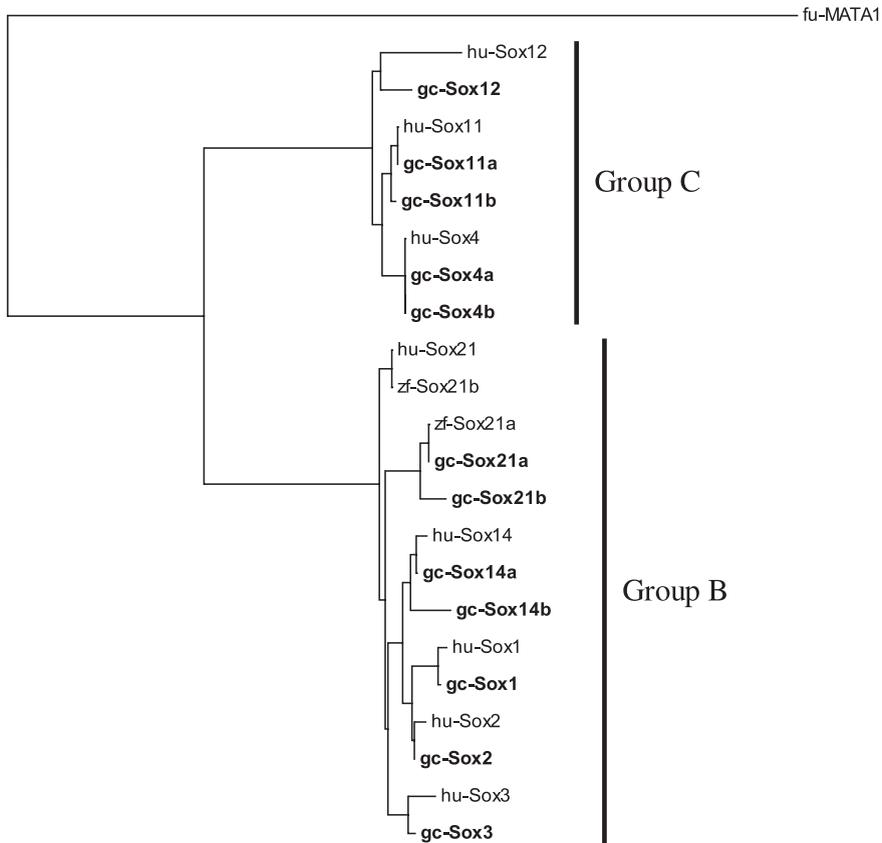
#### 3.2. Mapping duplication events onto phylogeny

Phylogenetic analysis shows that at least three of the ancestral vertebrate *Sox* genes (*Sox4*, *Sox11*, *Sox14*) are duplicated in grass carp (Fig. 2). Duplicates of *Sox4* and *Sox11* have also been identified in the zebrafish [6, 26]. In order to estimate the age of the paralogs, the number of nucleotide substitutions at third codon positions was used to construct a linearized tree from the data of *Sox4*,



