The Israeli Journal of Aquaculture - Bamidgeh, IJA_72.2020.1117447, 15 pages
CCBY-NC-ND-4.0 ● https://doi.org/10.46989/001c.21467

The *IJA* is a peer-reviewed open-access, electronic journal, freely available without charge to users
-
Produced by the AquacultureHub non-profit Foundation
Sale of *IJA* papers is strictly forbidden

Molecular characterization and expression of the related-male gene sox30 in the common carp *Cyprinus carpio*

Xianglin Cao[a,b], Xiaozhe Fu[b], Yidi Zhao[c], Baohua Li[a], Huajie Zhang[c], Nana Liu[c], Ningqiu Li[b], Jianjun Chen[c]*

[a] *College of Fisheries, Henan Normal University, Xinxiang, 453007, People's Republic of China*

[b] *Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Key Laboratory of Fishery Drug Development, Ministry of Agriculture and Rural Affairs, Key Laboratory of Aquatic Animal Immune Technology, Guangzhou, 510380, People's Republic of China*

[c] *College of Life Science, Henan Normal University, Xinxiang, 453007, People's Republic of China*

Key words: common carp *Cyprinus carpio*, sox30, identification, expression, sexual development

Abstract

The Sox (SRY-related HMG-box) family of transcription factors is involved in the regulation of embryonic development and determination of cell fate. Sox proteins serve as transcriptional regulators that are complexed with other proteins. For this study, we initially cloned and characterized the full-length cDNAs, DNA sequences, and 5'-flanking regions of the common carp *Cyprinus carpio* Sox30. The sequence analysis suggested that Ccsox30 carried a distinct HMG-box of the Sox family within *Cyprinus carpio*. Phylogenetic and gene structure analysis revealed that sox30 was homologous to mammalian Sox30, whereas chromosome synteny analysis demonstrated that the position of *Cyprinus carpio* Sox30 in the genome was different from that of other vertebrates. This might have been due to the split of the Sox30 flanking gene by several genes not yet found near the Ccsox30 in evolution, or because the genome sequencing data was not annotated. The results of Real Tim Quantitative-PCR (RT-qPCR) revealed that Sox30 expression was high in the testes, and the expression was traced in other tissues by researching the tissue distribution of *C. carpio* and ontogeny of Ccsox30 expression in the gonads. This expression pattern suggested that Ccsox30 may be involved in spermatogonial differentiation and spermatogenesis.

* Corresponding author: cjjjianjun@163.com

Introduction

The Sox (*Sry*-related high mobility group box) gene family encodes multiple transcription factors that share a DNA-binding motif known as the high mobility group (HMG) domain (Sinclair et al., 1990). Moreover, as they share an HMG-domain with more than an 80% sequence identity, Sox proteins are segmented into different sub-groups, which are denoted A to K (Sarkar and Hochedlinger, 2013; Sarraj et al., 2003). The Sox A group consists of Sry, and is specific to mammals, which is the testis-determining factor (TDF) that initiates male sex determination (Janssen et al., 2018). The SOXB1, SOXB2, SOXC, SOXD, SOXE, SOXF, and SOXH groups are found throughout vertebrates, which suggests that they are the core SOX groups (Heenan et al., 2016).

The Sox31 of *Xenopus* belongs to SOXI (Mizuseki et al., 1998), and SOXJ is currently only found in nematodes (Bowles et al., 2000). Furthermore, the Sox gene *Olsox32* was defined as the K subfamily by genome-wide analysis in *Oryzias latipes* (Cui et al., 2011). Beginning with the groundbreaking discovery of Sry as a determinant of mammalian testes (Gubbay et al., 1990), progressively more research has indicated that *Sox3, Sox5, Sox6, Sox8, Sox9, Sox11,* and *Sox30* facilitates spermatozoa formation, testis development, and maturity in vertebrates (Bai et al., 2018; Bettegowda and Wilkinson, 2010; Mata-Rocha et al., 2014; Melo et al., 2019; Wan et al., 2019; Weiss et al., 2003).

Sox30 may constitute an H subgroup of the Sox protein family, which was first cloned from mouse and human, and subsequently from the teleost Nile tilapia (Sarkar and Hochedlinger, 2013; Han et al., 2014a; Han et al., 2010b). In human and mouse, it was observed that Sox30 is involved in the regulation of embryonic development and the determination of cell fate (Osaki et al., 1999). Additionally, the antitumorigenic effect of Sox30 is mediated by directly binding to the CACTTTG motif of the p53 promoter region, thereby activating p53 transcription, which suggests that Sox30 is a novel transcriptional activating factor of p53 (Han et al., 2015). In the Nile tilapia, expression analysis revealed that Sox30 may be involved in female and male gonadal development at different stages via alternative splicing (Han et al., 2010b). Although the Sox30 of the above three vertebrates have been well documented for their essential roles in embryonic development, cell proliferation and disease, the functional roles and molecular mechanisms that regulate the sexual development of Sox30 in *C. carpio* are largely unexplored.

Cyprinus carpio is one of the most extensively cultured fish species, which currently has high economic value (Winfield, 2016). Moreover, it is important for genetic research toward the investigation of its cellular and molecular components (Xu et al., 2011). However, the study of sex determination and sex regulation in the common carp remains unknown.

In the present study, common carp Sox30 was characterized through bioinformatic, phylogenetic and syntenic analysis. Moreover, we investigated the tissue distribution of Sox30 in sexually mature carp via Real Tim Quantitative-PCR (RT-qPCR), and its spatiotemporal expression pattern in carp development, with the goal of expanding knowledge of the sex-determining mechanisms and functions of Sox genes.

Materials and methods

Ethics statement

All investigations in this study were performed according to the Animal Experimental Guidelines of the Ethical Committee of the University of China. Following deep and overdosed anesthesia with 2-phenxyethanol, the fish were euthanized by immediately severing the spinal cord adjacent to the head.

Samples

The common carp used in this study were obtained from the aquaculture base of Henan Normal University. Male and female *C. carpio* were reared in freshwater tanks under ambient photothermal conditions. To obtain different hatchling age groups, *in vitro* fertilization was conducted during the breeding season using mature males and gravid females that were intraperitoneally injected with carp pituitary extract dissolved in 0.86% saline (0.4 mg/kg BW). The fertilized eggs were transferred to small glass tanks containing

filtered water with aeration. *C. carpio* embryos hatched within 24 h, and the hatchlings could survive for 2-3 days by utilizing the yolk sac for nutrition without supplemental feed.

Following this period, the hatchlings were fed with yolk and soy milk until they were capable of eating pelletized feed. The primordial germ cells (PGC) and gonads of both sexes were collected at 45-50 days post-hatching (dph). It was recorded as O1 when PGCs surrounded by supporting cells were observed in the abdominal region near the coelomic cavity in the developing female gonad, and T1 when developing spermatogonia surrounded by supporting/sertoli cells were observed in testicular differentiation.

Testis and ovary were collected at early stages of sex differentiation (T1 and O1) and at the sexual maturity stage of adults (TA and OA). All samples were isolated under sterile and RNase-free conditions and immediately fixed in RNAlater (Applied Biosystems, Foster City, CA, USA) for total RNA extraction. Subsequently, the gonads, intestines, spleen, hepatopancreas, brain, muscle, kidney, adipose tissue, swim bladder, and gills were isolated from sexual mature *C. carpio* for the extraction of the total RNA.

