Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics 38 (2021) 100817

Contents lists available at ScienceDirect

Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics

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

Systematic identification and expression analysis of the *Sox* gene family in spotted sea bass (*Lateolabrax maculatus*)

Bingyu Li [1], Yuan Tian [1], Haishen Wen, Xin Qi, Lingyu Wang, Jingru Zhang, Jinku Li, Ximeng Dong, Kaiqiang Zhang, Yun Li [*]

Key Laboratory of Mariculture (Ocean University of China), Ministry of Education, Qingdao 266003, PR China

ARTICLE INFO

Keywords:
Spotted sea bass
Sox family
Gene expression
Development
Sex differentiation

ABSTRACT

The *Sox* gene family encodes a set of transcription factors characterized by a conserved *Sry*-related high mobility group (HMG)-box domain, which performs a series of essential biological functions in diverse tissues and developmental processes. In this study, the *Sox* gene family was systematically characterized in spotted sea bass (*Lateolabrax maculatus*). A total of 26 *Sox* genes were identified and classified into eight subfamilies, namely, SoxB1, SoxB2, SoxC, SoxD, SoxE, SoxF, SoxH and SoxK. The phylogenetic relationship, exon-intron and domain structure analyses supported their annotation and classification. Comparison of gene copy numbers and chromosome locations among different species indicated that except tandem duplicated paralogs of *Sox17/Sox32*, duplicated *Sox* genes in spotted sea bass were generated from teleost-specific whole genome duplication during evolution. In addition, qRT-PCR was performed to detect the expression profiles of *Sox* genes during development and adulthood. The results showed that the expression of 16 out of 26 *Sox* genes was induced dramatically at different starting points after the multicellular stage, which is consistent with embryogenesis. At the early stage of sex differentiation, 9 *Sox* genes exhibited sexually dimorphic expression patterns, among which *Sox3*, *Sox19* and *Sox6b* showed the most significant ovary-biased expression. Moreover, the distinct expression pattern of *Sox* genes was observed in different adult tissues. Our results provide a fundamental resource for further investigating the functions of *Sox* genes in embryonic processes, sex determination and differentiation as well as controlling the homeostasis of adult tissues in spotted sea bass.

1. Introduction

The *Sox* gene family encodes a group of transcription factors presenting throughout the animal kingdom (Gubbay et al., 1990). These *Sox* genes are characterized by the presence of the high mobility group (HMG) box domain and are involved in regulating diverse developmental processes (Sinclair et al., 1990; Heenan et al., 2016). Based on the HMG-box domain and a lesser extent the sequence features outside of the HMG-box, *Sox* genes are mainly classified into 12 subfamilies (from SoxA to SoxK). Of these subfamilies, the SoxB, SoxC, SoxD, SoxE, and SoxF subfamilies exist in almost all species, while the others seem to be species-specific (Bowles et al., 2000; Wegner, 2010).

Since the *Sox* gene was first detected in mice in 1990 (Gubbay et al., 1990), more than 40 *Sox* genes have been identified in a broad range of animal taxa by homologous screening approaches. However, the number of *Sox* genes varies greatly in different species. For instance, 8–9 *Sox* genes were discovered in insects (Cremazy et al., 2001; Wilson and Dearden, 2008; Wei et al., 2011), 11–14 were reported in echinoderms and coelenterates (Magie et al., 2005; Howard-Ashby et al., 2006), 18 were found in chickens (*Gallus gallus*) (Takada et al., 2005; Wei et al., 2016), and 20 were identified in mammals (Schepers et al., 2002). Due to teleost-specific whole-genome duplication (3R-WGD) events, fish species usually contain more *Sox* genes than other vertebrates. For example, 25 *Sox* genes have been identified in channel catfish (*Ictalurus punctatus*) (Zhang et al., 2018), 26 in large yellow croaker (*Larimichthys crocea*) (Wan et al., 2019), and 27 in both Nile tilapia (*Oreochromis niloticus*) (Wei et al., 2016) and Japanese flounder (*Paralichthys olivaceus*) (Yu et al., 2018). Notably, common carp (*Cyprinus carpio*) possess 49 *Sox* genes, while salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*) harbor 51 and 49 *Sox* genes, respectively, as they underwent additional genome duplication events during evolution (Berthelot et al., 2014; Xu et al., 2014).

* Corresponding author at: Ocean University of China, No 5 Yushan road, Qingdao 266003, PR China.
E-mail address: yunli0116@ouc.edu.cn (Y. Li).
[1] These authors equally contributed to this work.

https://doi.org/10.1016/j.cbd.2021.100817
Received 17 October 2020; Received in revised form 10 February 2021; Accepted 12 February 2021
Available online 2 March 2021
1744-117X/

Sox genes have been shown to be involved in regulating diverse developmental processes, including but not limited to sex determination and differentiation (Koopman, 2005), cell type specification (Avilion et al., 2003), neurogenesis (Hong and Saint-Jeannet, 2005) and organogenesis (Wegner, 1999). Because of their importance, *Sox* genes have been extensively studied to define their specific functions in higher vertebrates. In comparison, although studies of *Sox* genes in teleosts remain limited, a growing number of publications about the roles of *Sox* genes in fish species have been reported in the past decade. For instance, gene expression patterns indicated that several members of the *Sox* family in subgroups B, C, E and F may play various roles in the early embryonic development of medaka (*Oryzias latipes*) (Cui et al., 2011). In zebrafish (*Danio rerio*), *Sox17* is considered the master gene controlling the formation of the endoderm in early embryonic development (Kikuchi et al., 2001), and *Sox7* and *Sox18* have important functions in vascular development and arteriovenous specificity (Cermenati et al., 2008; Herpers et al., 2008). In terms of sex determination and differentiation, the essential functions of *Sox9* have been reported in several fish species. For example, in medaka, *Sox9* was demonstrated to play a role in testis cord differentiation rather than testis determination (Nakamoto et al., 2005). Moreover, *Sox9b* has been shown to be indispensable for the proper proliferation and survival of germ cells in gonads (Nakamura et al., 2012). Two homologs of *Sox9* genes play a pivotal role in the sexual differentiation, spermatogenesis and maintenance of gonadal function in Japanese flounder (Li et al., 2018). In addition, *Sox3*, *Sox19* and *Sox30* have been proven to play roles in gonadal development or sex differentiation in different teleosts (Han et al., 2010; Sutton et al., 2011; Navarro-Martin et al., 2012).

Spotted sea bass (*Lateolabrax maculatus*) is one of the most economically valuable species in the Chinese aquaculture industry, with production of over 180,000 tons last year (China Fishery Statistical Yearbook 2019). The draft genome sequence of spotted sea bass has been published (Shao et al., 2018; Chen et al., 2019); however, a systematic analysis of *Sox* genes has not been conducted. In the present study, a comprehensive analysis of gene identification, sequence structure, and evolutionary features was performed to systematically investigate the *Sox* gene family in spotted sea bass. Additionally, gene expression patterns were detected to study the potential functions of *Sox* genes in embryonic development stages, adult tissues, and the early stage of sex differentiation. The results will help us better understand the biological roles of *Sox* genes in spotted sea bass and will contribute substantially to further studies of *Sox* genes in other teleost species.

2. Materials and methods

2.1. Ethics statement

All experiments involving animals were conducted in accordance with the Animal Research and Ethics Committees of Ocean University of China (Permit Number: 20141201). Endangered or protected species were not involved in this field study, and experiments were performed according to the relevant guidelines.

2.2. Genome-wide identification of Sox genes in spotted sea bass

To identify *Sox* genes in spotted sea bass, reference genome (GCA_004028665.1), Iso-Seq (PRJNA515783) and RNA-Seq (PRJNA347604; PRJNA515986) databases were searched by TBLASTN using amino acid sequences of *Sox* genes from human and zebrafish retrieved from NCBI (https://www.ncbi.nlm.nih.gov/) and Ensembl (http://asia.ensembl.org/index.html) as queries with the threshold level of E-value of $1e^{-5}$. For the genes without annotation in the reference genome of spotted sea bass, the FGENESH program of Softberry software was employed to predict target genes based on the zebrafish model (Solovyev et al., 2006). The candidate *Sox* sequences were submitted to ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/) to predict the open reading frames (ORFs), and the ORFs were translated into amino acid sequences. The annotation results were further verified by performing a BLAST search against the NCBI non-redundant (NR) database. In addition, the theoretical isoelectric points (pI) and molecular weight (MW) of the Sox proteins in spotted sea bass were predicted using the online analysis tool EXPASY (https://web.expasy.org/compute_pi/).

2.3. Phylogenetic and comparative genomic analyses of Sox genes

The phylogenetic analysis was constructed using the amino acid sequences of *Sox* genes from human, zebrafish, tilapia, pufferfish (*Fugu rubripes*), channel catfish and spotted sea bass. All reference protein identifiers are shown in Supplementary Table 1. Multiple protein sequences were aligned using ClustalW software with default parameters. Then, the phylogenetic tree was built using MEGA 7.0 software based on the neighbor-joining (NJ) method and Jones-Taylor-Thornton (JTT) model with 1000 bootstrap replicates (Jones et al., 1992; van Megen et al., 2009). The ITOL website (https://itol.embl.de/login.cgi) was used to further adjust the phylogenetic tree. The collinearity analysis was conducted by comparing the chromosome distribution of zebrafish, tilapia and spotted sea bass. Chromosomal position information of Sox genes was obtained from the annotation files of the reference genomes.

2.4. Sequence structure analysis of Sox genes

The exon-intron structures of *Sox* genes in spotted sea bass were retrieved from the annotated files of the reference genomes and were further visualized using the web-based bioinformatics tool Gene Structure Display Server (GSDS: http://gsds.cbi.pku.edu.cn/). The conserved motifs of Sox proteins were predicted using the MEME online tool (http://meme-suite.org/tools/meme). The parameter settings were default values apart from the maximum number of motifs, which was set to 10. The amino acid sequences of the candidate *Sox* genes were submitted to the Simple Modular Architecture Research Tool (SMART) (http://smart.embl.de/) to predict the conserved domains. Then, multiple sequences of domains were aligned using DNAMAN software with the default parameters, and the WebLogo server (http://weblogo.berkeley.edu/logo.cgi) was employed to generate a domain logo for *Sox* genes in spotted sea bass.

2.5. Fish and sample collection

Four-year-old spotted sea bass adults (3 males and 3 females; body length: 48.26 ± 4.64 cm, body weight: 1160.35 ± 90.32 g) were obtained from Shuangying Aquatic Seedling Co., Ltd. (Lijin, Dongying, China). Fish were anesthetized with MS-222, and then tissue samples including brain, heart, muscle, liver, kidney, ovary and testis were quickly dissected, immediately frozen in liquid nitrogen and stored at −80 °C until RNA extraction.

