https://doi.org/10.17221/4/2022-CJAS

Evolution, characterization and expression analysis of *Sox* gene family in rainbow trout (*Oncorhynchus mykiss*)

Fang Ma, Yali Zou, Ruilin Ma, Xin Chen, Lanfang Ma*

Key Laboratory of Resource Utilization of Agricultural Solid Waste in Gansu Province, Tianshui Normal University, Tianshui, P.R. China

**Corresponding author: mafangfighting@163.com*

Citation: Ma F., Zou Y.L., Ma R.L., Chen X., Ma L.F. (2022): Evolution, characterization and expression analysis of *Sox* gene family in rainbow trout (*Oncorhynchus mykiss*). Czech J. Anim. Sci., 67: 157–166.

Abstract: The *Sox* transcription factor family plays an important role in various biological processes such as animal sex determination and multiple organ development. We used online databases to analyze the gene structure, chemical characteristics, and evolutionary relationship of *Sox* family genes through bioinformatics, and we studied the expression profiles and regulatory mechanisms of *Sox* family genes. A total of 29 rainbow trout *Sox* genes were identified. The phylogenetic analysis found that *Sox* genes of rainbow trout were clustered in seven subfamilies (B1, B2, C, D, E, F and H), and the gene structure of each subfamily was relatively conserved. Furthermore, *Sox1*, *Sox4*, *Sox6*, *Sox8*, *Sox9*, *Sox11*, *Sox17*, *Sox18*, and *Sox19* developed into two copies, which might be the result of teleost fish-specific genome replication. Multiple HMG box domain alignments indicated that the motifs for all *Sox* sequences are conserved. Gene expression studies reveal that *Sox* expression is tissue-specific and that multiple *Sox* genes are involved in rainbow trout gonad and central nervous system development. Our study provides valuable information on the evolution of teleosts, and will also help to further research the functional characteristics of *Sox* genes.

Keywords: bioinformatics; gene structure; genome duplication; gene expression

The *Sox* (Sry-related high-mobility group box) gene family composed of SRY (sex-determining region of the Y chromosome) related genes is one of the most essential components in mammalian sex determination (Pan et al. 2016). All of them have homologous sequences encoding a 79 amino acid domain, known as HMG-box (high mobility group-box motif) (Hett and Ludwig 2005). This homologous sequence has been found in all vertebrate mammals so far and is highly conserved during evolution. The HMG-box and SRY gene product sequences were compared, and the amino acid sequences with more than 60% similarity were named *Sox* genes. Therefore, according to the amino acid structure sequence of the HMG domain, the *Sox* gene family can be divided into eight subfamilies, namely, A to H, of which: A has only one member, SRY; B is divided into two subfamilies, B1 and B2, B1 subfamily members are *Sox1*, *Sox2* and *Sox3*, B2 subfamily members are *Sox14*, *Sox21*; C family members are *Sox4*, *Sox11* and *Sox12*; D family members are *Sox5*, *Sox6* and *Sox13*; E family mem-

Supported by the Major pre research projects of Gansu Province Educational Technology Innovation Project: Young Doctor Fund Project, China (Project No. 2021-QB106); Natural Science Foundation of Gansu Province China (Project No. 21JR7RE178); Higher Education Innovation Fund Project of Gansu Province China (Project No. 2021B-196).

https://doi.org/10.17221/4/2022-CJAS

bers are *Sox8*, *Sox9* and *Sox10*; F family members are *Sox7*, *Sox17* and *Sox18*; G family member is *Sox15*; H family members is *Sox30* (Yu et al. 2018).