2.3 Extraction of total RNA and genomic DNA

The total RNA was isolated from each sample using Trizol Reagent (Invitrogen, USA, http://www.lifetechnologies.com/cn/zh/home/brands/invitrogen) following the manufacturer's protocols. The extracted RNA was further treated with DNase I (10 U/mL, Ambion, USA, http://www.lifetechnologies.com/cn/zh/home/ brands/ambion) to eliminate contamination by genomic DNA. The genomic DNA was extracted from the common carp testis using the classical phenol-chlorophenol method. The quality and concentration of RNA and genomic DNA were determined by gel electrophoresis and the Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

Molecular cloning of Ccsox30

Three micrograms of total RNA were reverse transcribed to cDNA using the Oligo (dT) 20 primer and MMLV reverse transcriptase. According to the expressed sequence tag (EST) of Sox30 from the transcription database of the *C. carpio* gonad (unpublished), the rapid amplification of cDNA ends (RACE) was performed using the 5' RACE (TaKaRa Code: D315) and 3' RACE (TaKaRa Code: D314) systems to obtain the unknown 5' and 3' sequences. To obtain the 3' ends of the common carp Sox30, primer pairs sox30-outer 32/3′ RACE-outer primer and Sox30-inner 31/3′ RACE-inner primer (**Table 1**) were designed for the primary PCR and the nested PCR, respectively.

Similarly, the 5' end of the Sox30 gene was obtained by the primary PCR and nested PCR, using primer pairs Sox30-outer 52/5′ RACE--outer primer and Sox30-inner 51/5′ RACE-inner primer (**Table 1**). The full-length cDNA sequence was confirmed by sequencing the PCR product amplified by sox30cf and sox30cr primers (**Table 1**) in the 5' and 3' untranslated regions (UTR), respectively. The resulting PCR products were cloned into pMD19-T easy vector (TaKaRa, Japan) and sequenced.

Based on the Ccsox30 cDNA sequence, the primers were designed to gradually amplify the genomic sequence, where six pairs of primers worked well and six overlapping fragments that covered the full-length cDNA sequence were amplified and sequenced (**Table 1**). The introns were annotated by comparing the obtained sequence with the cDNA sequence.

Detection of promoter activity of the 5' flanking sequence of Ccsox30

The 5' flanking sequence of the Ccsox30 gene was PCR-amplified from genomic DNA using the Genome Walker™ Universal kit (Clontech). Briefly, in accordance with the manufacturer's instructions, four Genome Walker libraries were constructed. Two adjacent reverse primers, sox305-inner 1 and sox305-inner 2, were designed at the 5' UTR region of Ccsox30 (**Table 1**) and employed for a two-step extension of the 5' flanking sequence in combination with the forward adaptor primers AP1 and AP2 for each library. Then, the 5′-flanking DNA fragment was digested with BamH I/EcoR I enzymes, and subcloned into the pEGFP (Clontech) vector, which was digested with BamH I/EcoR I enzymes. The recombinant vector was designated as pSox30-EGFP and the 5′-flanking fragment of

Ccsox30 was located in the upstream of the Enhanced Green Fluorescent protein (EGFP) gene.

Caudal fin cells (CFCs) (purchased from the China Center for Type Culture Collection) were cultured in media containing 60% Dulbecco' s modified Eagle's medium, 20% Leibovitz medium (L-15), 10% fetal calf serum (Gibco BRL; www.gelcompany.com/gibco-br), and 10% tryptose phosphate (Difco) supplemented with 1% N-2-hydroxyethylpiperazine-N-2′-ethanesulfonic acid (HEPES), 100 IU/ml of penicillin (Sigma), and 100 µg/ml of streptomycin (Sigma). The cells were maintained at 28℃ with 5% CO_2 in 24 well tissue culture plates. The CFCs (3×106 cell/ml) were transfected with 0.5 µg of purified pSox30-EGFP and pCMV-EGFP (a promoterless plasmid, as a control) using the FuGENE HD Transfection Reagent (Roche). After 24h, the CFCs were observed under microscope (Nikon AZ100).

Sequence analysis

Sequence homology was determined using the basic local alignment search tool (BLAST) program (www.ncbi.nlm.nih.gov/blast) and matrix global alignment tool (MatGAT) (http://Bitincka.com/ledion/matgat/). The protein structure was predicted using the expert protein analysis system (www.expasy.org) and sequence manipulation suite programs (www.bioinformatics.org/sms). The protein domain features were predicted using the simple modular architecture research tool (SMART) (http://smart.embl-heidelberg.de/), Pfam database search (http://pfam.sanger.ac.uk/search/), and putative conserved domain database (Finn et al., 2008). Intradomain features were predicted by scanning the sequence against the PROSITE database (http://us.expasy.org/tools/scanprosite/).

The phylogenetics of animal group Sox30 was constructed using the neighbor-joining method by MEGA v. 7 (Molecular Evolutionary Genetics Analysis Version 7.0) (Kumar et al., 2016) with yeast MATA1 sequence as an outgroup. Furthermore, the nucleotide sequences and their deduced amino acid (aa) sequences were aligned to analyze molecular evolution by the multiple alignment software ClustalX.

The promoter region was predicted by PromoterScan (https://www-bimas.cit.nih.gov/molbio/proscan/) based on the flanking of the Ccsox30 genome sequence. The genomic DNA sequences of *Homo sapiens* (GeneID: 11063), *Mus musculus* (Gene ID: 214105), *Rattus norvegicus* (Gene ID: 689918), *Gallus gallus* (GeneID: 416243), *Anolis carolinensis* (Gene ID: 114591400), *Sus scrofa* (Gene ID: 100521345), *Oreochromis niloticus* (Gene ID: 100533980), and *Rhinatrema bivittatum* (Gene ID: 115079661) were downloaded from the NCBI (http://www.ncbi.nlm.nih.gov/) for comparison of the genetic structure of the Sox30 genes from various species.

2.7 Real Tim Qanititative-PCR

RT-qPCR (Real Tim Qanititative-PCR) was performed to investigate the mRNA expression patterns of the Ccsox30 gene in tissues, and ontogeny expression in the gonads of *C. carpio*. According to the method above, the cDNA templates for RT-qPCR were obtained and then stored at -20℃. The primers were designed using the Primer 5 software and synthesized by Sangon (Shanghai, China) (**Table 1**). The qRT-PCR was performed using an ABI 7500 system (Applied Biosystems, Foster City, USA), where beta-actin was employed as an internal reference gene for cDNA normalization in the liver and brain.

The RT-qPCR mixture was comprised of a 2 µl cDNA sample, 7.6 µl nuclease-free water, 10 µl 2 × SYBR Green PCR master mix (TaKaRa), and 0.2 µl of each gene-specific primer (10 mM). The PCR cycling conditions were: 95℃ for 30 s, and 40 cycles of 95℃ for 5 s, 60℃ for 30 s, and one cycle of 95℃ for 15 s, 60℃ for 30 s, and 95℃ for 15 s; followed by dissociation curve analysis (65-95℃ at increments of 0.5℃ for 5 s) to verify the amplification of a single product. The relative expression of the target genes of the treated group to the control group was calculated using the $2-\Delta\Delta CT$ method (Livak and Schmittgen, 2001).