Embryos samples of spotted sea bass were obtained in the second week of October 2019. Embryo samples were collected from the zygote stage (approximately 0 h post-fertilization (hpf)), 16-cell stage (approximately 3.5 hpf), multicellular stage (approximately 7.5 hpf), blastula stage (approximately 13 hpf), gastrula stage (approximately 19 hpf), neurula stage (approximately 23.5 hpf), eye sac formation stage (approximately 26 hpf), sarcomere formation stage (approximately 29 hpf), tail bud stage (approximately 38 hpf), heart beating stage (approximately 51 hpf), last hatching stage (approximately 60 hpf) and 1 day post-hatching (dph) stage (Wen et al., 2019). Images of embryos at different developmental stages during embryogenesis are shown in Supplementary Fig. 1. Approximately 100 embryos were pooled together for each sample and then frozen in liquid nitrogen and stored at −80 °C until RNA extraction.

The gonads were dissected from juvenile individuals (3 females and 3 males) at 205–215 dph, with an average body length of 15.48 ± 0.87 cm. The gonadal samples of each fish were divided into two parts: one was

fixed with 4% paraformaldehyde in 1× phosphate-buffered saline (PBS) at 4 °C overnight for histological observation, and the other was collected for RNA extraction.

2.6. Gonadal histology

The gonadal samples fixed with the 4% paraformaldehyde solution were dehydrated, embedded in paraffin, and serially sectioned at 7 μm (Leica RM2016, GER). Then, sections were stained with hematoxylin and eosin. The developmental stages of gonads were determined according to the histological morphology under a light microscope (Olympus BX53, Japan).

2.7. RNA extraction and qRT-PCR

Total RNA was extracted from tissues and embryo samples using TRIzol Reagent (Invitrogen, CA, USA) according to the manufacturer's instructions. The RNA concentration and quality were measured by determining the ratio of OD260/OD280 using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), and RNA integrity was evaluated by performing 1.5% agarose gel electrophoresis (AGE). The extracted RNA samples were reverse-transcribed into cDNA using the PrimeScript™ RT Master Mix Kit (Takara, Otsu, Japan). Primer5 software and Primer-BLAST of NCBI (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) were used to design specific primers according to the nucleotide sequences of *Sox* genes (Supplementary Table 2). *18S* rRNA was used as the internal reference gene (Wang et al., 2018). qRT-PCR experiments were performed on Applied Biosystems 7300 machines (Applied Biosystems, CA, USA). The total PCR volume was 20 μl and included 2 μl of cDNA templates, 10 μl of ChamQ™ SYBR Color qPCR Master Mix (Vazyme, NJ, China), 0.4 μl of forward primer, 0.4 μl of reverse primer, and 7.2 μl of ddH_2O. The qRT-PCR amplification reaction was carried out under the following conditions: 95 °C for 30 s and 40 cycles of 95 °C for 10 s, 60 °C for 30 s, and 72 °C for 30 s. All qRT-PCR experiments were performed in three biological replicates with three technical replicates. Relative gene expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method.

2.8. Statistical analysis

Data were presented as the means ± standard errors (SE). SPSS 17.0 software was employed for the statistical analyses. One-way analysis of variance (ANOVA) followed by Duncan's multiple range tests were used to analyze the experimental data. Differences were considered significant at $P < 0.05$.

3. Results

3.1. Identification of Sox genes in spotted sea bass

A total of 26 *Sox* genes were identified from the genomic and transcriptomic databases of spotted sea bass. The detailed characteristics of *Sox* genes in spotted sea bass are summarized in Table 1. The 26 *Sox* genes were further divided into 8 subfamilies, including SoxB1, SoxB2, SoxC, SoxD, SoxE, SoxF, SoxH and SoxK. The lengths of *Sox* genes in spotted sea bass ranged from 660 to 2625 bp. In addition, the predicted protein sizes varied from 219 to 874 amino acids (aa), the molecular weights (MWs) from 24.04 to 95.85 KDa, and the isoelectric points (pI) from 5.56 to 9.87. The cDNA sequences of these *Sox* genes have been submitted to the GenBank database, and their accession numbers are listed in Table 1.

3.2. Phylogenetic and comparative genomic analyses of Sox genes

To investigate the evolutionary relationship of *Sox* genes, a phylogenetic tree was constructed for *Sox* genes in selected vertebrates. As shown in Fig. 1, *Sox* genes of spotted sea bass clustered with their counterparts, as expected, and ten clades were generated that were consistent with the previous subfamily classifications, including SoxA, SoxB1, SoxB2, SoxC, SoxD, SoxE, SoxF, SoxG, SoxH and SoxK (SoxA and SoxG were specific in humans).

Comparative maps of *Sox* genes in spotted sea bass, zebrafish, and tilapia were constructed. For spotted sea bass, the 26 genes were unevenly distributed on 19 chromosomes (Chrs) (Fig. 2). Three *Sox* genes are positioned on Chr 15, two *Sox* genes each are located on Chr 2, 11, 18, 20, and 21, and the remaining 13 Chrs (Chr 1, 4, 6, 7, 9, 10, 12, 14, 16, 19, 22, 23, and 24) each harbor one unique *Sox* gene. Among these

Table 1
Sequence characteristics of *Sox* genes in spotted sea bass.

Subfamily classification	Gene name	CDS length (bp)	Predicted protein size (aa)	Isoelectric point	Molecular weight (kDa)	NCBI accession number
B1	*Sox1a*	864	287	9.81	30.822	MT919019
	Sox1b	1065	354	9.70	37.686	MT919020
	Sox2	969	322	9.77	34.353	MT919021
	Sox3	897	298	9.63	33.151	MT919022
	Sox19	660	219	9.79	24.038	MT919023
B2	*Sox14a*	714	237	9.68	26.652	MT919024
	Sox14b	726	241	9.87	26.260	MT919025
	Sox21	735	244	9.74	26.690	MT919026
C	*Sox4a*	1128	375	6.21	40.511	MT919027
	Sox4b	1389	462	6.86	47.920	MT919028
	Sox11a	1158	385	5.56	42.160	MT919029
	Sox11b	1326	441	6.55	48.616	MT919030
D	*Sox5*	2625	874	8.80	95.851	MT919031
	Sox6a	1971	656	8.53	72.873	MT919032
	Sox6b	2328	775	6.53	86.142	MT919033
	Sox13	1914	637	7.40	71.484	MT919034
E	*Sox8a*	1569	522	8.29	56.791	MT919035
	Sox8b	1407	468	6.56	51.233	MT919036
	Sox9a	1521	506	6.14	54.732	MT919037
	Sox9b	1458	485	6.23	53.505	MT919038
	Sox10	1494	497	6.54	52.366	MT919039
F	*Sox7*	1218	405	6.17	44.066	MT919040
	Sox17	1203	400	6.23	44.162	MT919041
	Sox18	1782	593	6.76	62.999	MT919042
H	*Sox30*	1179	392	6.18	43.484	MT919043
K	*Sox32*	984	327	6.37	36.277	MT919044

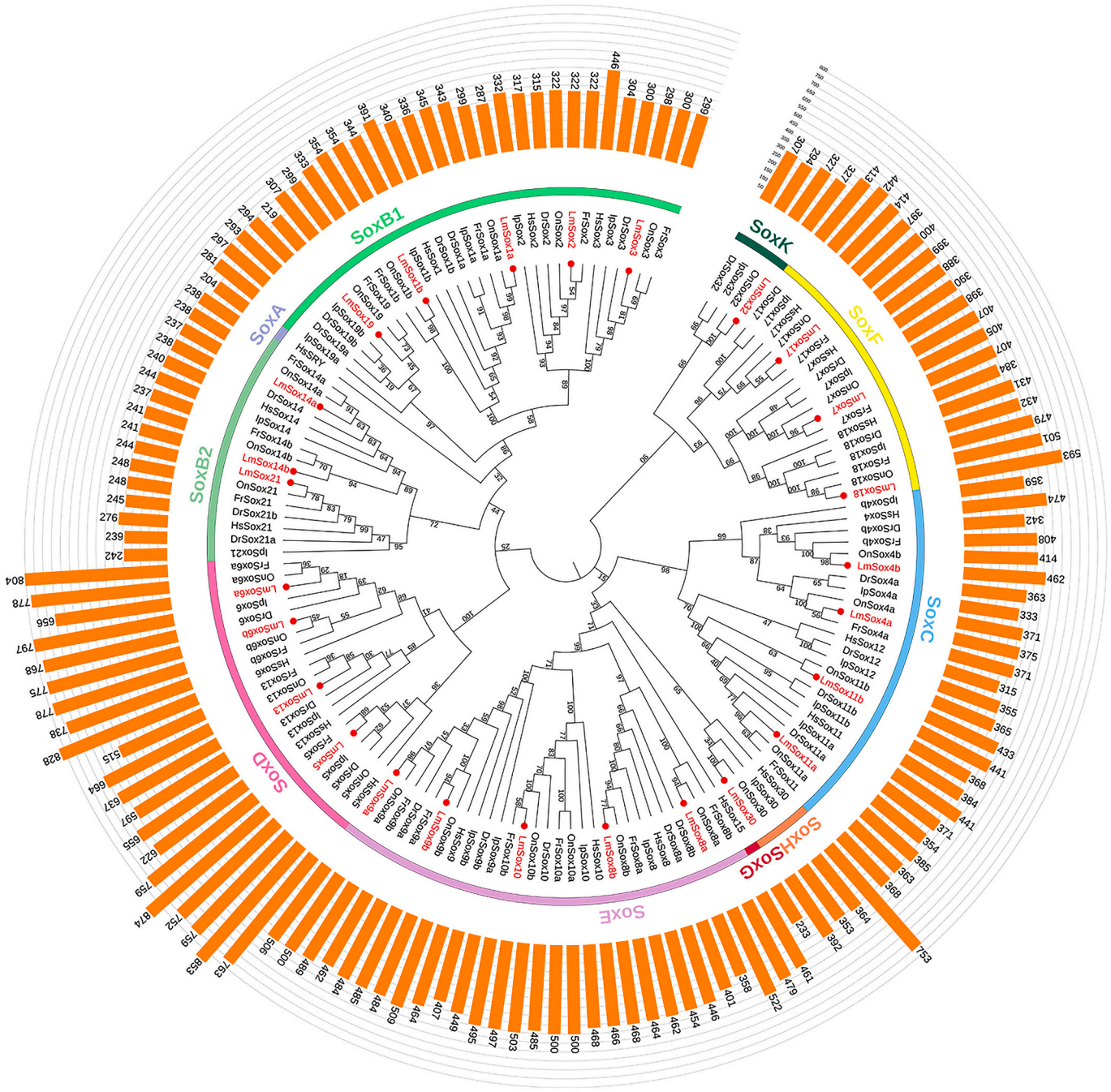

Fig. 1. Phylogenetic tree of *Sox* genes. The amino acid sequences of *Sox* genes from selected vertebrate species were used to construct a neighbor-joining phylogenetic tree with 1000 bootstrap replications using MEGA 7.0 software. *Sox* genes of spotted sea bass are labeled with red dots, and different subfamilies are represented by colored arcs. The simple orange bars around the phylogenetic tree represent the sizes of the predicted amino acid sequences of the *Sox* genes. Lm: *Lateolabrax maculatus*, Hs: *Homo sapiens*, Dr: *Danio rerio*, On: *Oreochromis niloticus*, Fr: *Fugu rubripes*, Ip: *Ictalurus punctatus*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Sox genes, *Sox17* and *Sox32*, which belong to subfamilies F and K, respectively, are tandemly arranged on Chr 20 of spotted sea bass. Similarly, tandem arrangements of *Sox17* and *Sox32* were also detected in the genomes of zebrafish and tilapia (Fig. 2). In addition, the genes included duplicate members, such as *Sox1* (*Sox1a* and *Sox1b*), *Sox4* (*Sox4a* and *Sox4b*), *Sox6* (*Sox6a* and *Sox6b*), *Sox8* (*Sox8a* and *Sox8b*), *Sox9* (*Sox9a* and *Sox9b*), *Sox11* (*Sox11a* and *Sox11b*) and *Sox14* (*Sox14a* and *Sox14b*), which are separately distributed on different Chrs of spotted sea bass. A similar phenomenon was observed in both zebrafish and tilapia genomes (Fig. 2), suggesting that these paralogs of *Sox* genes were mainly generated from whole genome duplication events.