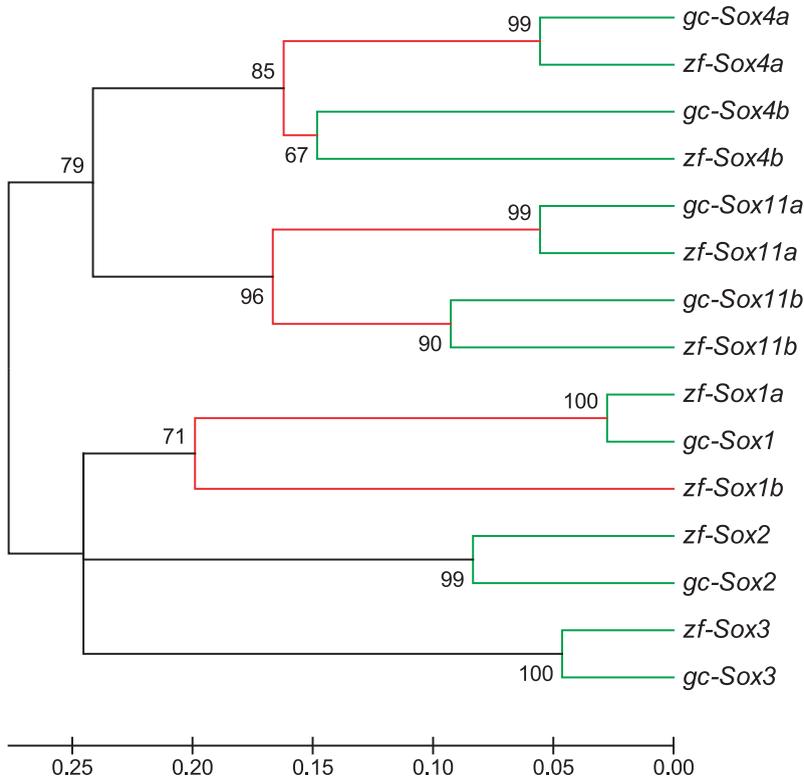
**Figure 1.** Alignment of the nucleotide sequences (a) and putative amino acid sequences (b) encoded by the HMG domains of cloned *Sox* genes in grass carp. Residues identical to those of *gc-Sox1* are indicated by dots (.). The amino acid sequences are ordered and shaded according to the group they occupy within the *Sox* family [2]. The sequences shown have been deposited in GenBank under the accession numbers DQ642604, DQ642605, DQ642606, DQ642607, DQ642608, DQ642609, DQ642610, DQ642611, DQ642612, DQ642613, DQ642614 and DQ642615.

*Sox11*, *Sox1*, *Sox2* and *Sox3* in grass carp and zebrafish since most third-codon position substitutions do not result in amino-acid replacements, the rate of fixation of these substitutions is expected to be relatively constant in different protein-coding genes (e.g. [29]) and to reflect the overall mutation rate [14].



**Figure 2.** Rooted, HMG domain phylogeny of grass carp and human Sox proteins. A distantly related HMG-box protein, yeast MATA-1 (GenBank accession number P36981) was used as the outgroup. fu, fungi, *Saccharomyces cerevisiae*; gc, grass carp, *Ctenopharyngodon idella*; hu, human, *Homo sapiens*; zf, zebrafish, *Danio rerio*. The GenBank accession numbers for the human *Sox* genes are: *Sox1*, NM\_005986; *Sox2*, BC013923; *Sox3*, NM\_005634; *Sox4*, NM\_003107; *Sox11*, NM\_003108; *Sox12*, NM\_006943; *Sox14*, NM\_004189; *Sox21*, NM\_007084, and for zebrafish *Sox* proteins: *Sox21a*, AAH56274 ; *Sox21b*, AAH95366.

The linearized tree (Fig. 3), with high bootstrap values, clearly demonstrated that the duplication of *Sox4* and *Sox11* occurred before the divergence of grass carp (subfamily *Leuciscinae*) and zebrafish (subfamily *Danioninae*), which belong to the different subfamilies of Cyprinidae of the order Cypriniformes, and are likely within the same time range. Although two copies of *Sox1* were reported in zebrafish, only one ortholog was cloned in grass carp in the present study. This may result from inadequate sampling and sequencing of the clones