Table 1 Primer sequences and their designated applications in this study.

Primer name	Sequence (5′-3′)	Amplicon length (nt) and primer information Ccsox30
Sox30-inner51	CGCCTCACTGGACAAATTGC	5′RACE
Sox30-outer52	CTCGCTGATTGGCTGATGG	
5′RACE Outer Primer	CATGGCTACATGCTGACAGCCTA	
5′RACE Inner Primer	CGCGGATCCACAGCCTACTGATGATCAGTCGATG	
Sox30-inner31	AGCAGAAGAAGCCATACTACGA	3′ RACE
Sox30-outer32	CAGAAAGTGAGGTGGCAAGAT	
3′RACE Outer Primer	TACCGTCGTTCCACTAGTGATTT	
3′RACE Inner Primer	CGCGGATCCTCCACTAGTGATTTCACTATAGG	
sox30cf	ACGACATACTGACAAGCGAATG	Sox30 full-length cDNA
sox30cr	GAACAAACAGGAACCCAGAAGT	
sox305-FR1	GTTCACTGACTCGTTGAGCTTTGACA	5'-flanking sequence
sox305-FR2	AGGCTGTTAACTTCAGTTGGTTCACT	
Sox30F11	CCACGACCATTACTGCTGAG	Intron detection 0 bp
Sox30R15	GTCTGTGAGTTGTTAGGCTGTT	
Sox30F23	CCGCTGATAACGCAATTCCT	Intron detection 290 bp
Sox30R24	CGCACTGATTGGCTGATGG	
Sox30F31	ACAGCCTAACAACTCACAGACA	Intron detection 76 bp
Sox30R33	GGAATTGCGTTATCAGCGGTTA	
Sox30F44	CTTCACTCAGCCGCCTTCT	Intron detection 487bp
Sox30R46	GACAGTTTGTGCCATTCTAACC	
Sox30F52	AGCAGAAGAAGCCATACTACGA	Intron detection 143bp
Sox30R53	CTGGGTAAGGGTTGGGTCTC	
Sos30F61	CAGAGACCCAACCCTTACCC	Intron detection 0bp
Sos30R65	GAAGATACATCGAGGAGCAGAG	
Sox30qF1	CCACGACCATTACTGCTGAG	qRT-PCR 234bp
Sox30qR1	CCACGACCATTACTGCTGAG	
AP1	GTAATACGACTCACTATAGGGC	Genome walking
AP2	ACTATAGGGCAAGAGTGGT	
Sox30-p-B	AGTCggatccGTTGCGGGTCAGTGTTT	Promoter amplification
Sox30-p-E	AGTCgaattcTGGGTGAACTATCCCTTTA	

Note: Lower case letters are introduced restriction sites

Results

Cloning and sequence analysis of Ccsox30 from C. carpio

The full-length *C. carpio* Sox30 cDNA sequence was 1469 bp in length, which contained an open reading frame (ORF) of 1167 bp flanked by a 106 bp 5' UTR and a 293 bp 3' UTR with a canonical polyadenylation signal (ATTAAA) (**Figure 1**). It revealed that Ccsox30 contained 5 exons (620, 237, 174, 127, and 311 bp) and four introns (290, 76, 487, and 143 bp) (Fi 1). The intron sequences began with GT and ended in AG. The cDNA encoded a polypeptide of 389 amino acids (aa) with a calculated molecular mass of 43.97 kDa, and an isoelectronic point of 8.27. The encoded protein of Ccsox30 contained a signal peptide and HMG-box motifs, and the motif of DNA binding sites was RPMNAFMVW.

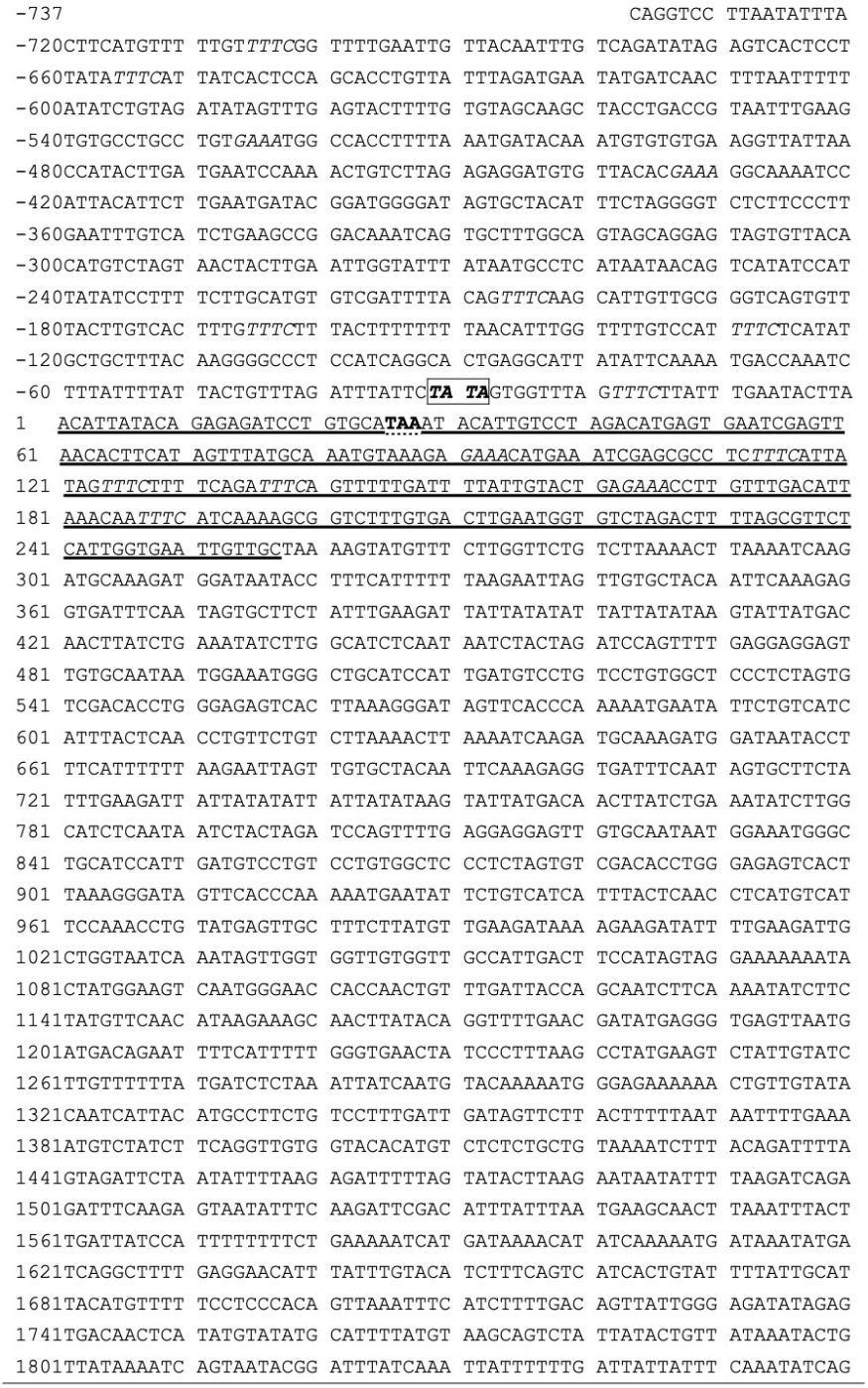

Figure 1 Full-length genomic sequence and predicted amino acid sequence of Ccsox30 gene. The nucleotide and amino acid sequences are numbered on the left of each line. "1" marks the transcription initiation site, and the introns are shown in lowercase. The predicted promoter region

