Pakistan J. Zool., vol. 50(2), pp 417-428, 2018. **DOI**: http://dx.doi.org/10.17582/journal.pjz/2018.50.2.417.428

Identification, Molecular Characterization and Expression Pattern Analysis of *SoxD* Subgroup Genes in Yellow River Carp (*Cyprinus carpio*)

RuiHua Zhang, YongFang Jia, TingTing Liang, QianWen Yang, QiYan Du and ZhongJie Chang*

Molecular and Genetic Laboratory, College of Life Science, Henan Normal University, 46# East of Construction Road, Xinxiang, Henan 453007, P.R. China

ABSTRACT

SoxD subgroup genes, which belong to the Sox transcription factor family, have been implicated in the developing nervous system. Their expression has been seen in neural stem cells as well as differentiating neurons. However, despite their importance in development, a relatively low number have been characterized for freshwater fish. In this study, we were able to gain three full-length *SoxD* sequences from transcriptome sequencing data of carp: Cc*Sox5*, Cc*Sox6* and Cc*Sox13 (Cyprinus carpio SoxD)*. These genes were verified by sequencing, blast and homology alignment. *CcSoxD* have two to four exons and encode a 780, 784 and 602 amino acid protein, respectively. Chromosome synteny analyses revealed that *CcSox5 and CcSox13* were tightly linked with the *etnk* gene, which was conserved in all species; however, there were no conserved regions flanking *CcSox6*. Numerous essential transcription factor binding sites (TFs) were predicted within the 2000 bp upstream of the 5' end of these genes. These TFs include BSX, BRN4 and NGN–NEUROD, which have been shown to be involved in the early stages of neuronal determination and neurogenesis in vertebrates. Tissue distribution analyses by Quantitative real-time RT-PCR (qRT-PCR) revealed that *CcSoxD* genes were abundant in the brain, showed sexual dimorphism, and were inconsistently expressed during embryogenesis. These results indicated that *CcSoxD* plays an important role in the development of the nervous system and may be involved in sexual development in carp. And they provide a foundation for further study of the function of *CcSoxD* genes during carp development and neurogenesis.

Article Information
Received 13 March 2017
Revised 02 May 2017
Accepted 08 June 2017
Available online 24 January 2018

Authors' Contribution
RZ, YJ and ZC designed research. RZ, YJ, TL and QY performed experiments and contributed new reagents and analytic tools. RZ and YJ analyzed the data. RZ, YJ, QD and ZC wrote the paper.

Key words
Carp, *CcSoxD*, Gene structure, Chromosome synteny, Neurogenesis.

INTRODUCTION

The sex-determining region (*SRY*)-related box (*Sox*) genes encode a group of transcription factors with a high mobility group (HMG)-type DNA-binding domain which consists of three α-helixes, permitting the proteins to bind to the minor groove of DNA and bend it (She and Yang, 2015; Daigle *et al*., 2015). Since their initial discovery in mice, there have been more than thirty *Sox* gene members isolated from various species, including tetrapods, fishes and insects (Lefebvre *et al*., 2007). The members of this family can be further divided into eight subgroups A to H based on the level of amino acid conservation within the HMG box and the presence of other motifs (She and Yang, 2015). Sox proteins were conserved during vertebrate evolution, and expressed in numerous tissues and regulated a variety of developmental processes (Wei *et al*., 2016). Although most Sox proteins fall predominantly into the transcriptional activator family, there is also evidence for transcriptional repression and architectural roles for these genes (Wegner, 2010). Sox proteins have been shown to play essential role and undertake key functions in cell fate decisions in neurogenesis, sex determination and gliogenesis, neural crest development, skeletogenesis, cardiogenesis and angiogenesis as well as in hematopoiesis (Lefebvre *et al*., 2007; Wegner, 2010). Their known functions have previously been compiled and reviewed by Bowles *et al*. (2000).

This paper focuses on the *SoxD* subgroup, which is composed of three genes *Sox5*, *Sox6*, and *Sox13* in most vertebrates and invertebrates. *SoxD* members are known to play a key role in multiple developmental pathways, including the development of the central nervous system (Lefebvre, 2010; Kiselak *et al*., 2010; Baroti *et al*., 2015; Reiprich and Wegner, 2015) and cartilage formation (Liu and Lefebvre, 2015).

During development, *Sox5* is expressed in subsets of cells in the central nervous system (CNS), cranial ganglia, neural crest and skeletal/cartilage tissues. Like *Sox5*, *Sox6* is expressed in CNS and skeletal/cartilage tissues, but in addition, it is also seen in cardiac myocytes and erythroid cells (Lefebvre, 2010; Hagiwara, 2011). *Sox13* is expressed in cells of the developing CNS and cartilage progenitors (Lefebvre, 2010; Wang *et al*., 2005), and is detected in the

* Corresponding author: 13837331530@163.com
0030-9923/2018/0002-0417 $ 9.00/0

developing artery, inner ear, hair follicle and a subset of T cells, either (Lefebvre, 2010; Wang *et al*., 2005; Roose *et al*., 1998; Melichar *et al*., 2007). The best characterized functions of *SoxD* genes demonstrate their involvement in cell fate determination and differentiation. However, the mechanisms underlying these specific expression patterns of the *SoxD* genes are virtually unknown.

The carp is an important freshwater commercial fish in China. The nervous system participates in the regulation of locomotion, food intake, injury repair as well as sexual differentiation amongst others. Therefore, understanding the structure, function and regulatory mechanisms of related genes in this specie is of both scientific and commercial interest. In this study, the *CcSoxD* (*SoxD* subgroup of *Cyprinus carpio*) genes of carp were identified and characterized by bioinformatics. The expression pattern of *SoxD* in five embryo development stages, thirteen adult tissues, and five different parts of the fish brain were obtained for further investigation and functional analysis.

MATERIALS AND METHODS

Animal and sample collection

The carp were obtained from Henan Provincial Research Institute of Aquaculture. Artificially fertilized eggs were incubated at 23 ± 2°C in hatching tanks with an open recirculation water system and continuous aeration. Samples from different embryonic stages were observed under a microscope to determine the particular developmental stage. Thirteen carp tissues (heart, liver, kidney, hindbrain, spleen, foregut, hindgut, muscle, gill, eye, scale, fin and gonad) were collected from six 2-year-old healthy adults (three females and three males), and five critical periods of embryonic development (blastula, gastrula, neurula, tail-bud and hatching) were obtained from embryos. Whole brains from adult carp were dissected and five parts of the brain (telencephalon, diencephalon, mesencephalon, epencephalon and macromyelon) were carefully separated. Tissue samples were immediately frozen in liquid nitrogen and stored at −80°C for further analysis. Animal experiments were performed according to the Regulations for the Administration of Affairs Concerning Experimental Animals.

Total RNA extraction

Total RNA samples were extracted from multiple tissues of carp and different stages of embryogenesis using RNA extraction kit and RNAiso reagent (Takara, Japan) according to the manufacturer's instructions. DNA contamination was removed by RNase-free DNase I (Takara) treatment. The concentration and purity of the total RNA were determined by electrophoretic gel imaging and spectrophotometry. RNA samples with a 28S: 18S ratio of approximately 2:1 and an OD260/OD280 ratio of 1.9-2.2 were considered of sufficient quality for further experimentation. The cDNA used for qRT-PCR was synthesized using PrimeScript Reverse Transcriptase reagent Kit (TaKaRa) according to the manufacturer's instructions.

Table I.- Primers used for verification.