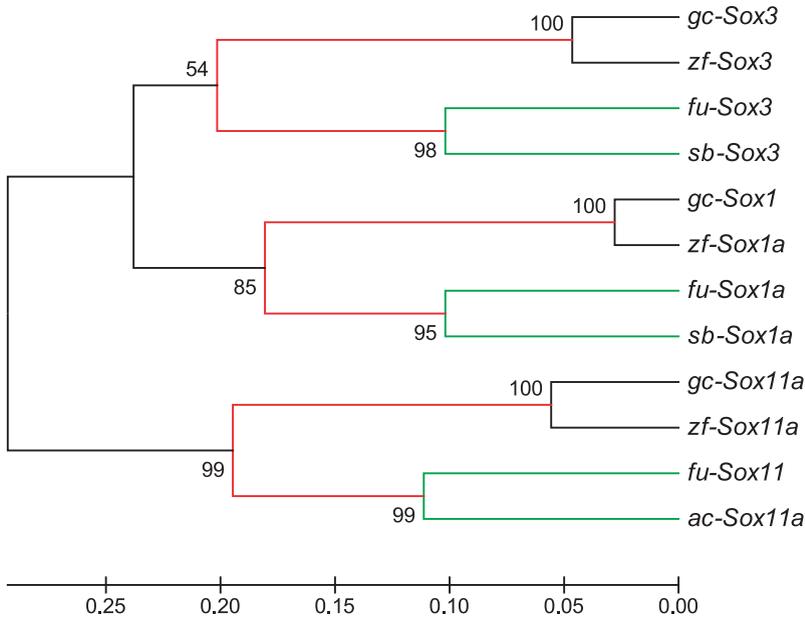


**Figure 3.** Linearized tree of *Sox4*, *Sox11* (Group C) and *Sox1*, *Sox2*, *Sox3* (Group B) for grass carp (gc-) and zebrafish (zf-). The numbers above or below the branches indicate bootstrap values from Minimum Evolution. The GenBank accession numbers for zebrafish *Sox* genes: AB242327, AB242328, NM\_213118, NM\_001001811, NM\_213122, NM\_200901, NM\_131336, NM\_131337.

or, with less probability, that the other copy of the *Sox1* gene was lost in the grass carp after the divergence of these two species.

### 3.3. Dating the divergence of grass carp and zebrafish

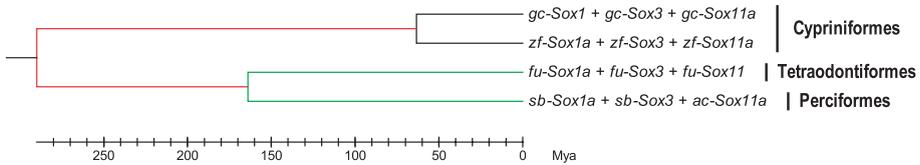
The number of synonymous substitutions per synonymous sites can be used to estimate divergence times based on the molecular clock model [28,29]. We constructed a linearized tree based on third codon positions of *Sox1*, *Sox3* and *Sox11* of grass carp, zebrafish, *Takifugu*, sea bass and *Amphilophus citrinellum*, respectively. It shows a fine phylogenetic resolution (Fig. 4) which is consistent with the previously reported phylogeny of these fish species. Furthermore, the



**Figure 4.** Linearized tree of *Sox1*, *Sox3* and *Sox11* using third-codon position substitutions of the nucleotide sequences. The numbers above or below the branches indicate bootstrap values from Minimum Evolution. gc, grass carp; zf, zebrafish (AB242327, NM\_001001811, NM\_131336); fu, *Takifugu rubripes* (AY277950, AY277953, AY277959); sb, sea bass, *Dicentrarchus labrax* (AY246990, AY246993); ac, *Amphilophus citrinellum* (AY333971).

patterns of the subtrees are nearly the same, which represent *Sox1*, *Sox3* and *Sox11*, respectively.

With a combined larger dataset, improved phylogenetic resolution is expected [20]. Combined datasets were concatenated, based on the assumption that the observed divergence for *Sox* genes corresponds to speciation events. Due to limited species availability, we combined sequences from different species belonging to the same larger fish taxon. Thus *Sox1a* and *Sox3* of sea bass were combined with *Sox11a* of *Amphilophus citrinellum* (Fig. 5) since these two species both belong to the Perciformes. The date of divergence of zebrafish and *Takifugu*, set at the root of the tree, was taken as  $\sim 290 (\pm 6)$  Mya (million years ago) from literature based on calibration from molecular data [19]. The result shows that the divergence time between grass carp and zebrafish is  $\sim 63 (\pm 2)$  Mya.



**Figure 5.** Linearized tree of concatenated dataset of *Sox1*, *Sox3* and *Sox11* using the number of synonymous substitutions per synonymous sites of the nucleotide sequences for grass carp (gc-), zebrafish (zf-), *Takiugu* (fu-), sea bass (sb-) and *Amphiphus citrinellum* (ac-). Mya, million years ago.