is underlined (-47/257nt)m and the TATA box is in bold and boxed. The "stop codon" in front of the start codon is in bold and under the dotted line. The start codon (ATG) is boxed and the stop codon (TAG) is designated by an asterisk. The canonical polyadenylation signal (AATAAA) is in bold and under the double lines. In the deduced amino acid sequence, the putative sequence of the HMG-box is marked with a bending line, and the conservative motif of RPMNAFMVW in the HMG-box is boxed.

Alignment and phylogenetic analysis

The Ccsox30 contained the characteristic HMG-box DNA binding domain with a conservative motif between species. The HMG-box of Sox30 between *C. carpio* and other fish showed a ~65.22% similarity, while the complete peptides of both showed ~28.81% similarity (**Figure 2**).

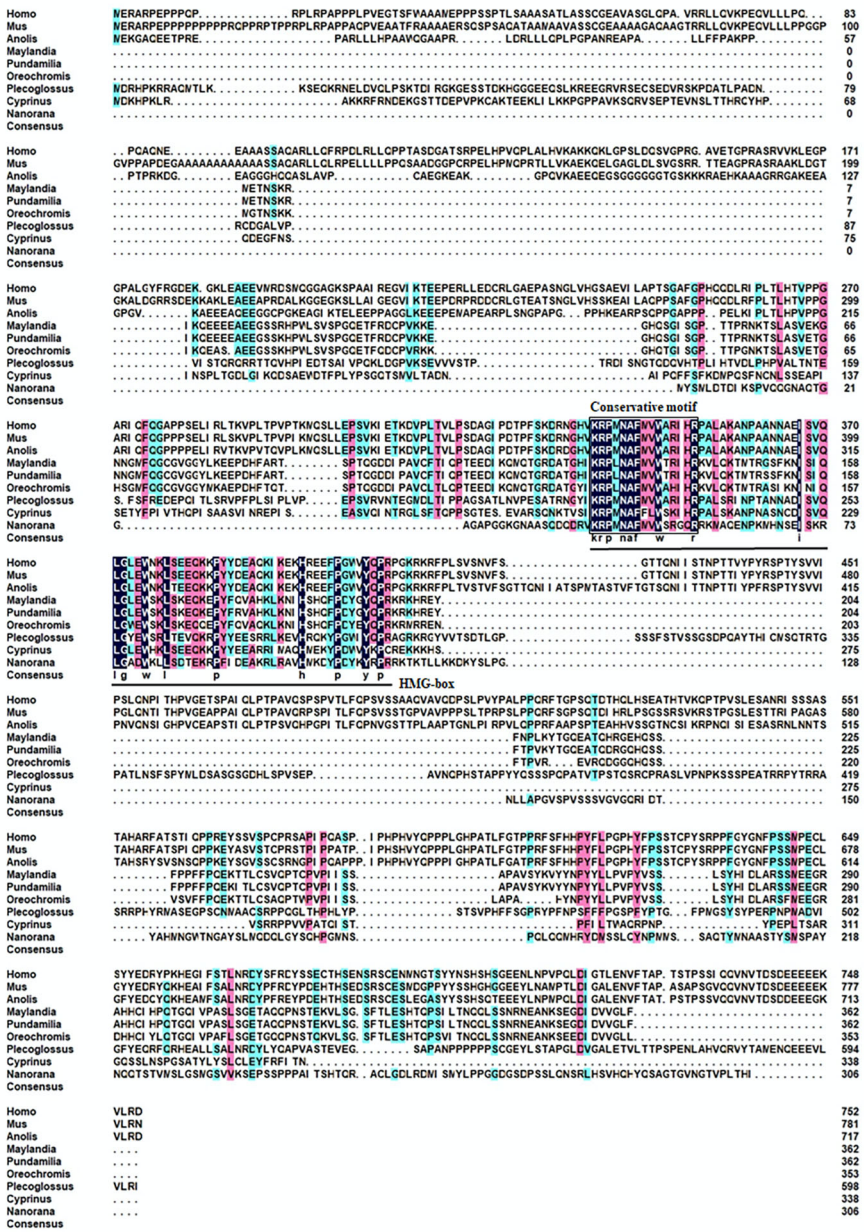

Figure 2 Alignment of the amino acid sequences of Sox30s. The HMG-box is underlined, and the conservative motif in the HMG-box is boxed. The protein IDs were as follows: *Homo sapiens* accession: BAA37146.1; *Mus musculus* accession: NP_775560.1; *Podarcis muralis* accession: XP_028574201.1; *Maylandia zebra* accession: XP_004550310.1, *Pundamilia nyererei* accession: XP_005727929.1; *Oreochromis niloticus* accession: ADM52739.1; *Plecoglossus altivelis* accession: AHK05944; *Nanorana parkeri* accession: XP_018419654.1.

To analyze the sequence homology and molecular evolution, all known and predicted Sox30 protein sequences were selected from GenBank to construct a phylogenetic tree and employed the yeast MATA1 sequence as outgroup (**Figure 3**). These ortholog proteins were divided into four groups: mammal, bird, reptile, and fish. The phylogenetic tree revealed that the common carp and *Clupea harengus* belonged to the same cluster, which was most closely related to the cluster including *Plecoglossus altivelis* and *Esox lucius*, which thus formed a separate cluster of fish Sox30 that differed from other vertebrate Sox30 genes.