All teleosts have undergone a teleost-specific whole-genome duplication (TSGD or 3R WGD) event during evolution. Therefore, teleosts contain more *Sox* genes than mammals and other vertebrates (Zhang et al. 2018). In addition, salmonids and some cyprinids experienced additional WGD (4R), implying that *Sox* genes accumulated during the fourth round of genome duplication (Lien et al. 2016; Zhang et al. 2018). Currently, although the *Sox* family genes have been found in mammals, birds, amphibians, reptiles, insects and fish, the number of *Sox* genes in different animals varies greatly due to the occurrence of WGD. For example, there are 20 *Sox* genes in mice (*Mus musculus*) and humans (*Homo sapiens*) (Schepers et al. 2002), 19 in medaka (*Oryzias latipes*) (Cui et al. 2011), 24 in pufferfish (*Fugu rubripes*) (Koopman et al. 2004), 27 in Nile tilapia (*Oreochromis niloticus*) (Han et al. 2010), 26 in large yellow croaker (*Larimichthys crocea*) (Wan et al. 2019) and 27 *Sox* genes in common carp (*Cyprinus carpio*) (Zafar et al. 2021). Therefore, genome replication can be considered as an important process in the origin and evolution of species.

Studies have found that *Sox* genes are involved in a variety of developmental processes in organisms, including sex determination and sex differentiation, organ formation, endoderm development, gonads (Barrionuevo and Scherer 2010), cartilage formation, ocular angiogenesis (Mukherjee et al. 2000), nerves, and cardiogenesis (Jiang et al. 2012). *Sox* family genes show different expressions at different stages of medaka embryonic development and may play a multiple roles in embryonic development (Cui et al. 2011). *Sox9b* plays an important role in the survival and proliferation of medaka germ cells (Cui et al. 2011). *Sox30* has been confirmed to be specifically expressed in the gonads of Nile tilapia (*Oreochromis niloticus*) (Han et al. 2010). *Sox21a* acts as a transcription repressor during embryonic development of zebrafish (*Danio rerio*) (Argenton et al. 2004).

Rainbow trout is one of the most important farmed freshwater fish in China (Ma et al. 2020). The current research on the role of *Sox* family genes in rainbow trout is still limited. In recent years, the development of genome and transcriptome sequencing technologies has provided better methods to identify rainbow trout *Sox* genes at the genome level. This study systematically describes the *Sox* gene family in rainbow trout, and analyzes their functional domains, motifs, gene organization, evolution and their expression patterns in different tissues, providing the theoretical basis for further research on the evolution and function of *Sox* family genes in rainbow trout.

MATERIAL AND METHODS

Experimental design and sampling

Rainbow trout used in this study were obtained from a commercial fish farm in Tianshui City, Gansu Province, China. Three-year-old healthy fish (weight 300 ± 20 g and length 20 ± 3 cm) were reared in large tanks with circulating fresh water at ambient temperature (18 °C) and natural photoperiod. Fish were anesthetized with MS-222 and then immediately nine tissue samples were collected including heart, liver, spleen, head kidney, muscle, brain, gill, ovary and testis. All samples were frozen in liquid nitrogen and stored at –80 °C for later use.

cDNA collection and RT-qPCR

Total RNA was extracted using TRIzol reagent (Invitrogen, Waltham, MA, USA). PrimerScript RT Reagent Kit with gDNA Eraser (TaKaRa, Kusatsu, Japan) was used to synthesize first-strand complementary DNA (cDNA). The Primer 5 (Premier Biosoft, Palo Alto, CA, USA) software was employed to design specific primers for all genes; the sequences of the primers are shown in Table S1 in electronic supplementary material (ESM) (for the supplementary material see the electronic version). Quantitative reverse transcription PCR (RT-qPCR) was performed on a 7 500 real-time PCR system (Applied Biosystems, Waltham, MA, USA) using a 10 µl reaction mixture containing 1 µl cDNA, 0.2 µl primer, 5 µl SYBR Premix Ex Taq II (TaKaRa, Kusatsu, Japan), and 3.6 µl RNase-free water. The thermal cycling conditions were as follows: 40 cycles of 95 °C for 20 s, 95 °C for 5 s, and 60 °C for 30 s. Each PCR experiment was performed in triplicate. The $2^{-\Delta\Delta Ct}$ method was used to calculate the relative mRNA expression.

https://doi.org/10.17221/4/2022-CJAS

Sequence bioinformatics analysis

Atlantic salmon (*Salmo salar*), zebrafish (*Danio rerio*), mouse (*Mus musculus*), human (*Homo sapiens*) and chicken (*Gallus gallus*) CDS and full-length sequences of *Sox* (Table S2 in ESM) were downloaded from NCBI (http://www.ncbi.nlm.nih.gov) as query sequences to search against the whole genome of rainbow trout. Secondly, sequences of the *Sox* family gene members of rainbow trout were obtained by BLAST (*e*-value = 1×10^{-5}) comparison of rainbow trout genomes. And then, Pfam (https://www.ebi.ac.uk/Tools/hmmer/) was used to confirm the presence of candidate protein domains.