Primers	Sequence(5′-3′)	Prod. size
Sox5-Fver-F	5' TGCATCTCAGACCCCTTGTT 3'	839
Sox5-Fver-R	5' CTGGACCTGCTGGATCTGTT 3'	
Sox5-Mver-F	5' AACTTCTGCAGCAACAACAC 3'	806
Sox5-Mver-R	5' TTATCCTTCTCTGAGCGACC 3'	
Sox5-Rver-F	5' CAAAGTAGCAGCAGTCAACAG 3'	868
Sox5-Rver-R	5' TGGTGAGATGGCTGTGATTGG 3'	
Sox6-Fver-F	5' TTGGGAGCTGGAGATAAAGT 3'	852
Sox6-Fver-R	5' TTGATTTTGTGCTGTTGCTG 3'	
Sox6-Mver-F	5' CACGCCAACAGCAAGAGCA 3'	872
Sox6-Mver-R	5' CCAATCTTTCCATCCTCGCC 3'	
Sox6-Rver-F	5' GGCAAACTAGGCGAGGATGG 3'	796
Sox6-Rver-R	5' GGATATTCTTCGCTGGCTGT 3'	
Sox13-Fver-F	5' ACGACGGAAAGACTGAAGGA 3'	1002
Sox13-Fver-R	5' CGGTACCCTGCTTGTAAGTG 3'	
Sox13-Rver-F	5' ACTGGAAATGGCCCACTTAC 3'	935
Sox13-Rver-R	5' CTCCAATCCTTTCGTCCTTCT 3'	

Bioinformatics and sequence analysis

Using a thorough search of the *de novo* transcriptome sequencing data of carp constructed by our laboratory, we discovered several sequence fragments which had high sequence similarity to the zebrafish *SoxD* genes. After assembling the sequences, we were able to construct the full-length mRNA sequences of *CcSoxD*. We then designed specific primers (Table I) to sequencing and verify the correctness of the sequences. Nucleotide sequence similarity analysis of the candidate *SoxD* genes was performed using BLAST software (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Homologous nucleotide and protein sequences were confirmed using the BLASTn and BLASTx search algorithm in NCBI (http://www.ncbi.nlm.gov/blast). Multiple alignments of amino acid sequences were performed using the online program ClustalW (http://www.genome.jp/tools/clustalw/) and DNAman programs. A phylogenetic tree was constructed using MEGA6 software based on the results of the protein sequence alignments. The deduced amino acid sequence was analyzed using DNAman to predict conserved domains.

Bioinformatics analysis of the promoter sequences and potential transcription factor binding sites within the 5' regulatory region of *CcSox5, CcSox6* and *CcSox13* were performed using the online program MatInspector (http://www.genomatix.de/matinspector.html). The data on Chromosome synteny and the carp genome were gathered from NCBI (https://www.ncbi.nlm.nih.gov/gene/).

Quantitative real-time PCR

The expression patterns of *CcSox5*, *CcSox6* and *CcSox13* in five critical periods of embryonic development, thirteen tissues and five parts of brain in female and male adult carp were measured by qRT-PCR. The expression levels of target genes were normalized to the levels of reference genes *40s* rRNA and *GAPDH* (Zhang *et al*., 2016). The cDNA templates used for qRT-PCR analysis were generated using the method described above. Primers for *CcSox5*, *CcSox6* and *CcSox13* for qRT-PCR (Table II) were designed outside the conserved domains to prevent any non-specific amplification, and meanwhile avoid hairpin and cross dimer. In addition, each cloned gene of the Yellow River carp presented a single band of expected size in 1% agarose gel electrophoresis and the melting curve appeared as a single peak. QRT-PCR was performed in a 10 µL reaction volume using quantitative real-time PCR detection system (LightCycler 96® Roche). Amplifications were conducted in a reaction mixture of 10 µL containing 5 µL of SYBR Premix 5x Taq (Takara), 0.2µL of each primer, 0.4 µL of diluted cDNA and 4.2 µL of H_2O, each assay was performed in triplicate. Expression levels were analyzed using the $2^{-\Delta\Delta CT}$ method. The data were expressed as the mean of RQ value ($2^{-\Delta\Delta CT}$) (ΔCT = CT of target gene minus CT of *40s* rRNA (or *GAPDH*), ΔΔCT = ΔCT of any sample minus calibrator sample) and analyzed with SPSS 15.0 using a one-way analysis of variance (ANOVA) and t-test. Statistically significant difference was set at $p < 0.05$.

Table II.- Primers used for quantitative real-time PCR.

Primers	Sequence(5′-3′)	Prod. size
40s-F	5' CCGTGGGTGACATCGTTACA 3'	119
40s-R	5' TCAGGACATTGAACCTCACTGTCT3'	
Gapdh- F	5' CCGTTCATGCTATCACAGCTACACA3'	159
Gapdh- R	5' CAGTAAGCTTGCCATTGAGCTC 3'	
Sox5-F	5' ACCTGCTATCATCCATCACC 3'	173
Sox5-R	5' TTCTCTGAGCGACCATTGTT 3'	
Sox6-F	5' AGCGCTGTTTGGAGATCAGG 3'	191
Sox6-R	5' CTCGCCTAGTTTGCCCAGGT 3'	
Sox13-F	5' CTTGAAGACGACGAATCAGG 3'	138
Sox13-R	5' GGAGATTGTGCATTAGGTGG 3'	

RESULTS

CDNA sequence analysis of Sox5, Sox6 *and* Sox13 *from carp*

Initially, we obtained three sequences from the carp transcriptome sequencing data. These were shown to be homologous to other *Sox5*, *Sox6* and *Sox13* genes using a BLAST search. This was then verified using sequence specific primers and sequencing of the carp genome. A putative 2800 bp *CcSox5* gene contained a 5' untranslated region (UTR) of 188 bp, a 3' UTR of 269 bp and an open reading frame (ORF) of 2343 bp. The ORF encoded a 780 amino acid protein. Similar to other Sox proteins, the predicted *CcSox5* contained a characteristic HMG-box DNA binding domain of 72 amino acids between positions 575 and 646. The nucleotide sequence and deduced amino acid sequence are shown in (Supplementary Fig. 1A)

The nucleotide sequence analysis indicated that the full-length cDNA of *CcSox6* is 3000 bp and was composed of a 555 bp 5' UTR, a 90 bp 3' UTR and a 2355bp open reading frame (ORF) that encoded a 784 amino acid protein. This putative ORF also contained the conserved characteristic HMG-box DNA binding domain, with the 72 amino acid motif appearing at positions 575 to 646 (Supplementary Fig. 1B).

For *CcSox13*, the full length cDNA sequence was assembled and demonstrated to be 2366 bp, with a 1809 bp ORF encoding a 602 amino acid protein, a 263 bp 5' UTR and a 294 bp 3' UTR. The conserved 72 amino acid HMG box domain was positioned between nucleotide 413 and 484 (Supplementary Fig. 1C).

Alignment and phylogenetic analysis

A multiple sequence alignment of the *CcSoxD* genes was assembled using sequence homology between these putative genes and known *SoxD* family members from different vertebrates including teleosts, amphibians, reptiles, birds and mammals using DNAman and ClustalW. Results showed that the predicted amino acid sequences of *CcSox5* shared higher identities with zebrafish *Sox5* (87.44%) and rainbow trout *Sox5* (74.27%), and lower identities with human, mouse, chicken and African clawed frog *Sox5* (61.71%−56.79%) (Table III). While the predicted amino acid sequence of *CcSox6* shared higher identities with zebrafish *Sox6* (86.66%) and Channel Catfish *Sox6* (73.29%), and lower identities with human, mouse, chicken and frog *Sox6* (52.43%−57.84%) (Table III). *CcSox13* showed total amino acid identities of 100% and 72.06% with zebrafish *Sox13* and Channel Catfish *Sox13* (Table III).