3.3. Analysis of Sox gene copy numbers

The copy numbers of *Sox* genes in several representative vertebrates are shown in Table 2. Overall, teleost species generally contained more *Sox* genes than higher vertebrates. There were 20 members of the *Sox* gene family in human and mouse, 18 in chicken, and 19–27 in teleosts. Notably, *Sox* genes present as singletons in the higher vertebrates examined. In contrast, duplication events of *Sox* genes are common in the teleost genome. In addition, *Sry* and *Sox15* were only detected in human and mouse among the vertebrate species we assessed, while *Sox19* and *Sox32*, which have not yet been discovered in the genomes of

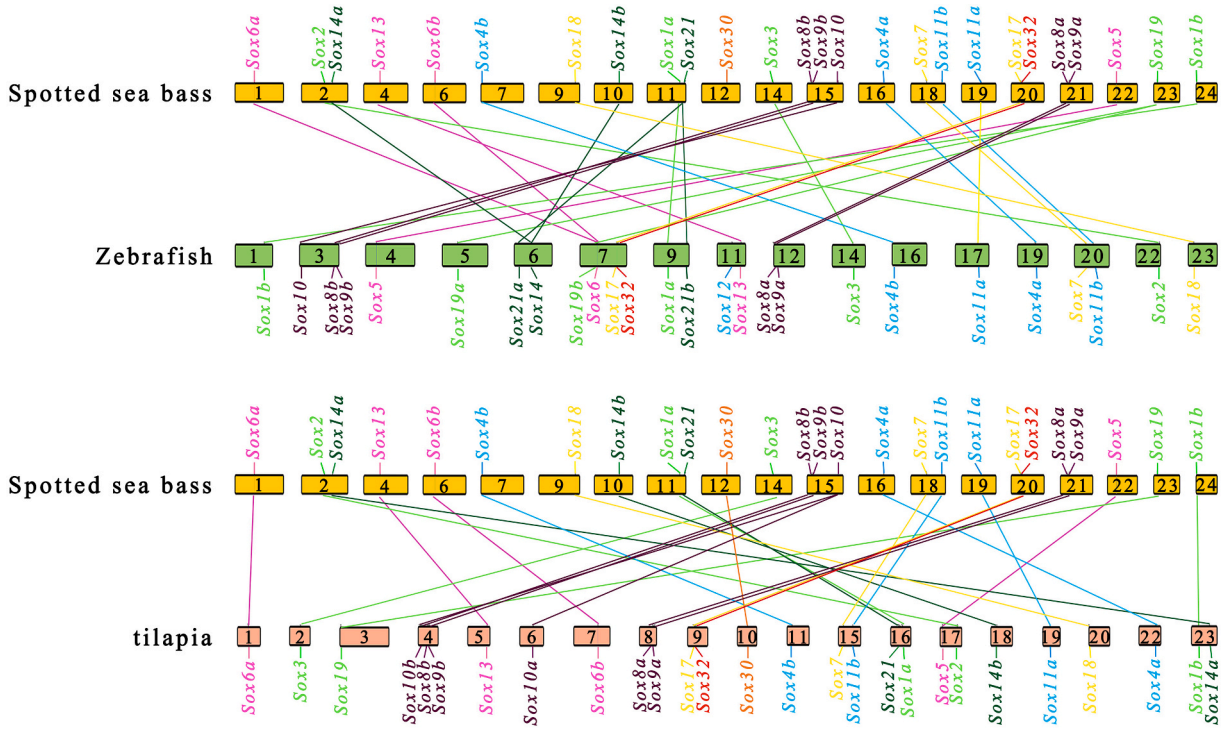

Fig. 2. Homologous relationships of *Sox* genes in spotted sea bass, zebrafish and tilapia. The boxes with numbers represent the reference chromosomes. Homologous genes between species are linked with lines. Lines of the same color represent the same subfamily.

Table 2
Copy number of *Sox* genes among spotted sea bass and other vertebrate species.

Subfamily classification	*Sox* genes	Human	Mouse	Chicken	Zebrafish	Medaka	Japanese flounder	Channel catfish	Large yellow croaker	tilapia	Spotted sea bass
A	*Sry*	1	1	0	0	0	0	0	0	0	0
B1	*Sox1*	1	1	1	2	1	2	2	2	2	2
	Sox2	1	1	1	1	1	1	1	1	1	1
	Sox3	1	1	1	1	1	1	1	1	1	1
	Sox19	0	0	0	2	0	1	2	1	1	1
B2	*Sox14*	1	1	1	1	1	1	1	2	2	2
	Sox21	1	1	1	2	1	1	1	1	1	1
C	*Sox4*	1	1	1	2	1	2	2	2	2	2
	Sox11	1	1	1	2	1	2	2	2	2	2
	Sox12	1	1	1	1	0	0	1	0	0	0
D	*Sox5*	1	1	1	1	1	1	1	1	1	1
	Sox6	1	1	1	1	2	2	1	2	2	2
	Sox13	1	1	1	1	1	1	1	1	1	1
E	*Sox8*	1	1	1	2	1	2	1	2	2	2
	Sox9	1	1	1	2	2	2	2	2	2	2
	Sox10	1	1	1	1	1	2	1	2	2	1
F	*Sox7*	1	1	1	1	1	1	1	1	1	1
	Sox17	1	1	1	1	1	1	1	1	1	1
	Sox18	1	1	1	1	1	1	1	1	1	1
G	*Sox15*	1	1	0	0	0	0	0	0	0	0
H	*Sox30*	1	1	1	0	0	0	1	0	1	1
K	*Sox32*	0	0	0	1	1	1	1	1	1	1
	Total	20	20	18	26	19	25	25	26	27	26

higher vertebrates, are likely teleost-specific. Additionally, a few *Sox* genes seem to be absent in particular teleost species: *Sox12* was not found in five tested fish genomes (spotted sea bass, medaka, Japanese flounder, large yellow croaker and tilapia), *Sox19* was not identified in medaka, and *Sox30* was absent in the genomes of zebrafish, medaka, Japanese flounder and large yellow croaker (Table 2).

3.4. Exon-intron structure and motif analyses of Sox genes

The gene structure analysis results showed that the exon-intron structure and the position of the HMG-box domain are relatively conserved within the same subfamily of *Sox* genes in spotted sea bass (Fig. 3A). Specifically, the genes in subfamilies SoxB1 (except for *Sox1a*), SoxB2 and SoxC harbor only one exon (no introns), while the SoxF and SoxK genes contain two exons. Three exons exist in all SoxE genes, and members of the SoxH subfamily have four exons. Notably, members of the SoxD subfamily possess significantly larger numbers of exons, which range from 13 (*Sox13*) to 18 (*Sox5*).

Similarly, genes that clustered into the same *Sox* subfamilies contained similar motif structures (Fig. 3B). Of the 10 detected motifs, motif

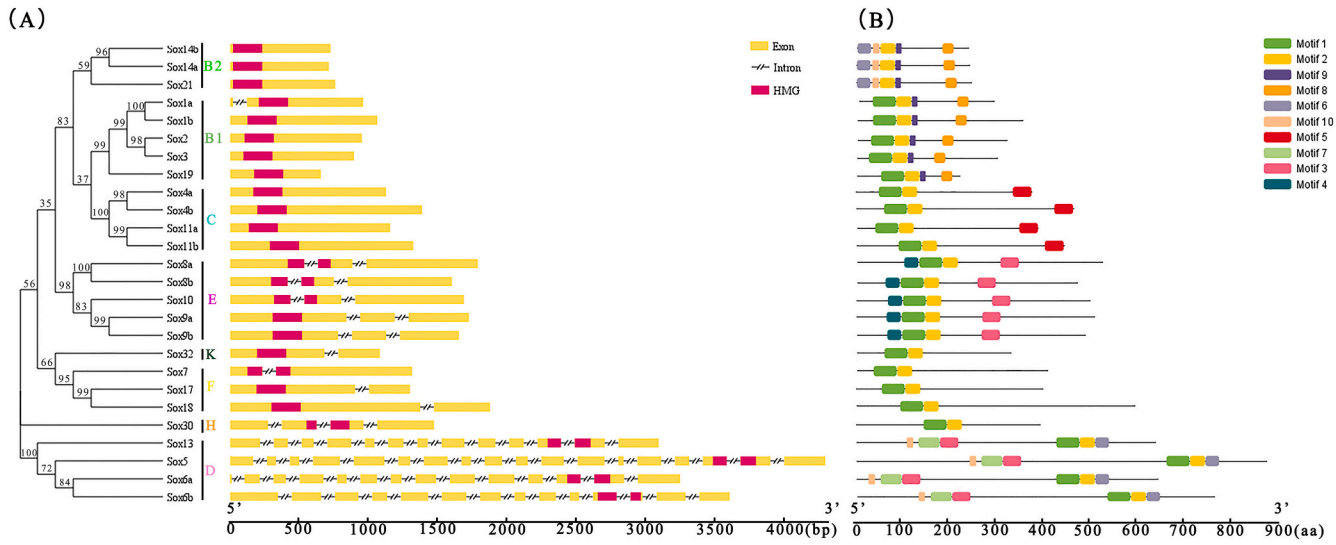

Fig. 3. Exon-intron structure and motif analyses of Sox genes in spotted sea bass. (A) Exon–intron structures of *Sox* genes in spotted sea bass. The tree was constructed based on the neighbor-joining method and JTT model with 1000 bootstrap replicates using MEGA 7.0 software. Yellow boxes and black horizontal lines indicate exons and introns, respectively. The HMG-box domain is marked with a fuchsia box. (B) Motif structure of *Sox* genes in spotted sea bass. The conserved motifs were identified using MEME online software. The different motifs are shown in different colors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