## 4. DISCUSSION

### 4.1. New evidence for the “fish-specific whole-genome duplication” theory

In total, 12 grass carp *Sox* genes including members of the *SoxB1*, *SoxB2* and *SoxC* groups [2] have been cloned and identified (Fig. 1). This is the first report describing *Sox* genes in the subfamily *Leuciscinae* and the data presented constitute one of the most complete analyses of the *Sox* gene family in fish to date.

The sequences described in this study are well conserved at the amino acid level and all the cloned *Sox* genes have mammalian orthologs. Several clones obtained in this study represent genes that are duplicated in grass carp with respect to the mammalian *Sox* gene family. Gene duplication appears to be very common in fish, although in many situations it has not yet been determined whether this is due to numerous independent segmental duplications or an ancestral whole-genome duplication [38].

Whereas the human genome contains a single copy of *Sox4*, *Sox11* and *Sox14*, each of these genes have two co-orthologs in grass carp (Fig. 2). The clustering of grass carp and human sequences illustrates the gene duplications that have occurred in grass carp, although nearly all duplicate genes showed an “outgroup” topology instead of the “sister-gene” topology expected for the duplicate genes. Occasionally, duplicate genes show such an “outgroup” topology because of the accelerated evolution rate of one copy of duplicated genes, because of its release from previous selective pressure. Evolutionary rates of two paralogs often differ enormously; usually one of the paralogs evolves considerably faster than the other one [37]. Exhaustive searches of orthologs in other fish from the literature and GenBank indicate that some species also have duplicates of certain *Sox* genes, especially *Sox1*, *Sox4*, *Sox11* and *Sox21* in zebrafish, *Sox1*, *Sox4*, *Sox14* in sea bass (Perciformes) [9] and *Sox1*, *Sox14* in

*Takifugu* (Tetraodontiformes) [17]. This suggests the occurrence of a large-scale duplication or whole-genome duplication event relatively early during teleost evolution with regards to *Sox* genes since these genes are distributed on different chromosomes in all genomes characterized so far [35]. Gene silencing and subsequent loss could happen within a short time after a gene duplication event [21–24], so the failure to detect duplicates for other *Sox* genes, *e.g.* *Sox3*, may be explained by the subsequent gene loss after duplication, considering the general high rate of loss of duplicate gene copies in vertebrates (~67% over ~500 million years [27]) and the time of the assumed fish-specific genome duplication event, which has been dated to 335–404 Mya [13].

If the fish-specific genome duplication before the teleost radiation [13, 38] is a real historical event, and the duplication of certain *Sox* genes in fish is the consequence of this event, then the divergence of the two copies of the duplicated *Sox* genes must have occurred before the divergence of grass carp and zebrafish, which both belong to Cyprinidae of the Cypriniformes. Our analyses using the number of nucleotide substitutions at third codon positions (Fig. 3) showed that the duplication of *Sox4* and *Sox11* took place before the divergence of grass carp and zebrafish, therefore further supporting the “fish-specific whole-genome duplication” theory.

The present data unexpectedly showed abnormal tree topology with regards to *gc-Sox21b*. It seems that *gc-Sox21b* is neither the ortholog of *hu-Sox21* nor *zf-Sox21b*, but the co-ortholog of *zf-Sox21a* (Fig. 1). Considering that the number of variation sites of the obtained nucleotide sequences between *gc-Sox21a* and *gc-Sox21b* is only 6, which is much less than that between other couples of duplicated *Sox* genes, *e.g.* 35 between *gc-Sox14a* and *gc-Sox14b*, the age of this gene, *gc-Sox21b*, must be relatively young and probably the consequence of a much recent segmental duplication rather than the ancient fish-specific whole-genome duplication event. Furthermore, among nucleotide differences between *gc-Sox21a* and *gc-Sox21b*, half of the substitutions are nonsynonymous substitutions in contrast to that almost all the substitutions are synonymous between other paralogs of duplicate *Sox* genes, indicating that the additional copy of *Sox21*, *gc-Sox21b*, is still in the period of relaxed selection experienced by most duplicated genes in their early history [24], again suggesting its recent origin.