Table III.- Amino acid sequence percent identities of *CcSox5*, *CcSox6* and *CcSox13* compared to other vertebrates SoxD proteins.

	Cc	Dr	Rt	To	Cs	Ol	Xl	Gg	Mus	Hs
Sox5										
Cc	100%									
Dr	87.44%	100%								
Rt	74.27%	80.80%	100%							
To	67.99%	69.34%	70.21%	100%						
Cs	67.38%	68.68%	72.36%	77.11%	100%					
Ol	66.37%	67.93%	64.82%	71.08%	66.90%	100%				
Xl	56.79%	59.51%	59.51%	56.91%	54.57%	49.51%	100%			
Gg	57.39%	60.20%	57.39%	55.56%	57.06%	47.50%	70.33%	100%		
Mus	57.80%	60.58%	59.51%	56.95%	57.06%	49.51%	76.83%	83.66%	100%	
Hs	61.71%	69.07%	66.37%	59.62%	53.59%	50.59%	83.38%	91.32%	92.80%	100%
Sox6										
Cc	100%									
Dr	86.66%	100%								
Ip	73.29%	74.50%	100%							
Km	68.21%	70.42%	68.76%	100%						
Xm	67.88%	69.76%	70.42%	81.35%	100%					
Cs	67.32%	57.84%	57.73%	54.53%	54.42%	100%				
Xl	57.84%	61.70%	59.93%	57.84%	56.62%	50.99%	100%			
Gg	51.88%	56.29%	55.63%	52.32%	52.87%	45.25%	64.05%	100%		
Mus	52.43%	56.73%	55.96%	52.54%	53.31%	45.25%	63.16%	81.08%	100%	
Hs	52.43%	56.84%	54.19%	52.98%	52.21%	45.58%	63.61%	79.98%	86.39%	100%
Sox13										
Cc	100%									
Dr	100%	100%								
Ip	72.06%	72.06%	100%							
Cs	60.57%	60.57%	53.26%	100%						
Pr	61.19%	61.19%	57.51%	67.58%	100%					
Ol	56.89%	56.89%	58.36%	68.25%	72.03%	100%				
Tr	51.97%	51.97%	45.14%	60.75%	61.09%	60.95%	100%			
Gg	49.84%	49.84%	44.74%	46.24%	47.01%	48.28%	42.81%	100%		
Mus	47.77%	47.77%	45.08%	46.01%	45.41%	46.23%	41.98%	71.34%	100%	
Hs	47.71%	47.71%	48.54%	44.68%	44.54%	46.83%	41.30%	74.28%	87.30%	100%

Cc, *Cyprinoid carp;* Dr, *Danio rerio;* Rt, Rainbow trout; To, *Takifugu obscurus;* Cs, *Cynoglossus semilaevis;* Ol, *Oryzias latipes;* Xl, *Xenopus laevis;* Ip, *Ictalurus punctatus;* Pr, *Poecilia reticulata;* Tr, *Takifugu rubripes;* Km, *Kryptolebias marmoratus;* Xm, *Xiphophorus maculatus;* Gg, *Gallus gallus;* Mus, *Mus musculus;* Hs, *Homo sapiens.*

To predict the evolutionary relationships between the *CcSoxD* genes and other species' *SoxD*, a phylogenetic tree was constructed based on the full-length amino acid sequences using the neighbor-joining method. The Sox proteins were grouped into two distinct clades, *Sox5* and *Sox6,* and a distinct *Sox13* subgroup. Furthermore, *CcSox5*, *CcSox6* and *CcSox13* were all most closely related to the teleost fish, and then with the tetrapods (Fig. 1).