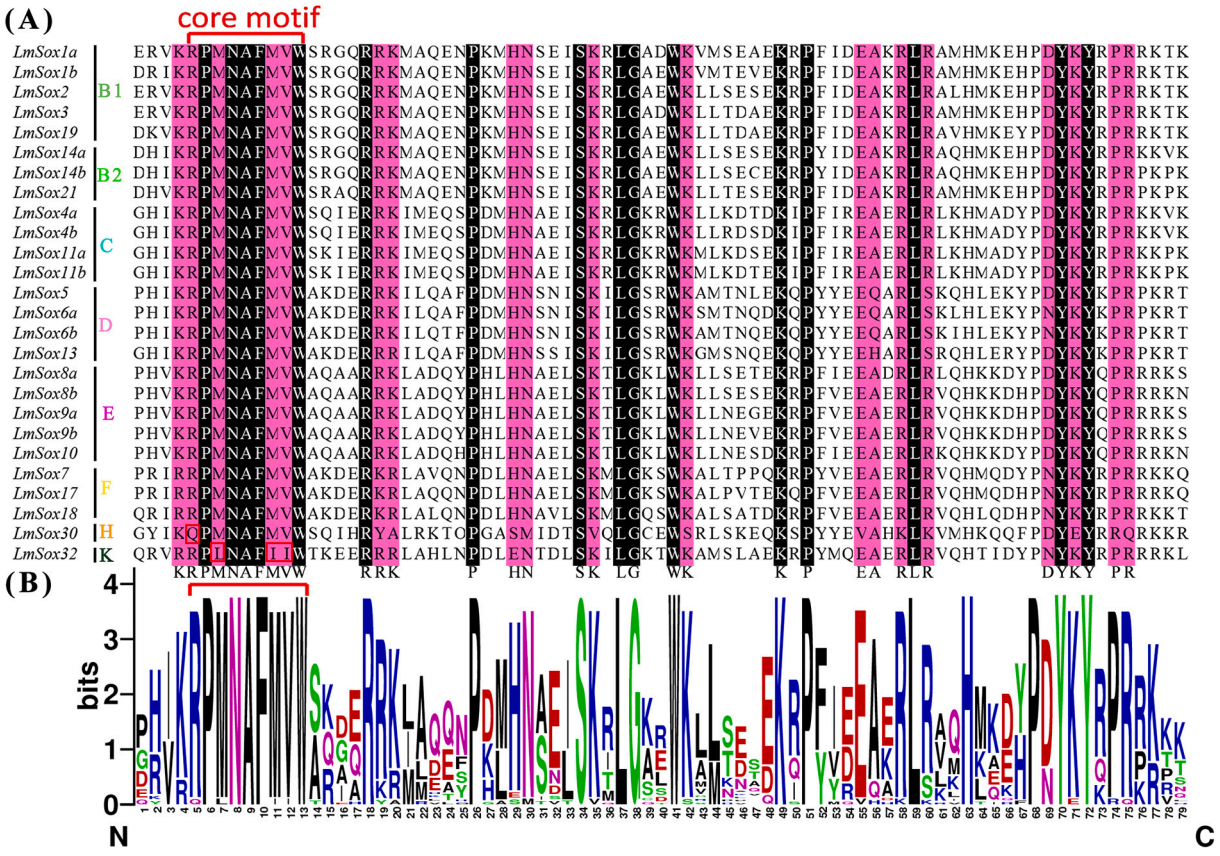

Fig. 4. Sequence analysis of the HMG-box domains of *Sox* genes in spotted sea bass. (A) Multiple alignment of HMG-box domain sequences. Acid residues that are identical among all *Sox* proteins are marked with a black box, and the amino acid residues with conservation ratesgreater than 75% are indicated in pink. (B) Sequence logo plots showing the sequence types and relative frequency at each position of the HMG-box domain. Numbers on the horizontal axis indicate the residue position numbering of the HMG-box domain. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2, which is part of the HMG-box, existed in all *Sox* genes. In addition, motif 1 is present in all genes, except for SoxB2 subfamily genes. The distribution patterns of the other motifs show a certain degree of specificity among subfamilies. For instance, motif 5 is mainly present in subfamily C, while motif 9 is only present in subfamilies B1 and B2. These specific motifs may also contribute to the functional divergence of *Sox* genes in spotted sea bass.

3.5. Multiple sequence alignment analysis of the HMG-box domain

Multiple sequence alignments were conducted to characterize the sequence features of the functional HMG-box domain among the 26 *Sox* genes in spotted sea bass (Fig. 4A). The HMG-box domain is a coherent and compact domain generally composed of 79 amino acid residues. Multiple sequence alignment results showed homology among *Sox* genes with an average identity of 70.87%. The core motif sequences (RPMNAFMVW) are located at positions 5–13 in the HMG-box domain of spotted sea bass and are highly conserved for all *Sox* sequences, except *Sox30* and *Sox32*. The 1st amino acid residue in the core motif RPMNAFMVW of *Sox30* is Gln (Q), and the amino acid residues at positions 3, 7 and 8 of the core motif of *Sox32* are substituted by Leu (L), Ile (I) and Ile (I), respectively. In addition, several other conserved amino acid sites are present in the functional HMG-box domain. Sequence logo plots were generated to clearly delineate the degree of conservation of these conserved sites (Fig. 4B).

3.6. Expression patterns of Sox genes during embryonic development stages

The temporal expression profiles of *Sox* genes during embryonic development were detected by qRT-PCR and are shown in Fig. 5. The results indicated that the expression of most *Sox* genes increased during embryo development in spotted sea bass, which is consistent with embryogenesis. Notably, the mRNA levels of half of the *Sox* genes were highly induced ($\log_2$FC $>$ 2) after the multicellular stage, although at different starting points. For example, the expression of *Sox7* was upregulated at the multicellular stage; *Sox2*, *Sox17* and *Sox32* were upregulated at the blastocyst stage; *Sox3*, *Sox6b*, *Sox9b*, *Sox10*, *Sox11a*, and *Sox21* were induced at the gastrula stage; the mRNA levels of *Sox1a*, *Sox1b*, *Sox9a and Sox13* increased dramatically at the neurula stage; and *Sox14a* and *Sox18* were induced before the heart beating stage. After the heart beating stage, the expression of these *Sox* genes increased slightly or was maintained at a relatively high level during the subsequent developmental stages. In contrast, the transcript abundance of the remaining *Sox* genes, including *Sox4b*, *Sox5*, *Sox6a*, *Sox8a*, *Sox11b*, *Sox14b*, *Sox19* and *Sox30*, did not change dramatically during embryogenesis (Fig. 5 and Supplementary Table 3).

3.7. Gonadal expression of Sox genes at the early sex differentiation stage

Prior to determining the expression patterns of *Sox* genes at the sex

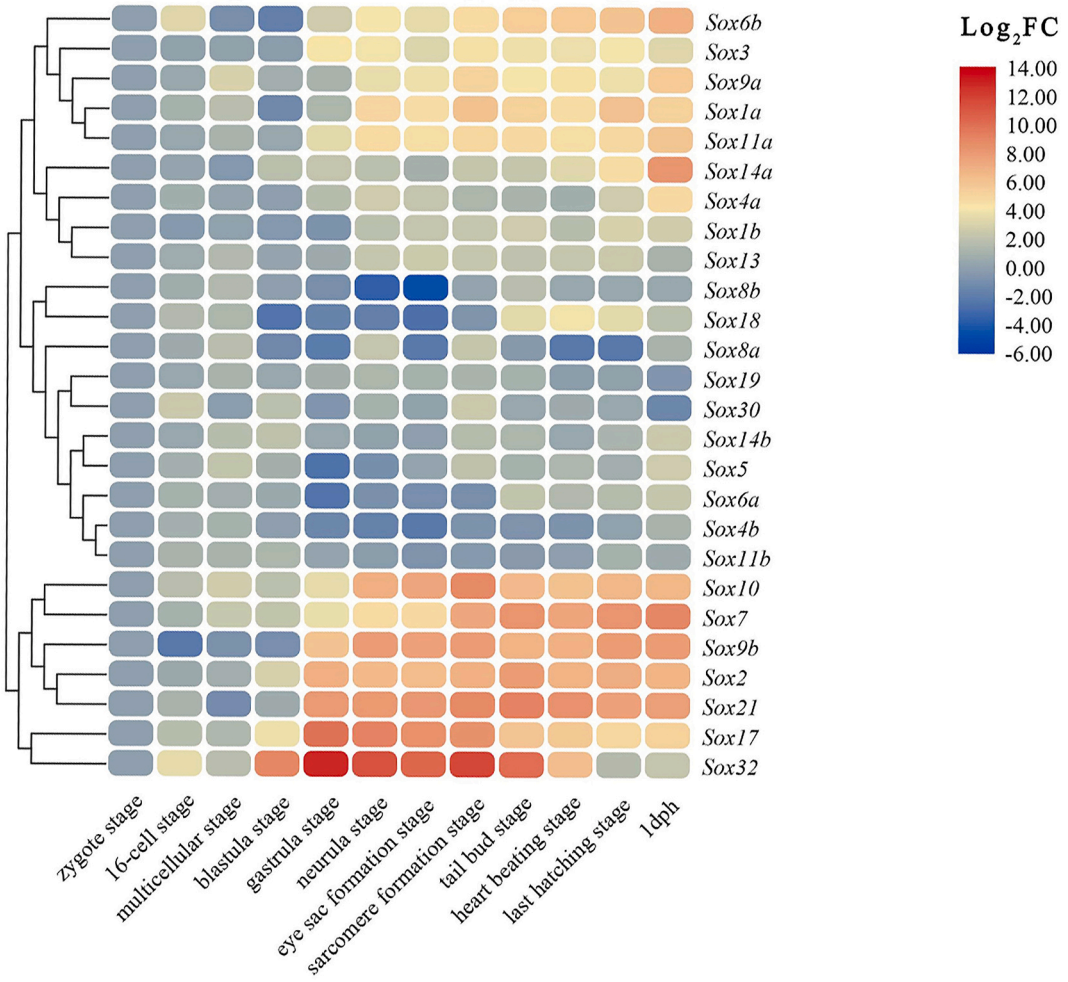

Fig. 5. Temporal expression patterns of *Sox* genes in spotted sea bass at different embryonic development stages. The log-transformed ($\log_2$) values of the relative expression levels of *Sox* family genes generated by qPCR were used to construct the heat map. The expression levels were calculated relative to the zygote stage. The color scale represents expression levels, and warmer colors indicate higher expression. Dph: days post-hatching.

differentiation stage by qRT-PCR, the histological structures of the juvenile gonads were observed to verify the morphological development stage. In this study, we focused on samples in the early sex differentiation stage, which were obtained from juvenile fish gonads at 205–215 dph with an average body length of 15.48 ± 0.87 cm. The results indicated that at the early stage of sex differentiation, only oogonia or spermatogonia of germ cells were observed in the ovary or testis. At this stage, the germ cells did not initiate meiosis, although the ovarian cavity and sperm duct had already formed (Fig. 6A and B).

qRT-PCR results showed that 9 of 26 *Sox* genes, namely, *Sox1a*, *Sox3*, *Sox19*, *Sox14a*, *Sox6a*, *Sox6b*, *Sox10*, *Sox18*, and *Sox32*, exhibited sexually dimorphic expression patterns (expression differences >2-fold) at the early sex differentiation stage of spotted sea bass. Among them, the relative expression levels of most *Sox* genes in the ovary were significantly higher than those in the testis, particularly *Sox3*, *Sox19* and *Sox6b*, which showed differences greater than 4-fold (Fig. 6C).