The physicochemical properties of rainbow trout *Sox* family proteins were analyzed by the software Prot Param (http://web.expasy.org/protparam/), including molecular weight, protein length, isoelectric point (pI) and total average hydrophilicity (GRAVY).

According to the amino acid sequence alignment generated by Clustal W, the neighbour joining method of MEGA v7.0 program (https://www.megasoftware.net/) was used to construct the phylogenetic tree of Sox protein. The exon-intron structure of this gene was analyzed by GSDS (http://gsds.cbi.pku.edu.cn/). The motif analysis of the *Sox* family genes was performed by MEME v5.1.1 (http://meme-suite.org/tools/meme) software on a bioinformatic platform and the result was visualized by TBtools and Adobe Illustrator CC 2018 software.

Data presentation and statistical analysis

There are three technical replicates for each experiment. The data are expressed as the mean ± SD and use SPSS Statistics v19 (IBM Corporation, Armonk, NY, USA) to calculate. The Shapiro-Wilk and Levene tests were used to check the normality and homogeneity of the variance, and the *t*-test was performed on the various variables obtained. All the statistical analyses considered significant were $P < 0.05$.

RESULTS

Identification and analysis of *Sox* genes in rainbow trout

In this study, 29 *Sox* family genes of rainbow trout were identified from the NCBI/Ensembl databases (Table 1). They are subfamily B1 (*Sox1a*, *Sox1b*, *Sox2*, *Sox3*, *Sox19a* and *Sox19b*), subfamily B2 (*Sox14* and *Sox21*), subfamily C (*Sox4a*, *Sox4b*, *Sox11a*, *Sox11b* and *Sox12*), subfamily D (*Sox5*, *Sox6a*, *Sox6b* and *Sox13*), subfamily E (*Sox8a*, *Sox8b*, *Sox9a*, *Sox9b* and *Sox10*), subfamily F (*Sox7*, *Sox17a*, *Sox17b*, *Sox18a*, *Sox18b* and *Sox32*), subfamily H (*Sox30*). The amino acid sizes of *Sox* genes ranged from 235 to 864, with the predicted pI varying from 5.20 to 9.81. The molecular weights (MWs) varied from 264.20 kDa to 964.51 kDa.

Phylogenetic and gene characteristic analysis of rainbow trout *Sox* genes

In order to clarify the evolutionary characteristics, we used the amino acid sequence of *Sox* genes to construct a phylogenetic tree, including mammals, birds, and fish. Our results confirm the division of *Sox* genes into seven subfamilies (including B1, B2, C, D, E, F and H). We found a high degree of consistency among different subfamilies. In addition, the evolutionary relationship between subfamily B1 and subfamily B2, subfamily D and subfamily C, subfamily E, subfamily F and subfamily H is closer (Figure 1).

Gene structure analysis of rainbow trout *Sox* genes

Ten different conserved motifs were found in all Sox proteins using the MEME program (Figure 2). Sox proteins within the same group have similar motifs, like B1, B2 subfamily, D subfamily and E subfamily. The C subfamily has three motifs (motif1, motif2, and motif5), while Sox12 has one more motif4. Subfamily F has two motifs (motif1 and motif2), except Sox17b, which has one more motif4.