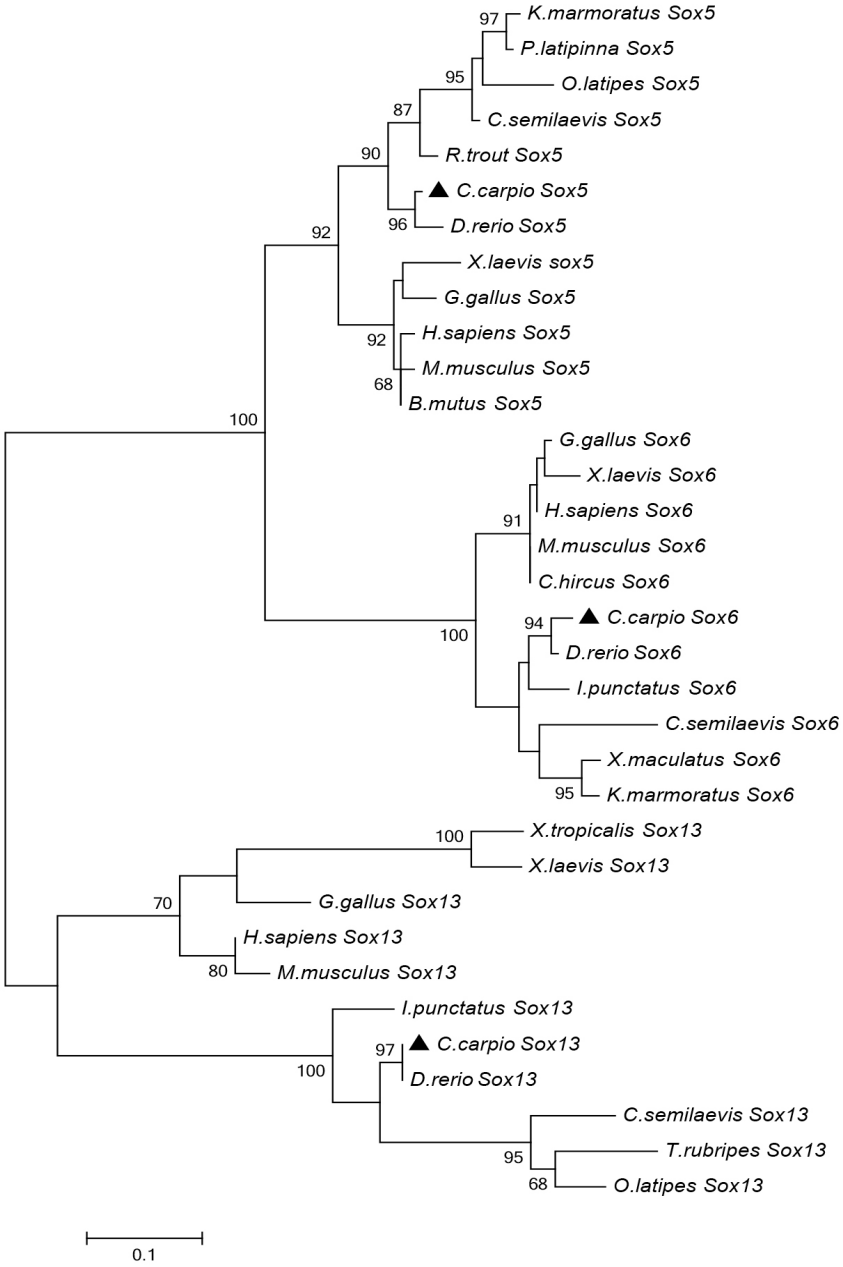

Fig. 1. Phylogenetic tree of *CcSox5*, *CcSox6* and *CcSox13* in comparison with *SoxD* proteins in other representative vertebrates using predicted amino acid sequences. The phylogenic tree was constructed by MEGA6 using the Jones–Thornton–Taylor (JTT) model based upon Neighbor-Joining method with 1000 bootstrap replicates. The scale bar is 0.1. The GenBank accession numbers are as follows: *Homo sapiens*: *Sox5*, BC060773.1; *Sox6*, AF309034.1; *Sox13*, NM_005686.2; *Mus musculus*: *Sox5*, AB006330.1; *Sox6*, U32614.1; *Sox13*, NM_011439.2; *Gallus gallus*; *Sox5*, NM_001004385.1; *Sox6*, XM_015286512.1; *Sox13*, XM_004934897.2; *Bos taurus*: *Sox5*, NM_001083471.1; *Bos mutus*: *Sox5*, XM_005897700.2; *Capra hircus:* *Sox6*, :XM_018059599.1; *Oryzias latipes*: *Sox5*, NM_001122910.1; *Sox13*, XM_011475310.1; *Xenopus laevis*: *Sox5*, AB682776.1; *Sox6*, NM_001280658.1; *Sox13*, NM_001087769.1; *Danio rerio*: *Sox5*, NM_001033585.1; *Sox6*, NM_001123009.1; *Xenopus tropicalis*: *Sox13*, XM_012963174.2; *Kryptolebias marmoratus*: *Sox5*, XM_017429903.1; *Sox6*, XM_017441909.1; *Poecilia latipinna*; *Sox5*, XM_015049850.1; *Cynoglossus semilaevis*; *Sox5*, XM_008315905.2; *Sox6*, XM_008311513.2; *Sox13*, XM_017036517.1; *Ictalurus punctatus*: *Sox6*, XM_017478071.1; *Oncorhynchus mykiss*: *Sox5*, FJ713023.1; *Xiphophorus maculatus*: *Sox6*, XM_005805573.1; *Ictalurus punctatus*: *Sox13*, XM_017450839.1

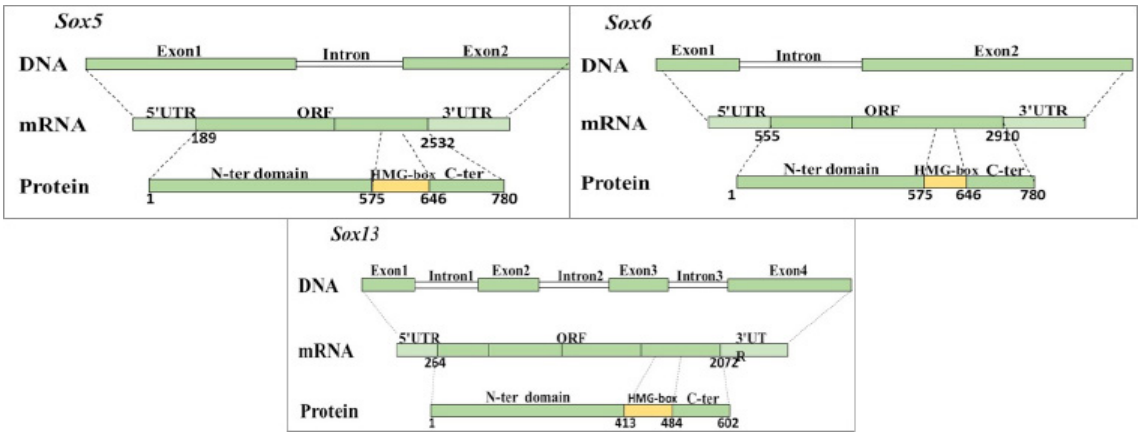

Fig. 2. Schematic presentation of *CcSox5*, *CcSox6* and *CcSox13*. The intron (thin box), 5' and 3' UTR (light green), and ORF (dark green) encoding the amino acid sequences are shown relative to their lengths in the cDNA sequences obtained. Protein domains are shown relative to their lengths and positions in the amino acid sequences. N-terminal domain (dark green); HMG-box, high mobility group box domain (orange); C-terminal domain (dark green).

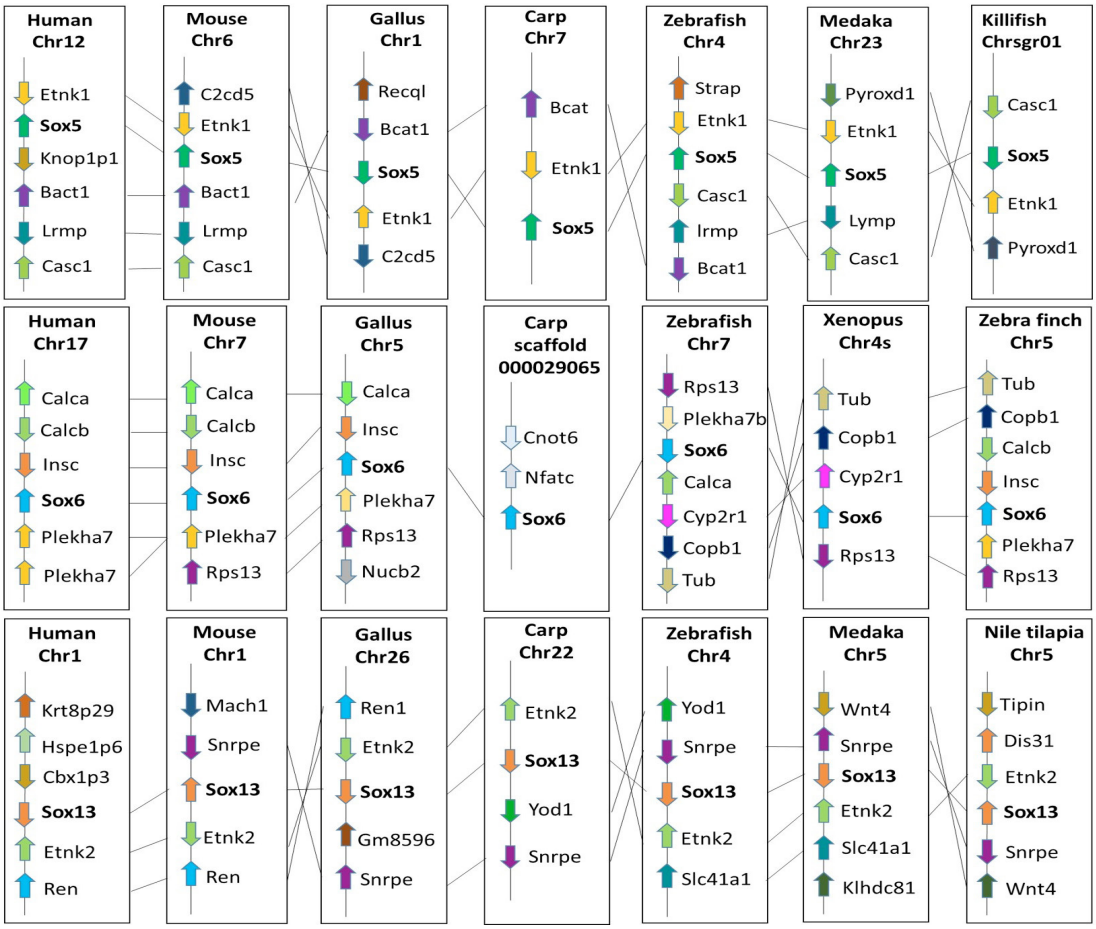

Fig. 3. Chromosome syntenic relationship of the *CcSox5*, *CcSox6* and *CcSox13* genes with their orthologs. Conserved syntenies are shown for chromosomal segments containing *Sox5*, *Sox6* and *Sox13*. Rectangles represent genes in chromosome/scaffold and arrows represent gene-coding direction. Chr, chromosome.