3.8. Expression patterns of Sox genes in adult tissues

To investigate the tissue distribution patterns of *Sox* genes in adult spotted sea bass, qRT-PCR was performed to determine the expression levels of 26 *Sox* genes in different tissues, including kidney, brain, heart, muscle, liver, testis and ovary, of adult spotted sea bass. As shown in Fig. 7, *Sox* genes exhibited distinct expression levels among various tissues in spotted sea bass. Notably, most *Sox* genes showed relatively higher expression levels in the brain. For example, the highest mRNA abundance of *Sox1a*, *Sox1b*, *Sox2*, *Sox3*, *Sox5*, *Sox8b*, *Sox9b*, *Sox10*, *Sox11a*, *Sox13*, *Sox17*, *Sox19* and *Sox21* was observed in the brain among all tested tissues (except the gonads). In addition, several genes exhibited relatively high expression levels in muscle and gonads. Notably, sex-biased expression was observed for several *Sox* genes in the adult gonads. For example, our results showed that *Sox3*, *Sox4b*, *Sox11b* and *Sox19* were predominantly expressed in the ovary, while *Sox4a*, *Sox5*, *Sox6a*, *Sox6b*, *Sox8a*, *Sox9a*, *Sox9b*, *Sox10*, *Sox14b*, *Sox17*, *Sox18*, *Sox21*, *Sox30* and *Sox32* showed higher expression levels in the testis (Fig. 7 and Supplementary Table 4).

4. Discussion

In the present study, a total of 26 *Sox* genes were identified in the spotted sea bass genome and further classified into 8 subfamilies: SoxB1 (*Sox1a*, *Sox1b*, *Sox2*, *Sox3*, and *Sox19*), SoxB2 (*Sox14a*, *Sox14b*, and

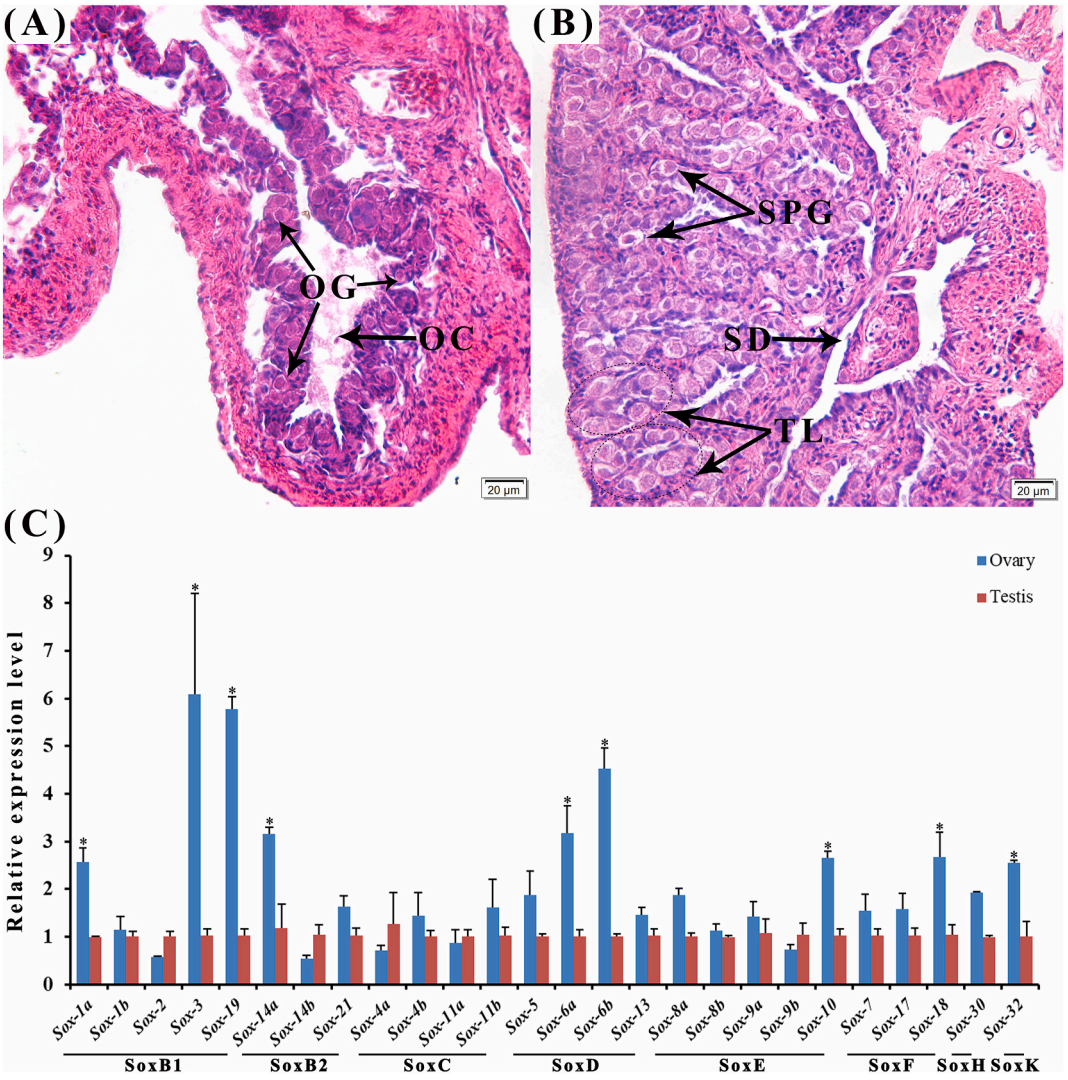

Fig. 6. Longitudinal sections of (A) ovary ($n = 3$) and (B) testis (n = 3) at the early stage of sexual differentiation stained with hematoxylin and eosin. OG: oogonia; OC: ovarian cavity; SPG: spermatogonia; TL: testicular lobules; SD: sperm duct. (C) Gonadal expression of *Sox* genes in spotted sea bass at the early sex differentiation stage. The expression levels of *Sox* genes in testes were set as internal controls. Asterisks indicate that the difference was significant ($P < 0.05$), and the differential expression was greater than 2-fold.

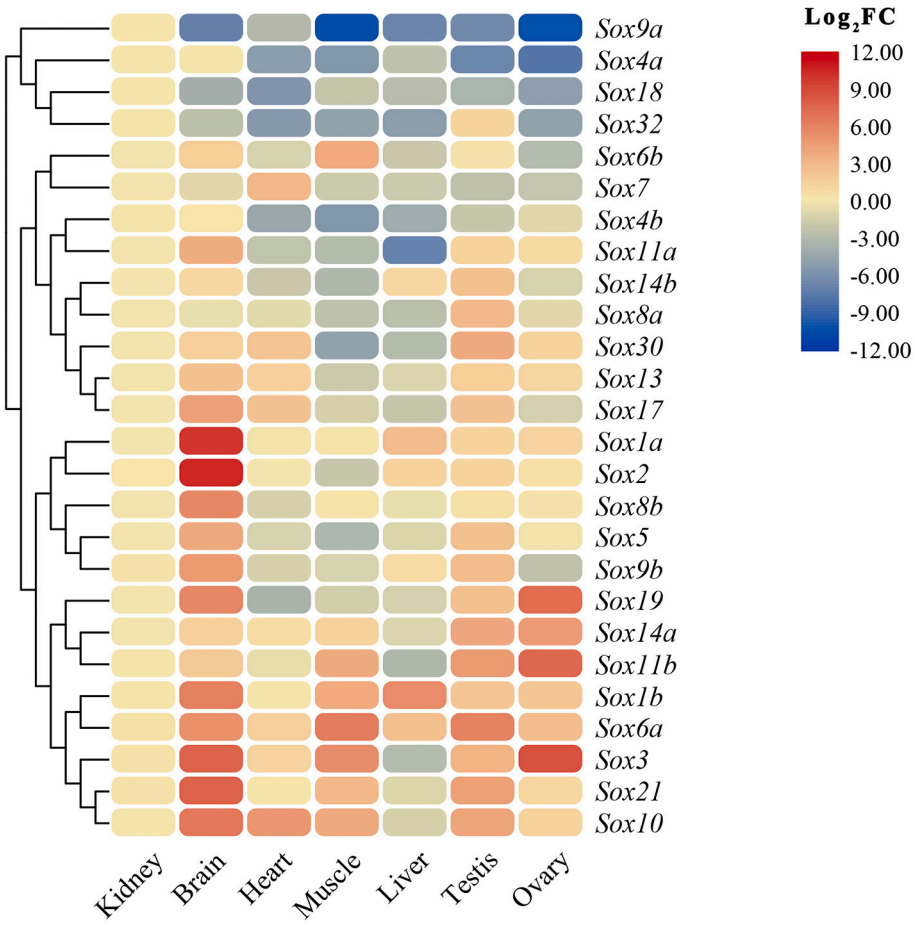

Fig. 7. Expression patterns of *Sox* genes in adult tissues of spotted sea bass (n = 3). The expression levels of *Sox* genes in the kidney were set as internal controls to calculate the relative expression levels of *Sox* genes. Log_2 values of the relative expression levels were used to construct the heat map, and the warmer colors indicate higher expression.

Sox21), SoxC (*Sox4a*, *Sox4b*, *Sox11a*, and *Sox11b*), SoxD (*Sox5*, *Sox6a*, *Sox6b*, and *Sox13*), SoxE (*Sox8a*, *Sox8b*, *Sox9a*, *Sox9b*, and *Sox10*), SoxF (*Sox7*, *Sox17*, and *Sox18*), SoxH (*Sox30*) and SoxK (*Sox32*). Their annotations were further confirmed by the results of domain homology, sequence structure and phylogenetic analyses. Compared with higher vertebrates, the numbers of *Sox* genes were significantly expanded in most tested teleosts (Table 2). Of these genes, *Sox19* and *Sox32* are exclusively present in teleosts. *Sox19* was first characterized in zebrafish (Vriz and Lovellbadge, 1995) and then identified in several bony fishes, such as pufferfish, European sea bass and large yellow croaker (Koopman et al., 2004; Navarro-Martin et al., 2012; Wan et al., 2019). *Sox19* is present as a single copy in spotted sea bass, while it is duplicated in zebrafish and channel catfish (Table 2) and derived from the teleost-specific WGD (Voldoire et al., 2017). The *Sox32* gene in spotted sea bass appears to be a duplicate of *Sox17* because it is localized on the same chromosome as tandem duplicated genes (Fig. 2). The same arrangement was reported in several teleost genomes, such as zebrafish and tilapia, suggesting that the paralogs of *Sox17* and *Sox32* detected in teleosts might have arisen through small-scale duplication (SSD) events instead of WGD (Voldoire et al., 2017). In contrast, all the other duplicated *Sox* genes in spotted sea bass, including *Sox1a/1b*, *Sox4a/4b*, *Sox6a/6b*, *Sox8a/8b*, *Sox9a/9b*, *Sox11a/Sox11b*, and *Sox14a/Sox14b*, probably derived from teleost-specific WGD (Fig. 2). This phenomenon was conserved among all teleost species tested in previous studies, as these duplicated genes were located on paralogous chromosomes derived from WGD (Voldoire et al., 2017), indicating that the teleost-specific WGD had an essential role in the expansion of the *Sox* gene family in this lineage. This observation corresponds to the hypothesis that the duplicated genes coding for transcription factors were preferably retained after WGD (Aury et al., 2006; Kassahn et al., 2009). In addition to gene duplication, loss of specific *Sox* genes occurred in particular species. For instance, *Sox30* was previously suspected to be specific to mammals (Koopman et al., 2004). However, the *Sox30* gene was detected in some teleost genomes, including tilapia (Han et al., 2010), channel catfish (Zhang et al., 2018), culter (*Culter alburnus*) (Zheng et al., 2020) and spotted sea bass in the present study. Meanwhile, *Sox10* and *Sox19* were not detected in the genomes of cavefish and medaka, respectively (Table 2).