#### 4.2. Duplicates of *Sox*: meaning for subfunction partitioning

After gene or genome duplication, each gene copy may follow a separate evolutionary trajectory. New gene duplicates face one of

three fates, *i.e.* non-functionalization, neo-functionalization, and sub-functionalization [31,32]. Since most of the nucleotide differences between duplicates of *gc-Sox* genes represent synonymous substitutions not altering the encoded amino acid, it would appear that the sequences are under considerable selective pressure, strongly suggesting their taking on function, and excluding the possibility that the duplicates are pseudogenes.

While the origin of a new function appears to be a very rare fate for a duplicate gene [24], subfunction partitioning is relatively common among duplicated genes arising from the fish-specific genome duplication event [31]. Previous studies have demonstrated that *Sox11a* and *Sox11b* may share the developmental domains of the single *Sox11* gene present in mice and chickens for their expressing in both overlapping and distinct sites [6].

The partitioning of ancestral subfunctions between gene copies arising from the whole-genome duplication could have contributed to the speciation and radiation of teleost fish. Beyond its importance for understanding mechanisms generating biodiversity, the partitioning of subfunctions between teleost co-orthologs of human genes can facilitate the identification of tissue-specific conserved noncoding regions, and can simplify the analysis of ancestral gene functions obscured by pleiotropy or haploinsufficiency [1, 31, 41]. It is known that the functions of some *Sox* genes in fish are related to those in humans in the key developmental process such as development of CNS and cartilage formation [6, 43]. Further studies on the partitioning of subfunctions between the *Sox* gene co-orthologs in grass carp will be worthwhile. Anyhow, the *Sox* genes presented in this study were proved to be good candidates for studies on the structure, function and evolution of genes and genomes in vertebrates.

#### **4.3. Estimation for the origin of grass carp from the *Sox* molecular clock**

The grass carp is one of the four major domestic carps in China, or so called “Chinese carps”, of which the origin is yet unclear, thus the historical information based on the calibration from molecular data of nuclear genes will be valuable. We constructed linearized trees based on the molecular clock theory using the nuclear genes *Sox1*, *Sox3* and *Sox11* as markers, which are located on different chromosomes thus representing independent unlinked loci in all of the genomes characterized so far. Our result (Figs. 4 and 5) indicates that grass carp (subfamily *Leuciscinae*) and zebrafish (subfamily *Danioninae*) shared a last common ancestor approximately 63 ( $\pm 2$ ) Mya, which falls in the Paleocene (57–65 Mya). This result shows agreement with the fossil record of fish

of *Leuciscinae* in the Oligocene (23–35 Mya) [33] and the previous notion that the fish of *Leuciscinae* originate from the ancient species of the original sub-family *Danioninae* in the early Paleogene [11]. It is known that the speciation of grass carp was related to the uplift of the Tibetan plateau, which began from the Neogene (1.6–23 Mya) [12]. Further studies on *Sox* genes of other cyprinid fish species with closer phylogenetic relationships with grass carp, *e.g.* silver carp (*Hypophthalmichthys molitrix*), will be informative in elucidating the accurate evolutionary history of grass carp.

### ACKNOWLEDGEMENTS

The authors thank X. Liao and L. Cheng for their technical assistance. This work was supported by the National Key Basic Researches (2004CB117405), National Natural Science Foundation of China (30271011, 30370225), and Natural Science Foundation of Hubei Province (2004ABC005).