Chromosome synteny and genomic analysis

CcSox5 and *CcSox6* were confirmed to contain two exons and a single intron when comparing the cDNA sequences with the genomic DNA sequences. While *CcSox13* was shown to contain four exons and three introns using the same method (Fig. 2). The introns were varying lengths, with a 61 bp intron in *CcSox5*, 981 bp intron in *CcSox6*, and introns of 1262 bp, 1621 bp and 118 bp in *CcSox13*, respectively. All exon-intron boundaries were conformed using the GT and AG splicing rule.

Further, based on up-to-date carp whole-genome sequencing data, a cross-species comparison of chromosome locations were applied to determine the homologous relationship between *CcSoxD* and other *SoxD* genes. Results revealed that *CcSox5* was on Chr7 (chromosome 7) flanked by *etnk1*, *CcSox13* was flanked by *etnk2* and *yod1* on Chr22, and *CcSox6* was on scaffold000029065 flanked by *nfatc* (Fig. 3). The chromosome syntenic relationships were highly conserved during evolution in the human and mouse, but gene rearrangement is common in fish. There is, however, a close linkage between *etnk* and *CcSox5/CcSox13* genes in different species. The flanking region of *Sox6* in carp, however, was shown to be unique from zebrafish, human and mouse.

For promoter analysis, we searched the Common Carp Genome Database (http://www.carpbase.org/) and BLASTn (http://www.ncbi.nlm.nih.gov/BLAST/), the 2000 bp upstream of the translational initiation site (ATG) was selected as the promoter region for futher analysis. The ATG was designated as +1 and 2000 bp upstream flanking sequences of *CcSox5*, *CcSox6* and *CcSox13* were analyzed by MatInspector. Numerous essential transcription factor binding sites (TFs) were predicted within the 5' regulatory region and those with a matrix score higher than 0.90 were drawn on the schematic diagram (Fig. 4). Some of these TFs, including BSX, BRN4, and NGN-NEUROD are involved in neurogenesis. TFs like Oct4, Nanog and FOXP1 have been linked to various pluripotency or stem cell properties. Some binding sites, which promote gene expression like AP1, CEBPB, NF-Y, and SF1, were also identified.

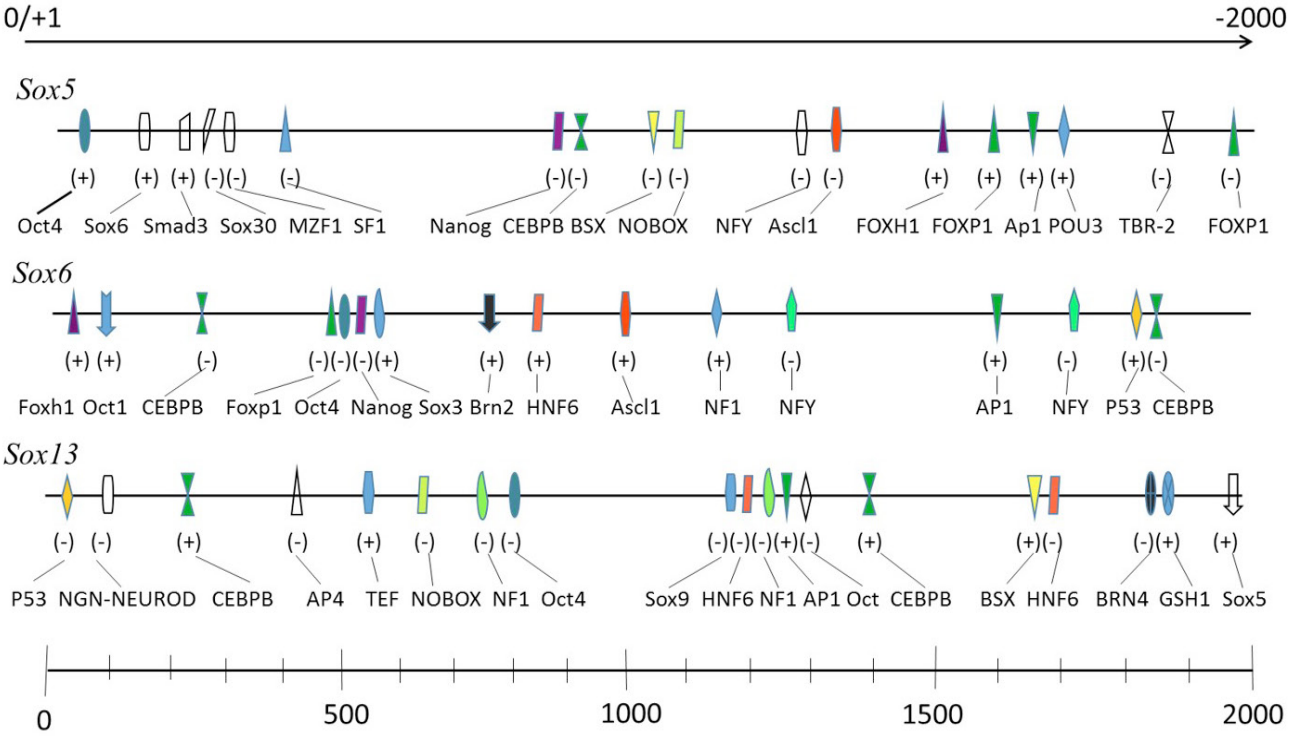

Fig. 4. A schematic diagram of putative regulatory motifs in the promoter of *CcSox5*, *CcSox6* and *CcSox13*. The scale is given above and the full names of the potential TF binding sites are provided at the bottom. Transcriptional initiate site (ATG) is designated as +1. BSX, Brain specific homeobox; BRN4, POU domain transcription factor brain 4; NGN–NEUROD, Neurogenin and NeuroD; Oct4, POU domain, class 5, transcription factor 1; Nanog, Homeobox transcription factor Nanog; FOXP1, Alternative splicing variant of FOXP1, activated in ESCs; AP1, Activator protein 1; CEBPB, CCAAT/enhancer binding protein beta; NF-Y, Nuclear factor Y; SF1, Steroidogenic Factor-1.

Embryo expression analysis of CcSoxD

The expression pattern of *CcSox5*, *CcSox6* and *CcSox13* during early embryonic development were analyzed by qRT-PCR using *40s* RNA and *GAPDH* as reference genes. The results revealed that *Sox5* transcript was detected at very low levels in embryos at early stages up to the gastrula, but was then upregulated and reached peak expression in the neurula stage, and then was slightly downregulated between tail-bud and hatching. *Sox13* showed an initial increase in expression at the blastula followed by a decline in expression during the in the rest of the embryonic development process. *Sox6* was maintained at very low levels during the whole embryonic development process (Fig. 5).