Sox transcription factors play essential roles in diverse physiological processes by activating or repressing the transcription of their target genes in a tissue- or developmental stage-specific manner (Lefebvre et al., 2007; Wegner, 2010; Wei et al., 2016). In this study, we performed qRT-PCR experiments to investigate the potential involvement of *Sox* genes during the development and adulthood of spotted sea bass. During the embryonic development stages, various processes, including cell fate specification, embryo patterning, and morphogenesis, occur concurrently. *Sox* transcription factor genes are thought to play a key role in these embryonic processes. Several *Sox* genes have been reported to play essential roles in gastrulation, which is central to animal development and involves the specification of three germ layers (endoderm, mesoderm and ectoderm) (Wolpert, 1992). For example, we detected high expression of *Sox17* and *Sox32* before the onset of gastrulation in spotted sea bass, with the highest expression level appearing at the gastrula stage (Fig. 5). Functional studies have proven that *Sox*17 plays key roles in endoderm formation in *Xenopus* (Hudson et al., 1997), zebrafish (Alexander and Stainier, 1999), and mouse (Engert et al.,

2013; Igarashi et al., 2018). As the teleost-specific gene was very likely duplicated from *Sox17*, the *Sox32* gene has been shown to play an essential cell-autonomous role in endoderm formation in zebrafish (Kikuchi et al., 2001). In zebrafish, both *Sox17* and *Sox32* have been proven to function as Nodal signaling pathway-related endoderm-specific transcription factors, and *Sox32* has been reported to be located downstream of Nodal signaling and upstream of *Sox17* (Chan et al., 2009). Regarding the differentiation of mesodermal lineages, *Sox7* has critical functions in hematopoiesis, vasculogenesis and cardiogenesis during mammalian embryonic development (Doyle et al., 2019). In our results, *Sox7* was upregulated beginning at the multicellular stage and its expression remained at a high level. SoxB1 transcription factor genes are thought to play a key role from the blastoderm stage to the neural stage, particularly during the development of the primordium of the central nervous system (Okuda et al., 2010). Correspondingly, our results showed that *Sox1a*, *Sox1b*, *Sox2*, and *Sox3*, which belong to the SoxB1 subfamily, exhibited significantly higher expression starting from the gastrula stage in spotted sea bass (Fig. 5). The SoxB2 subfamily, including *Sox14* and *Sox21*, is dissimilar to the SoxB1 genes in their non-HMG sequences. It has been reported that the *Sox21* expression pattern resembles that of *Sox2* in both chicken and mouse embryos (Uchikawa et al., 2011). Obviously, the most similar expression patterns were observed for *Sox21* and *Sox2* in spotted sea bass (Fig. 5). Evidence has shown that *Sox2* and *Sox21* are woven into highly interconnected regulatory networks that function at several levels to control the fate of embryonic stem cells (Mallanna et al., 2010). In addition, the neural crest, which contains a stem cell population unique to vertebrate embryos, gives rise to derivatives from multiple embryonic germ layers. Vertebrate SoxE genes are believed to be key regulators of neural crest cell (NCC) development, and at least two of these genes (*Sox9* and *Sox10*) are required for NCC specification and differentiation (Spokony et al., 2002; Carney et al., 2006; Lee et al., 2016). Therefore, it is reasonable that the *Sox9a*, *Sox9b* and *Sox10* genes in the SoxE subfamily of spotted sea bass were highly expressed during embryogenesis (Fig. 5). Overall, in spotted sea bass, the mRNA expression levels of most *Sox* genes change with embryo development, consistent with embryogenesis, suggesting that the *Sox* family plays a key role in embryonic processes.

In addition to regulating developmental processes, *Sox* family proteins also control homeostasis in adult tissues. In our study, the expression patterns of *Sox* genes in spotted sea bass were detected in adult tissues (Fig. 7). Members of the SoxB1 subfamily have been proven to be coexpressed during the proliferation of neural progenitor cells in the developing and adult central nervous system (CNS) (Buescher et al., 2002). Consistent with this finding, we observed high expression levels of SoxB1 subfamily members, such as *Sox1a*, *Sox1b*, *Sox2* and *Sox3*, in the brains of spotted sea bass (Fig. 7). In addition, high expression of *Sox21* from the SoxB2 subfamily was also detected in the brains of adult spotted sea bass, and the *Sox21* gene has been proven to play a critical regulatory role in adult neurogenesis in the mouse hippocampus (Matsuda et al., 2012). It can be deduced that the function of SoxB factors in the CNS might be evolutionarily conserved. In addition, relatively high expression of SoxE genes, including *Sox8b*, *Sox9b* and *Sox10*, was detected in the brain of spotted sea bass, and this subfamily has been reported to have vital functions in the CNS of mammals (Stolt et al., 2004; Anderson et al., 2015; He et al., 2016). In mammals, *Sox6* is involved in regulating the expression of muscle fiber type-specific genes. Adult mice lacking *Sox6* in the skeletal muscle show an increased number of slow myofibers, thus enhancing muscle endurance (Quiat et al., 2011). In addition, zebrafish *Sox6* is a crucial regulator of fast muscle fiber differentiation (Jackson et al., 2015). Accordingly, compared with other tissues, higher expression of *Sox6a* and *Sox6b* was detected in the muscle of adult spotted sea bass (Fig. 7), which may be necessary for the maintenance of specific biological processes in the muscle.

Members of the *Sox* gene family act as pivotal genes that control sex differentiation and gonad development (Koopman, 2005). However, the mechanism involved in gonadal development and sex differentiation in spotted sea bass is still unclear. In the present study, a total of 9 *Sox* genes (*Sox1a*, *Sox3*, *Sox19*, *Sox14a*, *Sox6a*, *Sox6b*, *Sox10*, *Sox18*, and *Sox32*) displayed significant differences in expression between the ovary and testis at the early stage of sex differentiation in spotted sea bass; in particular, the *Sox3*, *Sox19* and *Sox6b* genes exhibited more than 4-fold higher expression levels in the ovary than in the testis (Fig. 6C). Although several studies on *Sox3* have been conducted in teleosts, its roles in sex determination appear to be species-specific. For example, in Japanese eel (*Anguilla japonica*), *Sox3* was demonstrated to be related to ovarian differentiation, as its transcription increased substantially during ovary development (Jeng et al., 2018). This result was paralleled with findings in grouper (*Epinephelus coioides*) (Yao et al., 2007). In contrast, the *Sox*3 gene of large yellow croaker exhibited testis-biased expression during the sex determination process, suggesting that *Sox3* might be involved in testis development in this species (Xiao et al., 2019). In the current study, the ovarian expression levels of *Sox3* at the early sex differentiation stage and adult stage were both significantly higher than those in the testis, implying that it plays a role in ovarian development in spotted sea bass. Additionally, as a teleost-specific *Sox* member, *Sox19* displayed remarkably ovary-biased expression levels in spotted sea bass at both tested stages, consistent with previous findings in European sea bass (Navarro-Martin et al., 2012), turbot (*Scophthalmus maximus*) (Ribas et al., 2016) and zebrafish (Santos et al., 2007). However, as a teleost-specific gene, the mechanism underlying the role of *Sox19* in sex differentiation requires further investigation. It has been well documented that *Sox9* plays a crucial role in sex determination in mammals (Graves, 1998), and is also an important candidate gene for sex determination, as well as gonadal development and maintenance in teleosts (Chiang et al., 2001; Nakamoto et al., 2005; Li et al., 2018). *Sox9* orthologs, *Sox9a* and *Sox9b*, have been isolated from a number of fishes, although duplicate genes might have distinct functions. In adult Japanese flounder (Li et al., 2018) and juvenile sablefish (*Anoplopoma fimbria*) (Smith et al., 2013), both *Sox9a* and *Sox9b* are expressed at higher levels in the testis than in the ovary. In medaka, the *Sox9b* mRNA is expressed at similar levels in the gonads of both sexes during early sex differentiation. Later, the expression of *Sox9b* is only maintained in the testis at the initial stage of testicular tubule development and is remarkably reduced in the ovary (Nakamoto et al., 2005). Based on our results, the expression levels of *Sox9a* and *Sox9b* in spotted sea bass were not significantly different in the testis and ovary at the early stage of sex differentiation (Fig. 6A and B). Because we did not trace its expression level at later stages of sex differentiation, we are unable to speculate on the function of *Sox9* genes in sex determination and differentiation. However, the expression levels of *Sox9b* in the adult testis were significantly higher than those in the ovary (Fig. 7), suggestting that *Sox9b* has a potential role in the maintenance of the testis in spotted sea bass.

5. Conclusions

In conclusion, a complete set of 26 *Sox* genes was identified and characterized in spotted sea bass, and their expression patterns were examined during the embryonic development process, at the early sex differentiation stage and in adult tissues. Phylogenetic and comparative genomic analyses were conducted to provide sufficient evidence supporting the annotations and orthologies of these genes. The gene structures of the exon-intron and HMG-box domains of *Sox* genes were systematically analyzed, and they provided strong support for the classification of *Sox* genes into 8 subfamilies. qRT-PCR results showed distinct expression levels of *Sox* genes in spotted sea bass among adult tissues, temporal-specific expression profiles during embryonic development, and sex-biased expression patterns at the early sex differentiation stage. This study provides a comprehensive overview of the *Sox* gene family in spotted sea bass and establishes a foundation for further functional analyses of *Sox* genes.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would also like to thank Xiaojie Wang, Yalong Sun and Xuebin Mao for their assistance in the sampling process, as well as Junjie Li, Jianshuang Li, Chu Zeng, Likang Lv and Mengqun Liu for their guidance in the use of the software.

Funding

This research was funded by grants from National Natural Science Foundation of China (NSFC, 32072947), National Key Research and Development Program of China (2018YFD0900101) and the China Agriculture Research System (CARS-47).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cbd.2021.100817.

References

Alexander, J., Stainier, D.Y.R., 1999. A molecular pathway leading to endoderm formation in zebrafish. Curr. Biol. 9, 1147–1157. https://doi.org/10.1016/s0960-9822(00)80016-0.

Anderson, S.R., Lee, I., Ebeling, C., Stephenson, D.A., Schweitzer, K.M., Baxter, D., Moon, T.M., LaPierre, S., Jaques, B., Silvius, D., Wegner, M., Hood, L.E., Carlson, G., Gunn, T.M., 2015. Disrupted SOX10 function causes spongiform neurodegeneration in *gray tremor* mice. Mamm. Genome 26, 80–93. https://doi.org/10.1007/s00335-014-9548-5.