### REFERENCES

- [1] Amores A., Force A., Yan Y.-L., Joly L., Amemiya C., Fritz A., Ho R.K., Langeland J., Prince V., Wang Y.-L., Westerfield M., Ekker M., Postlethwait J.H., Zebrafish *hox* clusters and vertebrate genome evolution, *Science* 282 (1998) 1711–1714.
- [2] Bowles J., Schepers G., Koopman P., Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators, *Dev. Biol.* 227 (2000) 239–255.
- [3] Chiang E.F.-L., Pai C.-I., Wyatt M., Yan Y.-L., Postlethwait J., Chung B.-C., Two *Sox9* genes on duplicated zebrafish chromosomes: expression of similar transcription activators in distinct sites, *Dev. Biol.* 231 (2001) 149–163.
- [4] Crémazy F., Soullier S., Berta P., Jay P., Further complexity of the human *Sox* gene family revealed by the combined use of highly degenerate primers and nested PCR, *FEBS Lett.* 438 (1998) 311–314.
- [5] Dayhoof M.O., Schiowitz R.M., Orcutt B.C., A model of evolutionary change in proteins, in: Dayhoof M.O. (Ed.), *Atlas of protein sequence and structure*, National Biomedical Research Foundation, Silver Spring, MD, 1978, pp. 345–352.
- [6] De Martino S., Yan Y.-L., Jowett T., Postlethwait J.H., Varga Z.M., Ashworth A., Austin C.A., Expression of *Sox11* gene duplicates in zebrafish suggests the reciprocal loss of ancestral gene expression patterns in development, *Dev. Dynam.* 217 (2000) 279–292.
- [7] Denny P., Swift S., Brand N., Dabhade N., Barton P., Ashworth A., A conserved family of genes related to the testis determining gene, *SRY*, *Nucleic Acids Res.* 20 (1992) 2887.

- [8] Devlin R.H., Nagahama Y., Sex determination and sex differentiation in fish: an overview of genetic, physiological and environmental influences, *Aquaculture* 208 (2002) 191–364.
- [9] Galay-Burgos M., Llewellyn L., Mylonas C.C., Canario A.V.M., Zanuy S., Sweeney G.E. Analysis of the *Sox* gene family in the European sea bass (*Dicentrarchus labrax*), *Comp. Biochem. Phys. B* 137 (2004) 279–284.
- [10] Gubbay J., Collignon J., Koopman P., Capel B., Economou A., Münsterberg A., Vivian N., Goodfellow P., Lovell-Badge R., 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* 346 (1990) 245–250.
- [11] He S., Chen Y., Nakajima T., Sequences of cytochrome b gene for primitive cyprinid fishes in East Asia and their phylogenetic concerning, *Chinese Sci. Bull.* 46 (2001) 661–665.
- [12] He S., Liu H., Chen Y., Kuwahara M., Nakajima T., Zhong Y., Molecular phylogenetic relationships of Eastern Asian Cyprinidae (Pisces: Cypriniformes) inferred from cytochrome b sequences, *Sci. China Ser. C* 47 (2004) 130–138.
- [13] Hoegg S., Brinkmann H., Taylor J.S., Meyer A., Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish, *J. Mol. Evol.* 59 (2004) 190–203.
- [14] Hughes A.L., Adaptive evolution of genes and genomes, Oxford University Press, New York, 1999.
- [15] Kamachi Y., Cheah K.S.E., Kondoh H., Mechanism of regulatory target selection by the SOX high-mobility-group domain proteins as revealed by comparison of SOX1/2/3 and SOX9, *Mol. Cell. Biol.* 19 (1999) 107–120.
- [16] Kamachi Y., Uchikawa M., Kondoh H., Pairing SOX off: with partners in the regulation of embryonic development, *Trends Genet.* 16 (2000) 182–187.
- [17] Koopman P., Schepers G., Brenner S., Venkatesh B., Origin and diversity of the *Sox* transcription factor gene family: genome-wide analysis in *Fugu rubripes*, *Gene* 328 (2004) 177–186.
- [18] Kumar S., Tamura K., Nei M., MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment, *Brief. Bioinformatics* 5 (2004) 150–163.
- [19] Kumazawa Y., Yamaguchi M., Nishida M., Mitochondrial molecular clocks and the origin of euteleostean biodiversity: Familial radiation of perciforms may have predated the Cretaceous/Tertiary boundary, In: Kato M. (Ed.), *The Biology of Biodiversity*, Springer-Verlag, Hong Kong, 1999, pp. 35–52.
- [20] Lecointre G., Philippe H., V&agrave;n L&eacute; H.L., Le Guyader H., How many nucleotides are required to resolve a phylogenetic problem? The use of a new statistical method applicable to available sequences, *Mol. Phylogenet. Evol.* 3 (1994) 292–309.
- [21] Li W.-H., Rate of gene silencing at duplicate loci: A theoretical study and interpretation of data from tetraploid fishes, *Genetics* 95 (1980) 237–258.
- [22] Lynch M., Gene duplication and evolution, *Science* 297 (2002) 945–947.
- [23] Lynch M., Conery J.S., The evolutionary fate and consequences of duplicate genes, *Science* 290 (2000) 1151–1155.