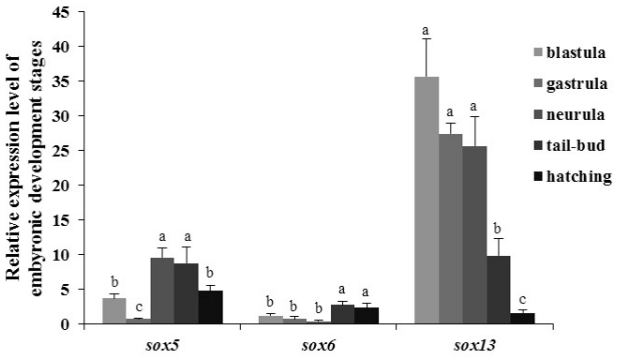

Fig. 5. Relative expression of *CcSox5*, *CcSox6* and *CcSox13* genes during embryonic development by qRT-PCR analysis. Error bars represent the mean ± SEM (n=6). Significant differences ($p < 0.05$) exist between any two samples labeled with different single letters.

Expression analysis of CcSoxD *in different tissues of both male and female fish*

We also analyzed the expression of *CcSoxD* genes in various adult tissues of both male and female fish. Expression analysis revealed that *CcSox5* was predominantly expressed in the brain, with low levels of expression in the other adult tissues including the eye, gill and heart, and was undetected in other tissues like the fin, liver, muscle and scale in both male and female samples. *CcSox5* showed obvious sexual dimorphism in the brain. *CcSox6* was abundantly expressed in the brain and muscle tissues, moderately expressed in the eye, heart, liver and spleen, and showed decreased expression in the fin, intestine, scale and gonads. *Sox13* was expressed at a high level in the brain, and its expression in the eye, gill and kidney in male fish and the eye and spleen in female fish were also high when compared with the other tissues whose expression levels were especially low. For *CcSox13*, sexual dimorphism was found in the gill, kidney, scale and spleen. In all cases, *CcSoxD* gene expression was relatively high in the brain and lower in the other studied tissues (Fig. 6).

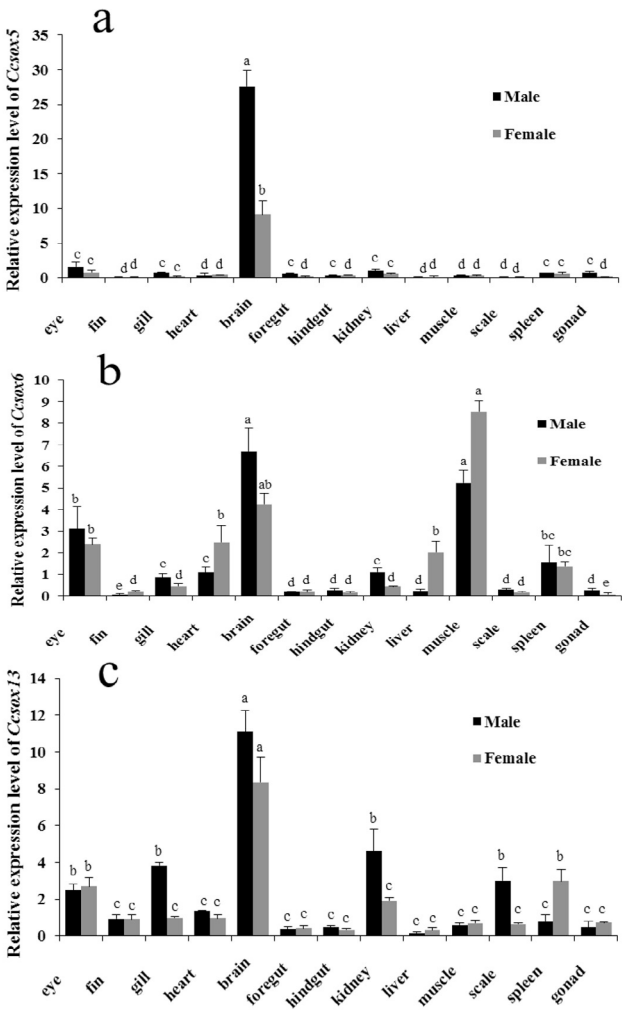

Fig. 6. Differential expression of male and female tissues of *CcSox5*, *CcSox6* and *CcSox13* genes (a, b, c) by qRT-PCR analysis. Error bars represent the mean ± SEM (n=6). Significant differences ($p < 0.05$) exist between any two samples labeled with different single letters.

Expression pattern of CcSoxD *in adult brains*

Because these three *Sox* genes were highly expressed in the brain compared to other tissues, we made a detailed analysis of their expression levels in five parts of the brain. *Sox5*, *Sox6*, and *Sox13* showed different expression patterns. *Sox5* transcript was abundantly expressed in the epencephalon and mesencephalon with a slightly lower level in the diencephalon, telencephalon, and macromyelon. The highest expression of the *Sox6* transcript was detected in the mesencephalon, with lower

levels in the telencephalon, epencephalon, macromyelon and diencephalon. *Sox13* was moderately expressed in all the regions analyzed (Fig. 7).

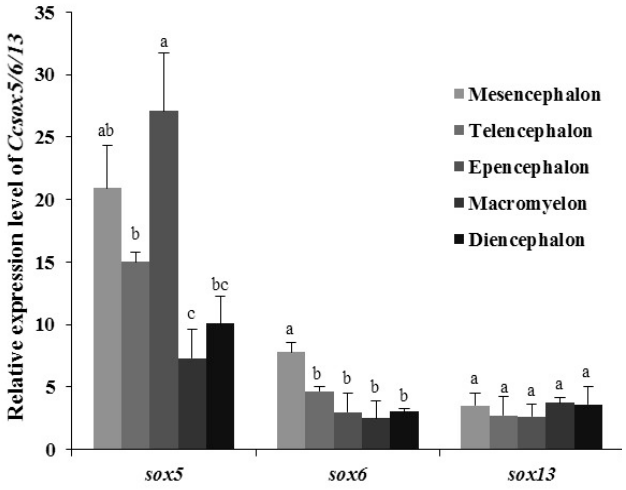

Fig. 7. Relative expression of *CcSox5*, *CcSox6* and *CcSox13* genes in different parts of the adult brain by qRT-PCR analysis. Error bars represent the mean ± SEM (n=6). Significant differences ($p < 0.05$) exist between any two samples labeled with different single letter.

DISCUSSION

Since the first discovery of the pluripotent *Sox* genes in mammalian tissues (Lefebvre *et al*., 2007), studies of the *Sox* gene have been undertaken in amphibians (Nordin and LaBonne, 2014), fish (Gao *et al*., 2015), and reptiles (Xin *et al*., 2012) amongst others. However, research methods have generally focused on a single *Sox* gene and most studies have focused on the human, mouse, zebrafish and the other model species. Although studies about sex determination in human and other mammals have been studied more closely (Ludbrook *et al*., 2016), there is some basic evidence to support the roles of the *Sox* genes in sex determination in fish; however, this is an area of research that is still under developed.

Multiple forms of *Sox* cDNAs have been reported from a variety of vertebrate species, including mammal, bird and fish. We are interested in the *SoxD* genes in carp, as *SoxD* genes have a well-established role in neurogenesis in other vertebrates, and play an important role in various aspects of development including cell fate specification (Lefebvre, 2010). In this study, we identified three full-length cDNAs encoding three different *CcSoxD* variants using a transcriptome library. We analyzed their mRNA expression pattern during embryogenesis and in adult tissues as well as in five parts of the brain. This is the first report that describes *SoxD* genes from this specie, and the data presented constitute a relatively reliable foundation for the evaluation of the *SoxD* gene family in a freshwater fish species.