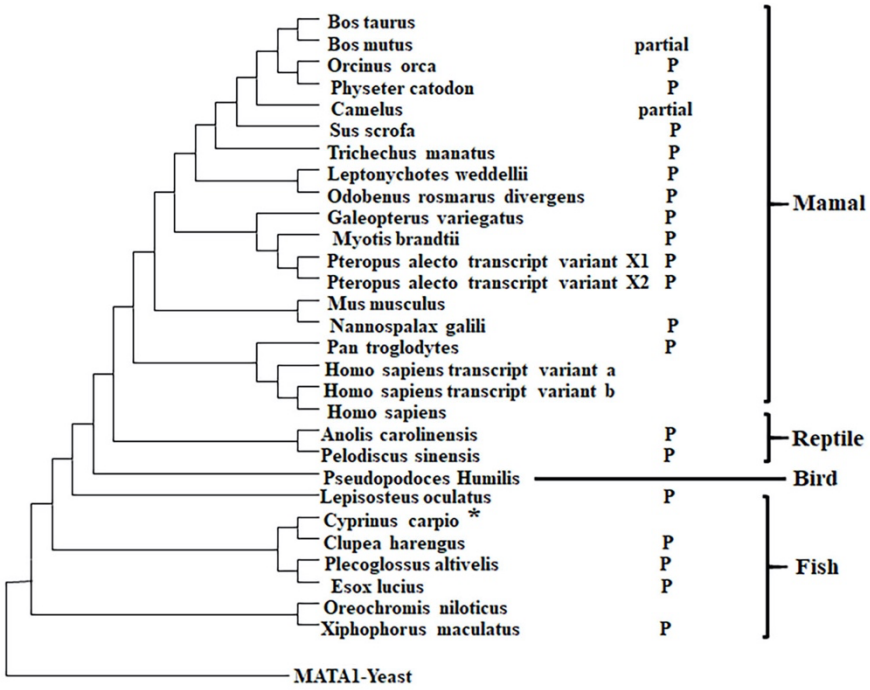

Figure 3 Phylogenetic tree showing the relationship between CcSOX30 and other vertebrate SOX30s. They include all known and predicted piscine SOX30 sequences, representative of reptile, birds, and mammalian sequences in GenBank. The bar indicates distance. The common carp SOX30 is marked with a pentagram (★). The protein IDs are listed in the supplementary data 1. "P" refers to predicted amino acid sequence.

Chromosome synteny and genomic analysis

The genomic sequence of CcSox30 was 6441 bp in length, which contained five exons (620, 237, 174, 127, and 311 bp) and four introns (290, 76, 487, and 143bp) (**Figure 1**) when comparing the cDNA sequences with genomic DNA sequences. The intron-exon structure of Ccsox30 was similar to that of other species, except for human, which contained six exons. Compared with the Sox30 of mammals, *C. carpio* Sox30 was shorter at both the N- and C-terminal regions. Furthermore, the HMG-box domain in the putative sox30 ORF was found to be interrupted by introns when compared to the cDNA sequence with the genomic DNA sequence (**Figure 4**).

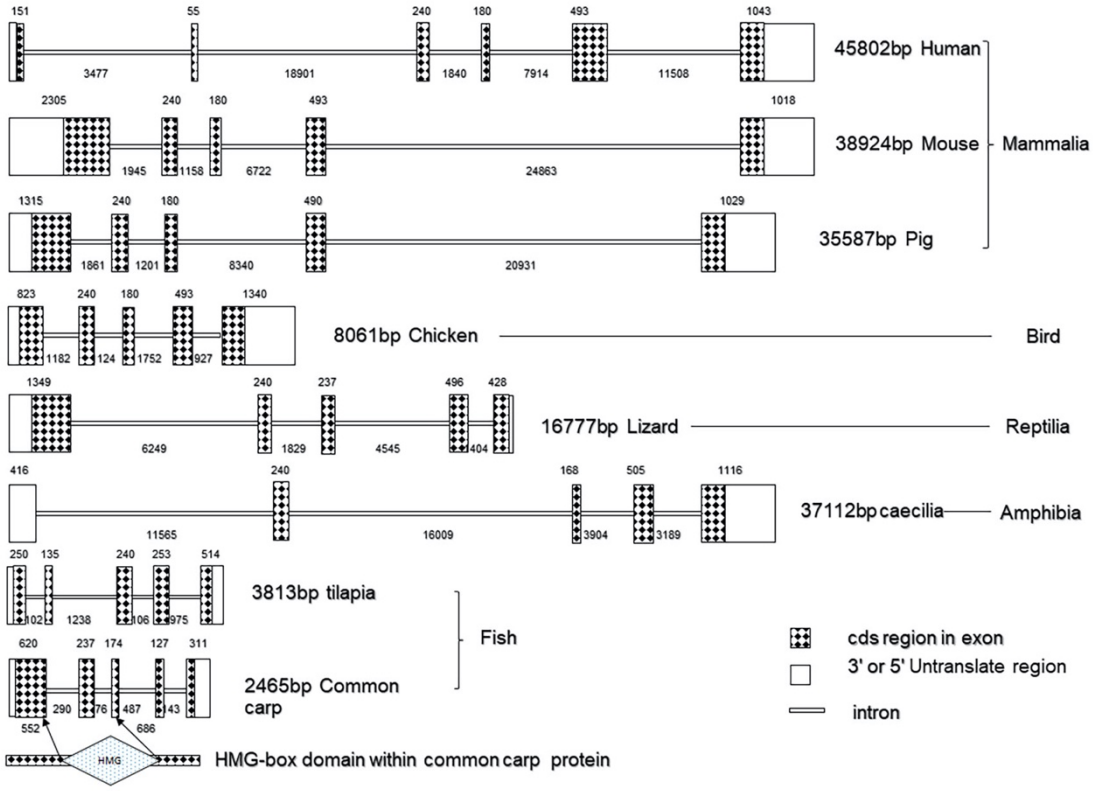

Figure 4 Diagrammatic comparison of the gene structure of the Sox30 gene from various species. Numbers on and under each schematic represent the lengths of exons and introns of the corresponding genes, respectively. The genomic DNA sequence of various species were obtained from the NCBI genomic database (https://www.ncbi.nlm.nih.gov/).

Subsequently, the chromosome syntenic relationship of the common carp Sox30 genes with other animal orthologs was obtained by comparing their chromosome locations. The orientation and chromosomal position of Sox30 were manually determined from the gene orientations listed in BIOMART from the NCBI database. In pig, chicken, anole lizard and coelacanth, Sox30 was adjacent to gene Thg1l (tRNA-histidine guanylyltransferase 1-like) and Adam19 (ADAM metallopeptidase domain 19). However, Sox30 in human, mouse, Nile tilapia, and common carp were not. The orientation and chromosomal/scaffold position of Sox30 from common carp were identical to that of Nile tilapia, which was on the scaffold flanked by clint1 (Clathrin interactor 1) and tcerg1 (Transcription elongation regulator 1) (**Figure 5**).