The exon-intron structures were compared according to the phylogenetic relationships of rainbow trout. As shown in Figure 3, the number of introns varies greatly among 29 *Sox* genes, ranging between 0 and 15. However, the *Sox* genes in the same group have a similar genetic structure, when subfamily F had one intron, subfamily C had zero intron. Subfamily B1, subfamily B2 and subfamily C have no intron, except *sox19a* and *sox19b*, which have one intron. In addition, the subfamily D has more introns than other subfamilies.

https://doi.org/10.17221/4/2022-CJAS

Table 1. Physical and chemical properties of *Sox* genes in rainbow trout

Gene	Accession No.	Protein	Chr	Location	AA	pI	MW	GRAVY
Sox1a	XM_021578776.2	XP_021434451.1	22	7719470-7736799	298	9.25	321.19	–0.825
Sox1b	XM_021608575.2	XP_021464250.1	7	13870080-13873318	338	9.81	366.15	–0.782
Sox2	XM_021599151.2	XP_021454826.1	30	21443912-21446237	315	9.69	346.23	–0.809
Sox3	XM_021583862.2	XP_021439537.1	31	29714139-29715857	297	9.63	330.89	–0.797
Sox4a	XM_021572761.2	XP_021428436.2	18	67108578-67112583	386	6.34	418.16	–0.880
Sox4b	XM_021593876.2	XP_021449551.2	2	2446374-2449761	377	6.34	407.96	–0.757
Sox5	NM_001164068.1	NP_001157540.1	15	53352748-53471763	762	7.75	842.23	–0.852
Sox6a	XM_036964052.1	XP_036819947.1	26	33354728-33556537	864	6.69	964.51	–0.698
Sox6b	NM_001124544.1	NP_001118016.1	6	73190979-73392249	767	6.43	852.12	–0.746
Sox7	XM_021601136.2	XP_021456811.2	4	42913317-42916945	451	6.37	484.47	–0.705
Sox8a	XM_036939138.1	XP_036795033.1	12	90186075-90190418	467	6.61	510.61	–0.856
Sox8b	XM_021575885.2	XP_021431560.1	20	16082734-16085908	471	6.98	513.39	–0.972
Sox9a	XM_021575875.2	XP_021431550.2	20	15278768-15282727	491	6.12	538.13	–0.899
Sox9b	XM_036941796.1	XP_036797691.1	13	61546265-61550711	489	6.28	535.82	–0.893
Sox10	XM_021625106.2	XP_021480781.2	12	75893477-75902213	496	6.28	521.43	–0.705
Sox11a	XM_021584477.2	XP_021440152.1	25	9121012-9123841	357	5.57	402.69	–0.843
Sox11b	NM_001124183.1	NP_001117655.1	4	10066566-10068867	367	5.20	413.58	–0.750
Sox12	XM_021610202.2	XP_021465877.1	7	65586599-65593329	402	7.77	443.13	–0.641
Sox13	XM_036950444.1	XP_036806339.1	17	64914606-64959680	661	9.12	742.34	–0.893
Sox14	XM_021604810.2	XP_021460485.1	5	95345148-95349127	235	9.68	264.20	–0.687
Sox17a	XM_021562941.2	XP_021418616.1	15	28003157-28005566	393	6.37	441.96	–0.981
Sox17b	XM_021619900.2	XP_021475575.2	11	36189424-36191073	325	6.54	375.59	–0.765
Sox18b	XM_021567155.2	XP_021422830.2	16	76164892-76168083	520	6.61	566.15	–0.864
Sox19a	XM_021618947.2	XP_021474622.1	10	50582305-50584800	319	9.63	353.10	–0.883
Sox19b	XM_021614680.2	XP_021470355.1	9	12436340-12441768	299	9.61	329.96	–0.849
Sox21	XM_021597289.2	XP_021452964.1	3	51228485-51230831	243	9.74	267.52	–0.552
Sox30	XM_021624173.2	XP_021479848.1	12	54485377-54494969	639	8.28	700.29	–0.696
Sox32	XM_021564366.2	XP_021420041.2	15	28035666-28036941	321	6.23	364.75	–0.814

AA = amino acid; Chr = chromosome; GRAVY = total average hydrophilicity; MW = molecular weight (kDa); pI = isoelectric point

Conserved sites of rainbow trout *Sox* genes

The HMG box of 29 Sox protein sequences was identified by SMART (http://smart.emblheidelberg.de/). Results showed that 29 Sox proteins had a conserved HMG box of 70 amino acid residues. Multiple sequence alignments revealed that the motif sequences were highly conserved for all Sox sequences except Sox32 and Sox17b. The 6th amino acid in Sox32 and Sox17b was L, and M in other Sox sequences. At the same time, the 11th and 12th amino acids of Sox32 and Sox17b are I, which is different from the highly conserved MVW of other Sox proteins in these positions (Figure 4).