In our studies, the deduced amino acid sequences of *Sox* were highly similar to other vertebrate *SoxD* genes. The relationship between different species was further confirmed by comparison of the chromosome synteny in various vertebrates. In carp, *etnk1* was located next to *Sox5* on Chr7, and further analysis revealed a number of other genes located around *Sox5,* all of which were present in both teleost fish and human, but interspersed. Similarly, *CcSox13* was located next to *etnk2* and *yod1* on Chr22, but interspersed in other species. Some genes were lost in carp, including *lrmp*, *casc1* around *Sox5* and *slc41a1* around *Sox13*. However, there were no conserved regions flanking *CcSox6*. One interpretation of the phenomenon is that chromosomal rearrangements including transpositions, translocations and deletions were different during carp evolution, and that the chromosomal rearrangements around *SoxD* are different in fish. However, *Sox5* and *etnk1*, *Sox13* and *etnk2* are always tightly linked. It revealed that conserved DNA domains are present around some *SoxD* genes, and that the conserved gene order remains, indicating that the genes arose from a common ancestral origin.

Evidence gathered from protein sequences, conserved and characteristic domains and phylogenetic analysis demonstrated that *CcSox5*, *CcSox6* and *CcSox13* were most closely related to the corresponding homologues of known *SoxD* proteins. *CcSox5*, *CcSox6* and *CcSox13* share high amino acid sequence identities with other species, especially within the conserved HMG-box domain. Results of homologous analysis indicated that *CcSoxD* genes all share high sequence similarity with teleosts such as zebrafish, rainbow trout and shared low identity with mammals. This may give a direction to study the evolutionary status of carp.

Here, we analyzed a 2000 bp 5' flanking region of *CcSox5*, *CcSox6* and *CcSox13* using bioinformatic software to identify a number of putative TFs, which might participate in the regulation of gene expression and function. These factors including BSX, BRN4, and NGN–NEUROD, are involved in the early stages of vertebrate neuronal determination and neurogenesis (Ma *et al*., 1996; Lee *et al*., 2013; Takahashi and Holland, 2004), and may interact with *CcSox5*, *CcSox6* and *CcSox13* to regulate their function in neural development. Other factors like Oct4, Nanog and FOXP1 may allow cells to retain certain stem cell properties (Loh *et al*., 2006; Gabut *et al*., 2011). It may explain the expression of *Sox5* in neural cells for it may be attributed to its interaction with these pluripotency. Moreover, some other putative binding sites for regulation,

including AP1, CEBPB, USF1, NF-Y and Sf1, have been identified. These transcription factors are ubiquitously expressed in eukaryotic cells and play important roles in diverse cellular processes. Previous studies have reported that several members of the *Sox* gene family, including *Sox2*, *Sox3* and *Sox14* participate in neurogenesis (Gao *et al*., 2015; Djurovic and Stevanovic, 2004; Dvorakova *et al*., 2016). This might suggest a conserved regulation mechanism of gene expression among different *Sox* members.

SoxD members are known to play a key role in multiple developmental pathways, particularly in the development of the central nervous system (Lefebvre, 2010; Ji and Kim, 2016; Baroti *et al*., 2015; Lefebvre *et al*., 1998). *Sox5* has been shown to play important roles in regulating processes of embryonic development and cell fate determination, including neural crest development (Martinez-Morales *et al*., 2010; Quintela *et al*., 2015). Some studies have shown that *Sox5* and *Sox6* jointly modulate oligodendrocyte development in the mouse spinal cord (Stolt *et al*., 2006; Baroti *et al*., 2015). A prior report suggested that *Sox13* was predominantly expressed in differentiating neurons of the CNS and argued against a role in glia (Wang *et al*., 2005). Correspondingly, in our paper, qRT-PCR revealed that the majority of the *Sox5* and *Sox13* transcripts were detected in the brain of carp. Importantly, the expression level of *CcSox5* was particularly high at the neurula stage during embryogenesis and in the brain of adult fish, which implied its significant role in neurogenesis and the central nervous system. *SoxD* genes are also studied to participate in gonad development (Daigle *et al*., 2015). What's more, there are some researches show that the gonadal differentiation are related to the brain steroidogenesis (Lin *et al*., 2015). In this study, tissue distribution analyses revealed that *CcSox5* and *Ccaox6* showed sexual dimorphism in brain. Therefore, we speculated that *CcSoxD* genes are involved in the process of sexual development in carp.

In summary, this study provided the full-length cDNA sequences of three *CcSoxD* genes in carp. By sequence comparison, phylogenetic analysis, gene structure and chromosomal linkage data were gathered. Several potential regulatory motifs were found in the promoter regions, which may suggest the functions of these three genes. In order to test the practicality and gain a profound understanding of the transcriptional mechanism of *CcSoxD*, further verifications are necessary. In addition, we have surveyed the expression patterns by qRT-PCR. Their upregulated expression in the adult brain suggests the potential functions of these transcripts in the regulation of neurogenesis in carp. These results will provide new information for further understanding the function of *SoxD* genes in teleost fish.

ACKNOWLEDGMENTS

This work is supported by grants from the National Natural Science Foundation of China (No. U1204329), Innovative Research Team (in Science and Technology) in University of Henan Province (No.17IRTSTHN017), The Henan Scientific and Technological Research Projects (No. 142300410164).

Supplementary material

There is supplementary material associated with this article. Access the material online at: http://dx.doi.org/10.17582/journal.pjz/2018.50.2.417.428

Statement of conflict of interest

There is no conflict of interest of any of the authors of this manuscript, and there is no financial relationship of any author with the grant funding agencies.

REFERENCES

Baroti, T., Schillinger, A., Wegner, M. and Stolt, C.C., 2015. *Sox13* functionally complements the related *Sox5* and *Sox6* as important developmental modulators in mouse spinal cord oligodendrocytes. *J. Neurochem*., **136**: 316-328. https://doi.org/10.1111/jnc.13414

Baroti, T., Zimmermann, Y., Schillinger, A., Liu, L., Lommes, P., Wegner, M. and Stolt, C.C., 2015. Transcription factors *Sox5* and *Sox6* exert direct and indirect influences on oligodendroglial migration in spinal cord and forebrain. *Glia*, **64**: 122-138. https://doi.org/10.1002/glia.22919

Bowles, J., Schepers, G. and 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

Daigle, M., Roumaud, P. and Martin, L.J., 2015. Expressions of *Sox9*, *Sox5*, and *Sox13* transcription factors in mice testis during postnatal development. *Mol. Cell. Biochem*., **407**: 209-221. https://doi.org/10.1007/s11010-015-2470-7

Djurovic, J. and Stevanovic, M., 2004. Structural and functional characterization of the human *Sox14* promoter. *Biochim. Biophys. Acta Gene Struct. Expression*, **1680**: 53-59. https://doi.org/10.1016/j.bbaexp.2004.08.011

Dvorakova, M., Jahan, I., Macova, I., Chumak, T., Bohuslavova, R., Syka, J., Fritzsch, B. and Pavlinkova, G., 2016. Incomplete and delayed *Sox2* deletion defines residual ear neurosensory

development and maintenance. *Scient. Rep*., **6**: 38253. https://doi.org/10.1038/srep38253

Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O'Hanlon, D., Sung, H.K., Alvarez, M., Talukder, S., Pan, Q., Mazzoni, E.O., Nedelec, S., Wichterle, H., Woltjen, K., Hughes, T.R., Zandstra, P.W., Nagy, A.,Wrana, J.L. and Blencowe, B.J., 2011. An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell*, **147**: 132-146. https://doi.org/10.1016/j.cell.2011.08.023

Gao, J., Li, P., Zhang, W., Wang, Z., Wang, X. and Zhang, Q., 2015. Identification, molecular characterization and gene expression analysis of *Sox1a* and *Sox1b* genes in Japanese flounder, *Paralichthys olivaceus*. *Gene*, **574**: 225-234. https://doi.org/10.1016/j.gene.2015.08.013