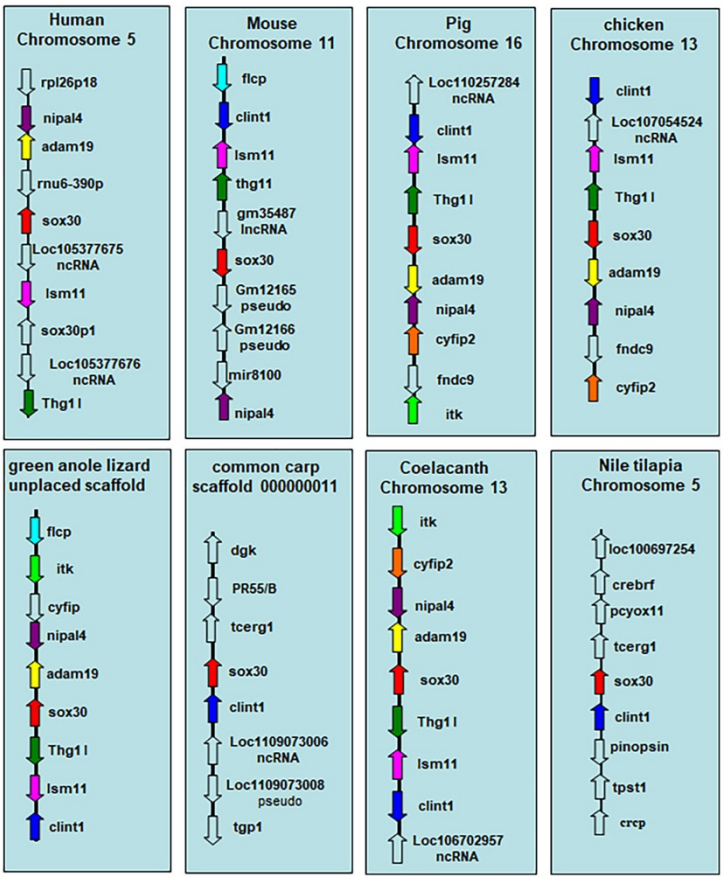

Figure 5 Chromosome syntenic relationship of common carp sox30 genes with other animal orthologs. Conserved syntenies are shown for chromosomal segments containing sox30. Rectangles represent genes in chromosome/scaffold and arrows represent gene-coding direction. Sox1, arhgef7, tubgcp3, atp11a and mcf2la orthologs are shown in red, green, yellow, blue and purple, respectively. Chr, chromosome; Sca, scaffold.

Verifying of promoter activity

The core sequence of the Ccsox30 gene promoter region was 250 base pairs in front of the gene initiation region, including TATA box (Figure 1). Based on the prediction results of the transcription initiation site, a 5'-flanking sequence was designed with two endonucleases (BamH Ⅰ/EcoR Ⅰ), and then ligated to the pEGFP vector (a promoterless report vector). The recombinant plasmid pSox30-EGFP was transferred into the caudal fin cells (CFCs), and the fluorescence was observed after 24 h. The green fluorescence was observed in the recombinant group, while the control group showed no fluorescence (Figure 6).

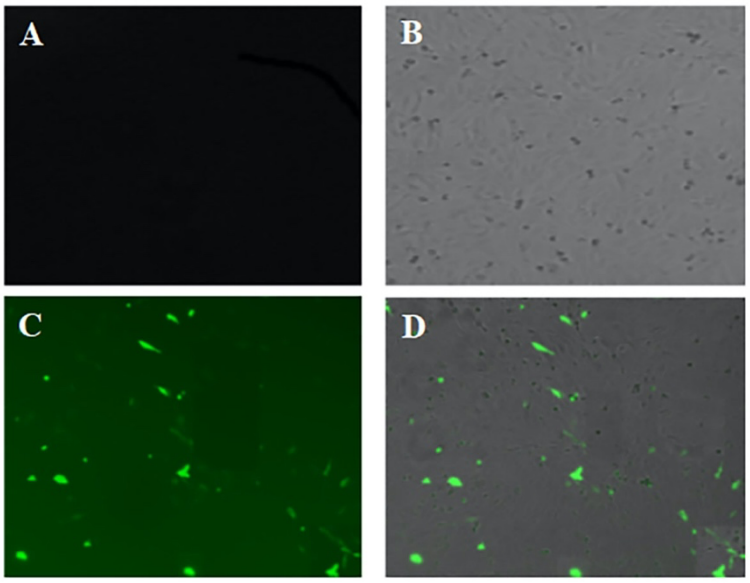

Figure 6 The observation results for verifying the promoter activity of the 5'-flanking DNA fragment of the *Ccsox30* gene by fluorescent microscopy. The fin cells transfected with pEGFP (promoterless plasmid as a control) were observed under fluorescent microscopy (A) and light microscopy (B), respectively. The fin cells transfected with pSox30-EGFP were observed under fluorescent microscopy (C), and the merging of fluorescent and light microscopy is shown in Figure D (100x magnification).

Tissue distribution and ontogeny expression of CcSox30 in gonads of C. carpio

The expression levels of CcSox30 in gonads, intestines, spleen, hepatopancreas, brain, muscle, kidney, adipose tissue, swim bladder, and gill were quantified by RT-qPCR. The results indicated that Sox30 expression levels in the gonads were highest compared with other tested tissues in both male and female fish. Further, the expression level in the testis was significantly higher than that of the ovary (**Figure 7**).

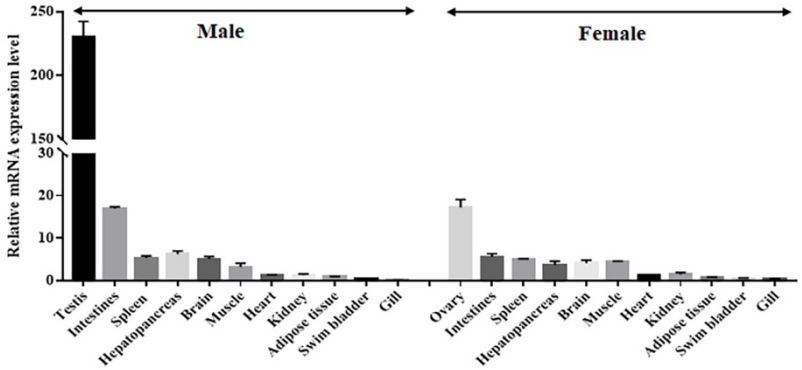

Figure 7 Expression patterns of CcSox30 mRNA in various adult tissues. The CcSox30 mRNA levels were measured via a real-time quantitative polymerase chain reaction (RT-qPCR) and normalized against the housekeeping gene β-actin. The values are shown as mean ±SE (n = 5). Each tissue sample was a mix from five individuals.

Sox30 expression was detected in the gonads of the common carp at the stages of PGC, T1, TA, O1, and OA via RT-qPCR. As shown in Figure 8, the expression levels of Sox30 at the initial (T1) and mature stages (TA) of testicular development were much higher than that in other developmental stages of ovary and PGC stages. Moreover, the expression level of Ccsox30 at sexual maturity was higher than that in the initial gender development stage.

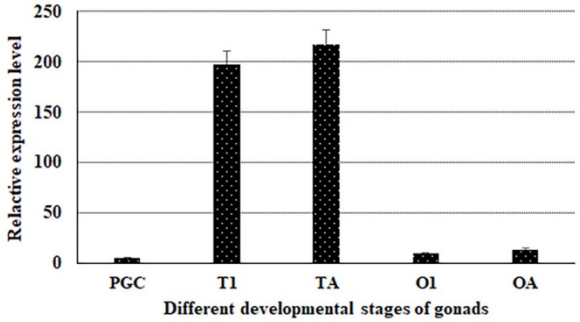

Figure 8 Expression levels of CcSox30 in different developmental stages of the gonads. CcSox30 mRNA levels were measured RT-qPCR and normalized by the housekeeping gene β-actin. The values are shown as mean ±SE (n = 5). PGC, Primordial germ cells; testis and ovary were collected at early stages of sex differentiation (T1 and O1) and at adult sexual maturity stage (TA and OA).