Aury, J.-M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Segurens, B., Daubin, V., Anthouard, V., Aiach, N., Arnaiz, O., Billaut, A., Beisson, J., Blanc, I., Bouhouche, K., Camara, F., Duharcourt, S., Guigo, R., Gogendeau, D., Katinka, M., Keller, A.-M., Kissmehl, R., Klotz, C., Koll, F., Le Mouel, A., Lepere, G., Malinsky, S., Nowacki, M., Nowak, J.K., Plattner, H., Poulain, J., Ruiz, F., Serrano, V., Zagulski, M., Dessen, P., Betermier, M., Weissenbach, J., Scarpelli, C., Schaechter, V., Sperling, L., Meyer, E., Cohen, J., Wincker, P., 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature 444, 171–178. https://doi.org/10.1038/nature05230.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., Lovell-Badge, R., 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev. 17, 126–140. https://doi.org/10.1101/gad.224503.

Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noel, B., Bento, P., Da Silva, C., Labadie, K., Alberti, A., Aury, J.-M., Louis, A., Dehais, P., Bardou, P., Montfort, J., Klopp, C., Cabau, C., Gaspin, C., Thorgaard, G.H., Boussaha, M., Quillet, E., Guyomard, R., Galiana, D., Bobe, J., Volff, J.-N., Genet, C., Wincker, P., Jaillon, O., Roest Crollius, H., Guiguen, Y., 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat. Commun. 5 (1), 3657. https://doi.org/10.1038/ncomms4657.

Bowles, J., Schepers, G., Koopman, P., 2000. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Dev. Biol. 227, 239–255. https://doi.org/10.1006/dbio.2000.9883.

Buescher, M., Hing, F.S., Chia, W., 2002. Formation of neuroblasts in the embryonic central nervous system of Drosophila melanogaster is controlled by SoxNeuro. Development 129, 4193–4203. doi:https://doi.org/10.1016/j.energy.2014.05.097.

Carney, T.J., Dutton, K.A., Greenhill, E., Delfino-Machin, M., Dufourcq, P., Blader, P., Kelsh, R.N., 2006. A direct role for Sox10 in specification of neural crest-derived sensory neurons. Development 133, 4619–4630. https://doi.org/10.1242/dev.02668.

Cermenati, S., Moleri, S., Cimbro, S., Corti, P., Del Giacco, L., Amodeo, R., Dejana, E., Koopman, P., Cotelli, F., Beltrame, M., 2008. Sox18 and Sox7 play redundant roles in vascular development. Blood 111 (5), 2657–2666. https://doi.org/10.1182/blood-2007-07-100412.

Chan, T.-M., Chao, C.-H., Wang, H.-D., Yu, Y.-J., Yuh, C.-H., 2009. Functional analysis of the evolutionarily conserved cis-regulatory elements on the *sox17* gene in zebrafish. Dev. Biol. 326, 456–470. https://doi.org/10.1016/j.ydbio.2008.11.010.

Chen, B., Li, Y., Peng, W., Zhou, Z., Shi, Y., Pu, F., Luo, X., Chen, L., Xu, P., 2019. Chromosome-level assembly of the Chinese seabass (*Lateolabrax maculatus*) genome. Front. Genet. 10, 275. https://doi.org/10.3389/fgene.2019.00275.

Chiang, F.L., Pai, C.I., Wyatt, M., Yan, Y.L., Postlethwait, J., Chung, B.C., 2001. Two sox9 genes on duplicated zebrafish chromosomes: expression of similar transcription activators in distinct sites. Dev. Biol. 231 (1), 149–163. https://doi.org/10.1006/dbio.2000.0129.

Cremazy, F., Berta, P., Girard, F., 2001. Genome-wide analysis of Sox genes in *Drosophila melanogaster*. Mech. Dev. 109, 371–375. https://doi.org/10.1016/s0925-4773(01)00529-9.

Cui, J., Shen, X., Zhao, H., Nagahama, Y., 2011. Genome-wide analysis of *Sox* genes in Medaka (*Oryzias latipes*) and their expression pattern in embryonic development. Cytogenet Genome Res 134 (4), 283–294. https://doi.org/10.1159/000329480.

Doyle, M.J., Magli, A., Estharabadi, N., Amundsen, D., Mills, L.J., Martin, C.M., 2019. *Sox7* regulates lineage decisions in cardiovascular progenitor cells. Stem Cells Dev. 28, 1089–1103. https://doi.org/10.1089/scd.2019.0040.

Engert, S., Burtscher, I., Liao, W.P., Dulev, S., Schotta, G., Lickert, H., 2013. Wnt/beta-catenin signalling regulates *Sox17* expression and is essential for organizer and endoderm formation in the mouse. Development 140, 3128–3138. https://doi.org/10.1242/dev.088765.

Graves, J.A.M., 1998. Interactions between SRY and SOX genes in mammalian sex determination. Bioessays 20, 264–269. https://doi.org/10.1002/(sici)1521-1878(199803)20:3.

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P., Lovellbadge, R., 1990. 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, 245–250. https://doi.org/10.1038/346245a0.

Han, F., Wang, Z., Wu, F., Liu, Z., Huang, B., Wang, D., 2010. Characterization, phylogeny, alternative splicing and expression of *Sox30* gene. BMC Mol. Biol. 11 (1), 98. https://doi.org/10.1186/1471-2199-11-98.

He, D., Marie, C., Zhao, C., Kim, B., Wang, J., Deng, Y., Clavairoly, A., Frah, M., Wang, H., He, X., Hmidan, H., Jones, B.V., Witte, D., Zalc, B., Zhou, X., Choo, D.I., Martin, D.M., Parras, C., Lu, Q.R., 2016. Chd7 cooperates with Sox10 and regulates the onset of CNS myelination and remyelination. Nat. Neurosci. 19, 678–689. https://doi.org/10.1038/nn.4258.

Heenan, P., Zondag, L., Wilson, M.J., 2016. Evolution of the *Sox* gene family within the chordate phylum. Gene 575, 385–392. https://doi.org/10.1016/j.gene.2015.09.013.

Herpers, R., van de Kamp, E., Duckers, H.J., Schulte-Merker, S., 2008. Redundant roles for *Sox7* and *Sox18* in arteriovenous specification in zebrafish. Circ. Res. 102, 12–15. https://doi.org/10.1161/circresaha.107.166066.

Hong, C.S., Saint-Jeannet, J.P., 2005. Sox proteins and neural crest development. Semin. Cell Dev. Biol. 16 (6), 694–703. https://doi.org/10.1016/j.semcdb.2005.06.005.

Howard-Ashby, M., Materna, S.C., Brown, C.T., Chen, L., Cameron, R.A., Davidson, E.H., 2006. Gene families encoding transcription factors expressed in early development of *Strongylocentrotus purpuratus*. Dev. Biol. 300, 90–107. https://doi.org/10.1016/j.ydbio.2006.08.033.

Hudson, C., Clements, D., Friday, R.V., Stott, D., Woodland, H.R., 1997. Xsox17 alpha and -beta mediate endoderm formation in Xenopus. Cell 91, 397–405. https://doi.org/10.1016/s0092-8674(00)80423-7.

Igarashi, H., Uemura, M., Hiramatsu, R., Hiramatsu, R., Segami, S., Pattarapanawan, M., Hirate, Y., Yoshimura, Y., Hashimoto, H., Higashiyama, H., Sumitomo, H., Kurohmaru, M., Saijoh, Y., Suemizu, H., Kanai-Azuma, M., Kanai, Y., 2018. *Sox17* is essential for proper formation of the marginal zone of extraembryonic endoderm adjacent to a developing mouse placental disk. Biol. Reprod. 99, 578–589. https://doi.org/10.1093/biolre/ioy079.

Jackson, H.E., Ono, Y., Wang, X., Elworthy, S., Cunliffe, V.T., Ingham, P.W., 2015. The role of Sox6 in zebrafish muscle fiber type specification. Skelet. Muscle 5 (1), 2. https://doi.org/10.1186/s13395-014-0026-2.

Jeng, S.-R., Wu, G.-C., Yueh, W.-S., Kuo, S.-F., Dufour, S., Chang, C.-F., 2018. Gonadal development and expression of sex-specific genes during sex differentiation in the Japanese eel. Gen. Comp. Endocrinol. 257, 74–85. https://doi.org/10.1016/j.ygcen.2017.07.031.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282. https://doi.org/10.1093/bioinformatics/8.3.275.

Kassahn, K.S., Dang, V.T., Wilkins, S.J., Perkins, A.C., Ragan, M.A., 2009. Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. Genome Res. 19, 1404–1418. https://doi.org/10.1101/gr.086827.108.

Kikuchi, Y., Agathon, A., Alexander, J., Thisse, C., Waldron, S., Yelon, D., Thisse, B., Stainier, D.Y.R., 2001. casanova encodes a novel Sox-related protein necessary and sufficient for early endoderm formation in zebrafish. Genes Dev. 15 (12), 1493–1505. https://doi.org/10.1101/gad.892301.

Koopman, P., 2005. Sex determination: a tale of two *Sox* genes. Trends Genet. 21 (7), 367–370. https://doi.org/10.1016/j.tig.2005.05.006.

Koopman, P., Schepers, G., Brenner, S., Venkatesh, B., 2004. Origin and diversity of the *Sox* transcription factor gene family: genome-wide analysis in *Fugu rubripes*. Gene 328, 177–186. https://doi.org/10.1016/j.gene.2003.12.008.

Lee, E.M., Yuan, T., Ballim, R.D., Nguyen, K., Kelsh, R.N., Medeiros, D.M., McCauley, D. W., 2016. Functional constraints on SoxE proteins in neural crest development: the importance of differential expression for evolution of protein activity. Dev. Biol. 418, 166–178. https://doi.org/10.1016/j.ydbio.2016.07.022.

Lefebvre, V., Dumitriu, B., Penzo-Mendez, A., Han, Y., Pallavi, B., 2007. Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. Int. J. Biochem. Cell Biol. 39, 2195–2214. https://doi.org/10.1016/j.biocel.2007.05.019.

Li, X., Yu, H., Wang, Y., Liu, X., Liu, Y., Qu, J., Wang, X., 2018. Roles of two *Sox9* genes during gonadal development in Japanese flounder: sex differentiation, spermatogenesis and gonadal function maintenance. Int. J. Mol. Sci. 19 (2), 512. https://doi.org/10.3390/ijms19020512.

Magie, C.R., Pang, K., Martindale, M.Q., 2005. Genomic inventory and expression of Sox and Fox genes in the cnidarian *Nematostella vectensis*. Dev. Genes Evol. 215, 618–630. https://doi.org/10.1007/s00427-005-0022-y.

Mallanna, S.K., Ormsbee, B.D., Iacovino, M., Gilmore, J.M., Cox, J.L., Kyba, M., Washburn, M.P., Rizzino, A., 2010. Proteomic analysis of Sox2-associated proteins during early stages of mouse embryonic stem cell differentiation identifies Sox21 as a novel regulator of stem cell fate. Stem Cells 28, 1715–1727. https://doi.org/10.1002/stem.494.

Matsuda, S., Kuwako, K.-i., Okano, H.J., Tsutsumi, S., Aburatani, H., Saga, Y., Matsuzaki, Y., Akaike, A., Sugimoto, H., Okano, H., 2012. Sox21 promotes hippocampal adult neurogenesis via the transcriptional repression of the *Hes5* gene. J. Neurosci. 32, 12543–12557. https://doi.org/10.1523/jneurosci.5803-11.2012.