Gao, J., Li, P., Zhang, W., Wang, Z., Wang, X. and Zhang, Q., 2015. Molecular cloning, promoter analysis and expression profiles of the *Sox3* gene in Japanese flounder, *Paralichthys olivaceus*. *Int. J. mol. Sci*., **16**: 27931-27944. https://doi.org/10.3390/ijms161126079

Hagiwara, N., 2011. *Sox6*, jack of all trades: a versatile regulatory protein in vertebrate development. *Devlop. Dynamics*, **240**: 1311-1321. https://doi.org/10.1002/dvdy.22639

Ji, E.H. and Kim, J., 2016. *SoxD* Transcription factors: Multifaceted players of neural development. *Int. J. Stem Cells*, **9**: 3-8. https://doi.org/10.15283/ijsc.2016.9.1.3

Kiselak, E.A., Shen, X., Song, J., Gude, D.R., Wang, J., Brody, S.L., Strauss, J.F. and Zhang, Z., 2010. Transcriptional regulation of an axonemal central apparatus gene, sperm-associated antigen 6, by a SRY-related high mobility group transcription factor, S-*Sox5*. *J. Biol. Chem.,* **285**: 30496-30505. https://doi.org/10.1074/jbc.M110.121590

Lee, B., Kim, S.G., Kim, J., Choi, K.Y., Lee, S., Lee, S.K. and Lee, J.W., 2013. Brain-specific homeobox factor as a target selector for glucocorticoid receptor in energy balance. *Mol. Cell Biol.,* **33**: 2650-2658. https://doi.org/10.1128/MCB.00094-13

Lefebvre, V., Li, P. and de Crombrugghe, B., 1998. A new long form of *Sox5* (L-Sox5), *Sox6* and *Sox9* are co-expressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO J.,* **17**: 5718-5733. https://doi.org/10.1093/emboj/17.19.5718

Lefebvre, V., Dumitriu, B., Penzo-Méndez, A., Han, Y. and 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

Lefebvre, V., 2010. The *SoxD* transcription factors–*Sox5*, *Sox6*, and *Sox13*–are key cell fate modulators. *Int. J. Biochem. Cell Biol*., **42**: 429-432. https://doi.org/10.1016/j.biocel.2009.07.016

Lin, C.J., Fan-Chiang, Y.C., Dufour, S. and Chang, C.F., 2015. Activation of brain steroidogenesis and neurogenesis during the gonadal differentiation in protandrous black porgy, *Acanthopagrus schlegelii*. *Develop. Neurobiol*., **76**: 121-136. https://doi.org/10.1002/dneu.22303

Liu, C.F. and Lefebvre, V., 2015. The transcription factors *Sox9* and *Sox5*/*Sox6* cooperate genome-wide through super-enhancers to drive chondrogenesis. *Nucl. Acids Res*., **43**: 8183-8203. https://doi.org/10.1093/nar/gkv688

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B. and Liu, J., 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet*., **38**: 431-440. https://doi.org/10.1038/ng1760

Ludbrook, L., Alankarage, D., Bagheri-Fam, S. and Harley, V. 2016. Dataset of differentially expressed genes from *Sox9* over-expressing nt2/d1 cells. *Data in Brief*, **9**: 194-198. https://doi.org/10.1016/j.dib.2016.08.047

Ma, Q., Kintner, C. and Anderson, D.J., 1996. Identification of neurogenin, a vertebrate neuronal determination gene. *Cell*, **87**: 43-52. https://doi.org/10.1016/S0092-8674(00)81321-5

Martinez-Morales, P.L., Quiroga, A.C., Barbas, J.A. and Morales, A.V., 2010. *Sox5* controls cell cycle progression in neural progenitors by interfering with the WNT-beta-catenin pathway. *EMBO Rep*., **11**: 466-472. https://doi.org/10.1038/embor.2010.61

Melichar, H.J., Narayan, K., Der, S.D., Hiraoka, Y., Gardiol, N., Jeannet, G., Held, W., Chambers, C.A. and Kang, J., 2007. Regulation of gamma delta versus alpha beta T lymphocyte differentiation by the transcription factor *Sox13*. *Science*, **315**: 230-233. https://doi.org/10.1126/science.1135344

Nordin, K. and LaBonne, C., 2014. *Sox5* Is a DNA-Binding Cofactor for *BMP* R-Smads that directs target specificity during patterning of the early ectoderm. *Dev. Cell*, **31**: 374-382. https://doi.org/10.1016/j.devcel.2014.10.003

Quintela, I., Barros, F., Lago-Leston, R., Castro-Gago, M., Carracedo, A. and Eiris, J., 2015. A maternally inherited 16p13.11-p12.3 duplication concomitant

with a de novo *Sox5* deletion in a male patient with global developmental delay, disruptive and obsessive behaviors and minor dysmorphic features. *Am. J. med. Genet. A*, **167**: 1315. https://doi.org/10.1002/ajmg.a.36909

Reiprich, S. and Wegner, M., 2015. From CNS stem cells to neurons and glia: *Sox* for everyone. *Cell Tissue Res*., **359**: 111-124. https://doi.org/10.1007/s00441-014-1909-6

Roose, J., Korver, W., Oving, E., Wilson, A., Wagenaar, G., Markman, M., Lamers, W. and Clevers, H., 1998. High expression of the HMG box factor *Sox13* in arterial walls during embryonic development. *Nucl. Acids Res*., **26**: 469. https://doi.org/10.1093/nar/26.2.469

She, Z.Y. and Yang, W.X., 2015. Sox family transcription factors involved in diverse cellular events during development. *Eur. J. Cell Biol*., **94**: 547-563. https://doi.org/10.1016/j.ejcb.2015.08.002

Stolt, C.C., Schlierf, A., Lommes, P., Hillgärtner, S., Werner, T., Kosian, T., Sock, E., Kessaris, N., Richardson, W.D., Lefebvre, V. and Wegner, M., 2006. *SoxD* proteins influence multiple stages of oligodendrocyte development and modulate *SoxE* protein function. *Dev. Cell*, **11**: 697-709. https://doi.org/10.1016/j.devcel.2006.08.011

Takahashi, T. and Holland, P.W., 2004. Amphioxus and ascidian Dmbx homeobox genes give clues to the vertebrate origins of midbrain development. *Development,* **131**: 3285-3294. https://doi.org/10.1242/dev.01201

Wang, Y., Bagheri-Fam, S. and Harley, V.R., 2005. *Sox13* is up-regulated in the developing mouse neuroepithelium and identifies a sub-population of differentiating neurons. *Brain Res. Dev. Brain Res*., **157**: 201-208. https://doi.org/10.1016/j.devbrainres.2004.12.010

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., Yang, C., Tao, W. and 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

Xin, Y., Tang, X., Yue, F., Zhang, D., Yan, X., Wang, C. and Chen, Q., 2012. Isolation and sequence analysis of *Sox* genes from lizard *Eremias multiocellata*. *Russian J. Genet*., **48**: 89-96. https://doi.org/10.1134/S102279541201019X

Zhang, W., Jia, Y., Ji, X., Zhang, R., Liang, T., Du, Q. and Chang, Z., 2016. Optimal reference genes in different tissues, gender, and gonad of yellow river carp (*Cyprinus carpio var*) at various developmental periods. *Pakistan J. Zool*., **48**: 1615-1622.