Tissue expression profiling of *Sox* genes in rainbow trout

RT-qPCR was used to study the mRNA expression pattern of each ACSL gene in nine tissues, including liver, brain, gills, spleen, heart, muscle, head kidney, ovary, and testis. The expression of *Sox* genes was related to different tissues, and the expression pattern of *Sox* genes varied in different tissues. Most *Sox* genes are relatively high in the brain, except for *Sox6a*, *Sox11b* and *Sox32* and most *Sox* genes are low expressed in the gills, except for *Sox4b*, *Sox9a* and *Sox9b*. Compared with other tissues, *Sox2*, *Sox5*, *Sox4a*, *Sox7*, *Sox11a*, *Sox11a* and *Sox11b* have

https://doi.org/10.17221/4/2022-CJAS

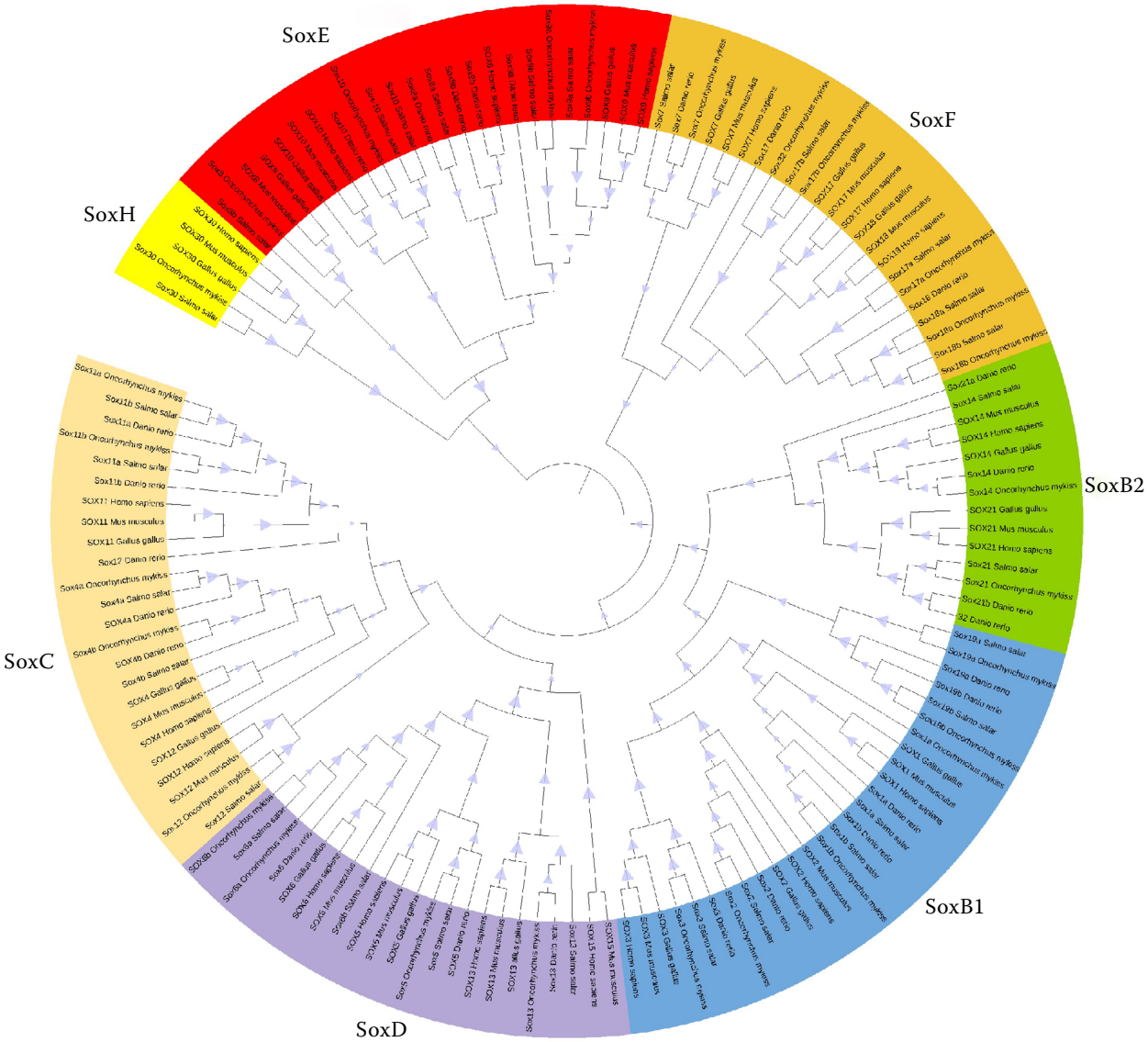

Figure 1. Phylogenetic analysis of the Sox proteins of rainbow trout

The phylogenetic tree was constructed by MEGA v7.0 using the neighbour-joining method. The statistical robustness of the tree was estimated by bootstrapping with 1 000 replicates. For Genbank accession numbers and full-length amino acid sequences of these genes see Table S1 and Table S2 in electronic supplementary material (for the supplementary material see the electronic version)

relatively higher expression in the ovary. Expression levels of *Sox8a*, *Sox8b*, *Sox9a*, *Sox30* and *Sox32* in the testis were all relatively higher (Figure 5).

DISCUSSION

In recent years, the *Sox* family genes have been studied in some species, such as Japanese medaka (*Oryzias latipe*) (Cui et al. 2011), mouse (*Mus musculus*) (Schepers et al. 2002), Nile tilapia (*Oreochromis niloticus*) (Wei et al. 2016), channel catfish (*Ictalurus punctatus*) (Zhang et al. 2018), zebrafish (*Danio rerio*) (Howe et al. 2013) and common carp (*Cyprinus carpio*) (Zafar et al. 2021). Several *Sox* genes have been studied in rainbow trout, such as *Sox9*, *Sox24* and *Sox23* (Takamatsu et al. 1997; Kanda et al. 1998; Yamashita et al. 1998), but the genome-wide identification of the rainbow trout *Sox* gene family and their expression regulation in different tissues have not been reported. In the present study, a total of 29 *Sox*