Discussion

To date, more than 30 Sox family transcription factors have been divided in A-K groups. Accumulating evidence has documented that the Sox family shares a characteristic high-mobility-group (HMG) domain that binds DNA in a sequence-specific manner, which is then involved in embryonic development and cell fate decisions, such as sex differentiation, maturation and/or sex changes (Bergstrom et al., 2000; Navarro-Martin et al., 2009; Zhang et al., 2010; Zhou et al., 2002).

Sox30 was initially cloned from mouse and human, and then from the teleost Nile tilapia, where several studies have identified its essential roles in embryonic development. However, the gene structure of Sox30 and its potential regulatory roles in sex development have remain unexplored. Therefore, this study initially cloned and characterized the full-length cDNAs, DNA sequence, and the 5′-flanking regions of the common carp *Cyprinus carpio* Sox30. Subsequently, various bioinformatics analyses were performed to further detect its expression patterns in different *C. carpio* tissues, which suggested its key roles in spermatogonial differentiation and spermatogenesis.

Based on data from the common carp gonad expression library, this study cloned the full-length cDNA of Sox30, which was a novel testis-specific Sox family member (Osaki et al., 1999). Accumulating evidence revealed that Sox proteins share a characteristic high-mobility-group (HMG) domain that binds DNA in a sequence-specific manner, which is involved in embryonic development, cell fate decisions, cell proliferation, and disease processes (Kamachi and Kondoh, 2013).

As expected, protein structure analysis using the SMART program showed that Ccsox30 contained 389 amino acids with the HMG-box domain. On the basis of the Sox30 amino acid sequence alignment across species, Ccsox30 showed the specific amino acid (RPMNAFMVW) motif in the HMG-box domain, which was consistent with the Sox gene family designated by Ralf (Janssen et al., 2018). The HMG-box of Sox30 between *C. carpio* and other fish showed a ~65.22% similarity, while the complete peptides of both showed a ~28.81% similarity, which suggested that the Ccsox30 amino acid sequence and HMG-boxes domains were comparatively conserved.

This was consistent with the notion that SOX sequences are highly variable outside the HMG domain, although common non-HMG domains can be identified among a number of SOX proteins; further suggesting a recent shared ancestry (Bowles et al., 2000; Uchikawa et al., 1999). Furthermore, the constructed phylogenetic tree of the Sox30 and Sox families showed that Sox30 was evolutionarily conserved from fish to mammal. The genomic structure of Ccsox30 contained five exons and four introns, which was similar to human, mouse, chicken, pig, platypus, coelacanth, and Nile tilapia, except for the green anole lizard. Compared with the Sox30 of mammals, the genomic organization of fish, including the common carp was much shorter. Similar phenomenon were found for Sox30 in Nile tilapia (Han et al., 2010b). Moreover, the HMG-box domain of the putative Sox30 ORF was interrupted by introns, which was characterized similarly to the Sox30 of Nile tilapia (Han et al., 2010b).

The genetic order and orientation of Sox30 from the common carp were the same as that of Nile tilapia, which was located on scaffold 0 000 00011 and flanked by clint1 (Clathrin interactor 1) and tcerg1 (Transcription elongation regulator 1). However, these three genes (sox30-Thg1l-Adam19) were clustered in pig, chicken, green anole lizard, and coelacanth, while interspersed in human and mouse. There were considerable variabilities in the genetic arrangement of different species. The results suggested that chromosome rearrangements, such as translocations and inversions, had frequently occurred during genomic propagation in fish. However, the conserved genetic order remained, which indicated that the genes originated from a common ancestor.

Based on the amino acid sequence alignment and genomic comparison, although the Sox30 protein is highly conserved from fish to human in primary structure, it is not similar to the finding in gene synteny. These findings suggested that the *C. carpio* sox30 flanking gene might be split by several genes not yet found in evolution, or the genome sequencing data was not annotated. This is consistent with previous studies that human and mouse Sox30 were assigned to chromosomes 5q33 and 11, respectively (Osaki et al., 1999). Further, Sox30 and its adjacent genes Thg1l and Adam19 were located on the same chromosome/scaffold in the same gene order in human, mouse, chicken, green anole lizard, and frog (Han et al., 2010b).

The 5’ upstream sequences play a critical role in the regulation of gene expression. Based on applicable technology established with the GFP gene driven by the predicted promoter regions of Ccsox30, green fluorescence was observed in the recombinant plasmid pSox30-EGFP, while the control group showed no fluorescence. This result suggested that the sequence with 250 base pairs in front of the Ccsox30 gene initiation region possessed promoter activity. However, further verification is required to test this practicality, as well as to gain a profound understanding of the transcriptional regulation mechanism of Ccsox30. This, by identifying the transcription factor binding sites in the 5’ upstream flanking region of Ccsox30.

Gene expression pattern analysis is the basis of the study of gene function. Human and mouse Sox30 are reported to be exclusively expressed in normal adult testis, specifically in germ cells (Han et al., 2014a; Osaki et al., 1999). This expression pattern suggests that Sox30 may be involved in mammalian spermatogonial differentiation and spermatogenesis (Bai et al., 2018). In the Nile tilapia, the gonad specific expression and spatiotemporal expression patterns of the alternative splicing isoforms suggested that Sox30 may play a key role in gonadal differentiation and development (Han et al., 2010b).

In contrast to mammals, the Sox family genes involved in sex determination and sex differentiation are more obscured in fish. Therefore, the present study evaluated the mRNA expression of Ccsox30 in different tissues of female and male *C. carpio*. The results revealed that Ccsox30 began to express in gonads from the PGC stage, and was significantly higher than that of other configurations (male and female), which showed a gonad-specific expression pattern, at least in adults. Furthermore, the expression level of Ccsox30 in the testis was much higher than in the ovary. This study suggested that Ccsox30 may be involved in spermatogonial differentiation and spermatogenesis in male fish, as in human and mouse (Han et al., 2014a; Osaki et al., 1999). Additional studies on Sox30 will be required to elucidate its precise function.

Conclusions

In summary, the full-length cDNA sequence of Sox30 in *Cyprinus carpio* was obtained. Amino acid sequence alignment revealed that Sox30 was poorly conserved except in the HMG-box, which suggested the high diversity and rapid evolution of these proteins. The construction of a phylogenetic tree revealed that Ccsox30 was homologous to the genes in other species. Ccsox30 was confirmed to be co-ortholog derived from gene duplication through genetic structural analysis. However, results of chromosome synteny indicated that the DNA surrounding Sox30 in different species was active in the rearrangement. Moreover, the expression patterns were surveyed by RT-qPCR, which showed potential functionality in spermatogonial differentiation and spermatogenesis. These results provide new information for further study of the potential function of Sox30 in the common carp.

Acknowledgement
We are thankful to the Henan Provincial Key Scientific and Technological Project in China (no. 192102110195, 152102210081), and Henan Normal University Ph.D. Startup Fund (no. qd17143) for the support of this study.

Declaration of competing interest
The authors declare that they have no conflict of interest.

CRediT authorship contribution statement
Xianglin Cao, Xiaozhe Fu: Data curation, Investigation, Writing - review & editing. Yidi Zhao: Data curation, Formal analysis, Investigation. Baohua Li: Validation. Huajie Zhang: Formal analysis. Nana Liu, Ningqiu Li: Investigation, Methodology. Jianjun Chen: Conceptualization, Supervision, Funding acquisition.