Nakamoto, M., Suzuki, A., Matsuda, M., Nagahama, Y., Shibata, N., 2005. Testicular type *Sox9* is not involved in sex determination but might be in the development of testicular structures in the medaka, *Orydas latipes*. Biochem. Biophys. Res. Commun. 333 (3), 729–736. https://doi.org/10.1016/j.bbrc.2005.05.158.

Nakamura, S., Watakabe, I., Nishimura, T., Toyoda, A., Taniguchi, Y., Tanaka, M., 2012. Analysis of Medaka *sox9* orthologue reveals a conserved role in germ cell maintenance. PLoS One 7, e29982. https://doi.org/10.1371/journal.pone.0029982.

Navarro-Martin, L., Galay-Burgos, M., Piferrer, F., Sweeney, G., 2012. Characterisation and expression during sex differentiation of *Sox19* from the sea bass *Dicentrarchus labrax*. Comp Biochem Physiol B Biochem Mol Biol 163 (3–4), 316–323. https://doi.org/10.1016/j.cbpb.2012.08.004.

Okuda, Y., Kondoh, H., Kamachi, Y., 2010. B1 SOX coordinate cell specification with patterning and morphogenesis in the early zebrafish embryo. Dev. Biol. 344, 490–491. https://doi.org/10.1016/j.ydbio.2010.05.429.

Quiat, D., Voelker, K.A., Pei, J., Grishin, N.V., Grange, R.W., Bassel-Duby, R., Olson, E.N., 2011. Concerted regulation of myofiber-specific gene expression and muscle performance by the transcriptional repressor Sox6. Proc. Natl. Acad. Sci. U. S. A. 108, 10196–10201. https://doi.org/10.1073/pnas.1107413108.

Ribas, L., Robledo, D., Gomez-Tato, A., Vinas, A., Martinez, P., Piferrer, F., 2016. Comprehensive transcriptomic analysis of the process of gonadal sex differentiation in the turbot (*Scophthalmus maximus*). Mol. Cell. Endocrinol. 422, 132–149. https://doi.org/10.1016/j.mce.2015.11.006.

Santos, E.M., Workman, V.L., Paull, G.C., Filby, A.L., Van Look, K.J.W., Kille, P., Tyler, C. R., 2007. Molecular basis of sex and reproductive status in breeding zebrafish. Physiol. Genomics 30, 111–122. https://doi.org/10.1152/physiolgenomics.00284.2006.

Schepers, G.E., Teasdale, R.D., Koopman, P., 2002. Twenty pairs of *Sox*: extent, homology, and nomenclature of the mouse and human *Sox* transcription factor gene families. Dev. Cell 3, 167–170. https://doi.org/10.1016/s1534-5807(02)00223-x.

Shao, C., Li, C., Wang, N., Qin, Y., Xu, W., Liu, Q., Zhou, Q., Zhao, Y., Li, X., Liu, S., Chen, X., Mahboob, S., Liu, X., Chen, S., 2018. Chromosome-level genome assembly of the spotted sea bass, *Lateolabrax maculatus*. Gigascience 7 (11), 114. https://doi.org/10.1093/gigascience/giy114.

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovellbadge, R., Goodfellow, P.N., 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346, 240–244. https://doi.org/10.1038/346240a0.

Smith, E.K., Guzman, J.M., Luckenbach, J.A., 2013. Molecular cloning, characterization, and sexually dimorphic expression of five major sex differentiation-related genes in a Scorpaeniform fish, sablefish (*Anoplopoma fimbria*). Comp Biochem Physiol B Biochem Mol Biol 165, 125–137. https://doi.org/10.1016/j.cbpb.2013.03.011.

Solovyev, V., Kosarev, P., Seledsov, I., Vorobyev, D., 2006. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. 7 (Suppl. 1), S10.11–S10.12. https://doi.org/10.1186/gb-2006-7-s1-s10.

Spokony, R.F., Aoki, Y., Saint-Germain, N., Magner-Fink, E., Saint-Jeannet, J.P., 2002. The transcription factor Sox9 is required for cranial neural crest development in Xenopus. Development 129, 421–432.doi:https://doi.org/10.1159/000047931.

Stolt, C.C., Lommes, P., Friedrich, R.P., Wegner, M., 2004. Transcription factors Sox8 and Sox10 perform non-equivalent roles during oligodendrocyte development despite functional redundancy. Development 131, 2349–2358. https://doi.org/10.1242/dev.01114.

Sutton, E., Hughes, J., White, S., Sekido, R., Tan, J., Arboleda, V., Rogers, N., Knower, K., Rowley, L., Eyre, H., Rizzoti, K., McAninch, D., Goncalves, J., Slee, J., Turbitt, E., Bruno, D., Bengtsson, H., Harley, V., Vilain, E., Sinclair, A., Lovell-Badge, R., Thomas, P., 2011. Identification of *SOX3* as an XX male sex reversal gene in mice and humans. J. Clin. Invest. 121, 328–341. https://doi.org/10.1172/jci42580.

Takada, S., Mano, H., Koopman, P., 2005. Regulation of Amh during sex determination in chickens: Sox gene expression in male and female gonads. Cell. Mol. Life Sci. 62, 2140–2146. https://doi.org/10.1007/s00018-005-5270-5.

Uchikawa, M., Yoshida, M., Iwafuchi-Doi, M., Matsuda, K., Ishida, Y., Takemoto, T., Kondoh, H., 2011. B1 and B2 Sox gene expression during neural plate development in chicken and mouse embryos: universal versus species-dependent features. Develop. Growth Differ. 53, 761–771. https://doi.org/10.1111/j.1440-169X.2011.01286.x.

van Megen, H., van den Elsen, S., Holterman, M., Karssen, G., Mooyman, P., Bongers, T., Holovachov, O., Bakker, J., Helder, J., 2009. A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. Nematology 11 (6), 927–950. https://doi.org/10.1163/156854109x456862.

Voldoire, E., Brunet, F., Naville, M., Volff, J.-N., Galiana, D., 2017. Expansion by whole genome duplication and evolution of the *sox* gene family in teleost fish. PLoS One 12. https://doi.org/10.1371/journal.pone.0180936.

Vriz, S., Lovellbadge, R., 1995. 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 153, 275–276. https://doi.org/10.1016/0378-1119(94)00815-a.

Wan, H., Han, K., Jiang, Y., Zou, P., Zhang, Z., Wang, Y., 2019. Genome-wide identification and expression profile of the *Sox* gene family during embryo development in large yellow croaker, *Larimichthys crocea*. DNA Cell Biol. 38 (10), 1100–1111. https://doi.org/10.1089/dna.2018.4586.

Wang, H., Wen, H., Li, Y., Zhang, K., Liu, Y., 2018. Evaluation of potential reference genes for quantitative RT-PCR analysis in spotted sea bass (Lateolabrax maculatus) under normal and salinity stress conditions. PeerJ 6, e5631. https://doi.org/10.7717/peerj.5631.

Wegner, M., 1999. From head to toes: the multiple facets of Sox proteins. Nucleic Acids Res. 27 (6), 1409–1420. https://doi.org/10.1093/nar/27.6.1409.

Wegner, M., 2010. All purpose Sox: the many roles of Sox proteins in gene expression. Int. J. Biochem. Cell Biol. 42, 381–390. https://doi.org/10.1016/j.biocel.2009.07.006.

Wei, L., Cheng, D., Li, D., Meng, M., Peng, L., Tang, L., Pan, M., Xiang, Z., Xia, Q., Lu, C., 2011. Identification and characterization of Sox genes in the silkworm, *Bombyx mori*. Mol. Biol. Rep. 38, 3573–3584. https://doi.org/10.1007/s11033-010-0468-5.

Wei, L., Yang, C., Tao, W., Wang, D., 2016. Genome-wide identification and transcriptome-based expression profiling of the Sox gene family in the Nile Tilapia (*Oreochromis niloticus*). Int. J. Mol. Sci. 17, 270. https://doi.org/10.3390/ijms17030270.

Wen, H.S., Li, J.F., Zhang, M.Z., Li, Y., Qi, X., Zhang, K.Q., Liu, Y., Zhang, X.Y., Tian, Y., Wang, J.H., 2019. Physiology and Breeding Technology of Sea Bass in Mariculture. China Agricultural Press, Beijing, China, pp. 337–341.

Wilson, M.J., Dearden, P.K., 2008. Evolution of the insect Sox genes. BMC Evol. Biol. 8 (1), 120. https://doi.org/10.1186/1471-2148-8-120.

Wolpert, L., 1992. Gastrulation and the evolution of development. Development (Cambridge) 0, 7–13.

Xiao, J., Cao, K., Zou, Y., Xiao, S., Wang, Z., Cai, M., 2019. Sex-biased gene discovery from the gonadal transcriptomes of the large yellow croaker (*Larimichthys crocea*). Aquac Fish 4, 9–16. https://doi.org/10.1016/j.aaf.2019.01.001.

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G., Zhang, Y., Huo, L., Zhao, Z., Cao, D., Lu, C., Li, C., Zhou, Y., Liu, Z., Fan, Z., Shan, G., Li, X., Wu, S., Song, L., Hou, G., Jiang, Y., Jeney, Z., Yu, D., Wang, L., Shao, C., Song, L., Sun, J., Ji, P., Wang, J., Li, Q., Xu, L., Sun, F., Feng, J., Wang, C., Wang, S., Wang, B., Li, Y., Zhu, Y., Xue, W., Zhao, L., Wang, J., Gu, Y., Lv, W., Wu, K., Xiao, J., Wu, J., Zhang, Z., Yu, J., Sun, X., 2014. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. Nat. Genet. 46, 1212–1219. https://doi.org/10.1038/ng.3098.

Yao, B., Zhou, L., Wang, Y., Xia, W., Gui, J.-F., 2007. Differential expression and dynamic changes of SOX3 during gametogenesis and sex reversal in protogynous hermaphroditic fish. J Exp Zool A Ecol Integr Physiol 307A, 207–219. https://doi.org/10.1002/jez.361.

Yu, H., Du, X., Li, X., Qu, J., Zhu, H., Zhang, Q., Wang, X., 2018. Genome-wide identification and transcriptome-based expression analysis of sox gene family in the Japanese flounder Paralichthys olivaceus. Journal of Oceanology and Limnology 36, 1731–1745. https://doi.org/10.1007/s00343-018-7216-4.

Zhang, S., Chen, X., Wang, M., Zhang, W., Pan, J., Qin, Q., Zhong, L., Shao, J., Sun, M., Jiang, H., Bian, W., 2018. Genome-wide identification, phylogeny and expressional profile of the *Sox* gene family in channel catfish (*Ictalurus punctatus*). Comp Biochem Physiol Part D Genomics Proteomics 28, 17–26. https://doi.org/10.1016/j.cbd.2018.03.001.

Zheng, J., Jia, Y., Liu, S., Chi, M., Cheng, S., Gu, Z., 2020. Molecular characterization and expression profiles of transcription factor *Sox* gene family in *Culter alburnus*. Gene Expr. Patterns 36. https://doi.org/10.1016/j.gep.2020.119112.