- [24] Lynch M., Conery J.S., The evolutionary demography of duplicate genes, *J. Struct. Funct. Genomics* 3 (2003) 35–44.
- [25] Maere S., De Bodt S., Raes J., Casneuf T., Van Montagu M., Kuiper M., Van de Peer Y., Modeling gene and genome duplications in eukaryotes, *Proc. Natl. Acad. Sci. USA* 102 (2005) 5454–5459.
- [26] Mavropoulos A., Devos N., Biemar F., Zecchin E., Argenton F., Edlund H., Motte P., Martial J.A., Peers B., Sox4b is a key player of pancreatic alpha cell differentiation in zebrafish, *Dev. Biol.* 285 (2005) 211–223.
- [27] Nadeau J.H., Sankoff D., Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution, *Genetics* 147 (1997) 1259–1266.
- [28] Nei M., Kumar S., *Molecular evolution and phylogenetics*, Oxford University Press, New York, 2000.
- [29] Nei M., Rogozin I.B., Piontkivska H., Purifying selection and birth-and-death evolution in the ubiquitin gene family, *Proc. Natl. Acad. Sci. USA* 97 (2000) 10866–10871.
- [30] Otto S.P., Whitton J., Polyploid incidence and evolution, *Ann. Rev. Genet.* 34 (2000) 401–437.
- [31] Postlethwait J., Amores A., Cresko W., Singer A., Yan Y.-L., Subfunction partitioning, the teleost radiation and the annotation of the human genome, *Trends Genet.* 20 (2004) 481–490.
- [32] Prince V.E., Pickett F.B., Splitting pairs: the diverging fates of duplicated genes, *Nat. Rev. Genet.* 3 (2002) 827–837.
- [33] Romer A.S., *Vertebrate Paleontology*, 3rd edn., The University of Chicago Press, Chicago & London, 1966, pp. 356–357.
- [34] Rzhetsky A., Nei M., Theoretical foundation of the minimum-evolution method of phylogenetic inference, *Mol. Biol. Evol.* 10 (1993) 1073–1095.
- [35] 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* 3 (2002) 167–170.
- [36] Sneath P.H.A., Socol R.R., *Numerical taxonomy*, Freeman, San Francisco, CA, 1973.
- [37] Van de Peer Y., Taylor J.S., Braasch I., Meyer A., The ghost of selection past: Rates of evolution and functional divergence of anciently duplicated genes, *J. Mol. Evol.* 53 (2001) 436–446.
- [38] Vandepoele K., De Vos W., Taylor J.S., Meyer A., Van de Peer Y., Major events in the genome evolution of vertebrates: Paralogy age and size differ considerably between ray-finned fishes and land vertebrates, *Proc. Natl. Acad. Sci. USA* 101 (2004) 1638–1643.
- [39] Wegner M., From head to toes: the multiple facets of Sox proteins, *Nucleic Acids Res.* 27 (1999) 1409–1420.
- [40] Wilson M., Koopman P., Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators, *Curr. Opin. Genet. Dev.* 12 (2002) 441–446.
- [41] Wittbrodt J., Meyer A., Schartl M., More genes in fish?, *Bioessays* 20 (1998) 511–515.

- [42] Wright E.M., Snopek B., Koopman P., Seven new members of the *Sox* gene family expressed during mouse development, *Nucleic Acids Res.* 21 (1993) 744.
- [43] Yan Y.-L., Willoughby J., Liu D., Crump J.G., Wilson C., Miller C.T., Singer A., Kimmel C., Westerfield M., Postlethwait J.H., A pair of Sox: distinct and overlapping functions of zebrafish Sox9 co-orthologs in craniofacial and pectoral fin development, *Development* 132 (2005) 1069–1083.
- [44] Zhou R., Cheng H., Zhang Q., Guo Y., Cooper R.K., Tiersch T.R., *SRY*-related genes in the genome of the rice field eel (*Monopterus albus*), *Genet. Sel. Evol.* 34 (2002) 129–137.