Data Availability Statement
The research data is confidential.

References
Bai, S., Fu, K., Yin, H., Cui, Y., Yue, Q., Li, W., 2018. Sox30 initiates transcription of haploid genes during late meiosis and spermiogenesis in mouse testes. *Development.* 145(13). DOI: 10.1242/dev.164855
Bergstrom, DE., Young, M., Albrecht, KH., Eicher, EM., 2000. Related function of mouse SOX3, SOX9, and SRY HMG domains assayed by male sex determination. *Genesis*. 28(3-4), 111-24. https://doi.org/10.1002/1526-968X(200011/12)28:3/4<111::AID-GENE40>3.0.CO;2-5
Bettegowda, A., Wilkinson, MF., 2010. Transcription and post-transcriptional regulation of spermatogenesis. *Philos Trans R Soc Lond B Biol Sci.* 365(1546), 1637-51. doi : 10.1038 / s41467-019-09972-7
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(2), 239-55. doi: 10.1006/dbio.2000.9883
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), 283-94. DOI: 10.1159/000329480
Finn, RD., Tate, J., Mistry, J., Coggill, PC., 2008. Sammut SJ, Hotz HR, et al. The Pfam protein families database. *Nucleic Acids Res.* 36(Database issue), D281-8. DOI: 10.1093/nar/gkh121
Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., 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(6281), 245-50. DOI: 10.1038/346245a0
Han, F., Dong, Y., Liu, W., Ma, X., Shi, R., Chen, H., 2014a. Epigenetic regulation of sox30 is associated with testis development in mice. *PLoS One*. 9(5), e97203. DOI: 10.1038/346245a0
Han, F., Liu, W., Jiang, X., Shi, X., Yin, L., Ao, L., 2015c. SOX30, a novel epigenetic silenced tumor suppressor, promotes tumor cell apoptosis by transcriptional activating p53 in lung cancer. *Oncogene.* 34(33), 4391-402. DOI: 10.1038/onc.2014.370
Han, F., Wang, Z., Wu, F., Liu, Z., Huang, B., Wang, D., 2010b. Characterization, phylogeny, alternative splicing and expression of Sox30 gene. *BMC Mol Biol.* 11, 98. DOI: 10.1186/1471-2199-11-98
Heenan, P., Zondag, L., Wilson, MJ., 2016. Evolution of the Sox gene family within the chordate phylum. *Gene.* 575(2), 385-92. DOI: 10.1016/j.gene.2015.09.013
Janssen, R., Andersson, E., Betner, E., Bijl, S., Fowler, W., Hook, L., 2018. Embryonic expression patterns and phylogenetic analysis of panarthropod sox genes: insight into nervous system development, segmentation and gonadogenesis. *BMC Evol Biol.* 18(1), 88. DOI: 10.1186/s12862-018-1196-z

Kamachi, Y., Kondoh, H., 2013. Sox proteins: regulators of cell fate specification and differentiation. Development. 140(20), 4129-44. DOI: 10.1242/dev.091793
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 33(7), 1870-4. DOI: 10.1093/molbev/msw054
Livak, KJ., Schmittgen, TD., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25(4), 402-8. DOI: 10.1006/meth.2001.1262
Mata-Rocha, M., Hernandez-Sanchez, J., Guarneros, G., de la Chesnaye, E., Sanchez-Tusie, AA., Trevino, CL., 2014. The transcription factors Sox5 and Sox9 regulate Catsper1 gene expression. *FEBS Lett.* 588(18), 3352-60. DOI: 10.1016/j.febslet.2014.07.024
Melo, LH., Melo, RMC., Luz, RK., Bazzoli, N., Rizzo, E., 2019. Expression of Vasa, Nanos2 and Sox9 during initial testicular development in Nile tilapia (*Oreochromis niloticus*) submitted to sex reversal. *Reprod Fertil Dev*. DOI: 10.1071/rd18488
Mizuseki, K., Kishi, M., Shiota, K., Nakanishi, S., Sasai, Y., 1998. SoxD: an essential mediator of induction of anterior neural tissues in Xenopus embryos. *Neuron.* 21(1), 77-85. DOI: 10.1016/s0896-6273(00)80516-4
Navarro-Martin, L., Galay-Burgos, M., Sweeney, G., Piferrer, F., 2009. Different sox17 transcripts during sex differentiation in sea bass, Dicentrarchus labrax. *Mol Cell Endocrinol*. 299(2), 240-51. DOI: 10.1016/j.mce.2008.11.013
Osaki, E., Nishina, Y., Inazawa, J., Copeland, NG., Gilbert, DJ., Jenkins, NA., 1999. Identification of a novel Sry-related gene and its germ cell-specific expression. *Nucleic Acids Res.* 27(12), 2503-10. DOI: 10.1093/nar/27.12.2503
Sarkar, A., Hochedlinger, K., 2013. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell.* 12(1), 15-30. DOI: 10.1016/j.stem.2012.12.007
Sarraj, MA., Wilmore, HP., McClive, PJ., Sinclair, AH., 2003. Sox15 is up regulated in the embryonic mouse testis. *Gene Expr Patterns.* 3(4), 413-7. DOI: 10.1016/s1567-133x(03)00085-1
Sinclair, AH., Berta, P., Palmer, MS., Hawkins, JR., Griffiths, BL., Smith, MJ., 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346(6281), 240-4. DOI: 10.1038/346240a0
Uchikawa, M., Kamachi, Y., Kondoh, H., 1999. Two distinct subgroups of Group B Sox genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken. *Mech Dev.* 84(1-2), 103-20. DOI: 10.1016/s0925-4773(99)00083-0
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), 1100-11. DOI: 10.1089/dna.2018.4586
Weiss, J., Meeks, JJ., Hurley, L., Raverot, G., Frassetto, A., Jameson, JL., 2003. Sox3 is required for gonadal function, but not sex determination, in males and females. *Mol Cell Biol.* 23(22), 8084-91. DOI: 10.1128/mcb.23.22.8084-8091.2003
Winfield, IJ., 2016. Biology and Ecology of Carp. Journal of Fish Biology. 88(2), 825-6.
Xu, P., Li, JT., Li, Y., Cui, RZ., Wang, JT., Wang, J., 2011. Genomic insight into the common carp (*Cyprinus carpio*) genome by sequencing analysis of BAC-end sequences. *BMC Genomics.* 12. DOI: 10.1186/1471-2164-12-188
Zhang, L., Zhu, T., Lin, D., Zhang, Y., Zhang, W., 2010. A second form of Sox11 homologue identified in the orange-spotted grouper *Epinephelus coioides*: analysis of sequence and mRNA expression patterns. *Comp Biochem Physiol B Biochem Mol Biol*. 157(4), 415-22. DOI: 10.1016/j.cbpb.2010.09.004
Zhou, R., Cheng, H., Zhang, Q., Guo, Y., Cooper, RK., Tiersch, TR., 2002. SRY-related genes in the genome of the rice field eel (*Monopterus albus*). *Genet Sel Evol.* 34(1), 129-37. DOI : 10.1186 / 1297-9686-34-1-129