https://doi.org/10.17221/4/2022-CJAS

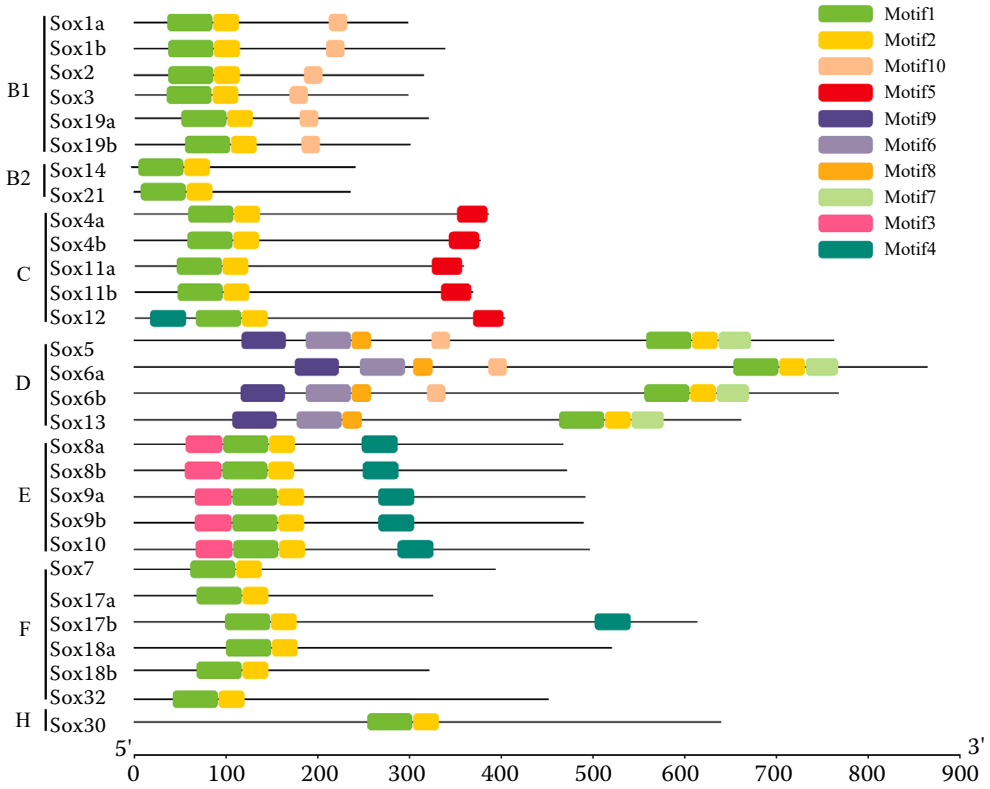

Figure 2. The conserved motifs of the Sox protein

These motifs were identified using Multiple EM for Motif Elicitation (MEME), and boxes of different colours represented different motifs

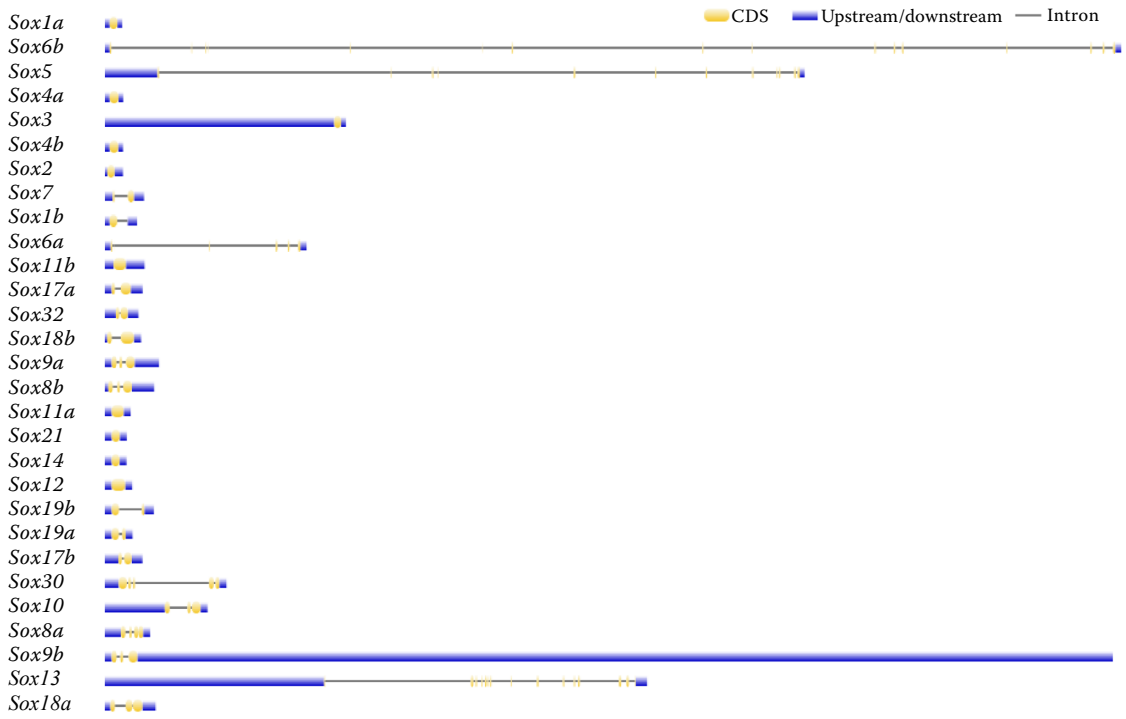

Figure 3. The exon-intron structure of *Sox* family genes

Exon-intron analysis was performed using GSDS (http://gsds.cbi.pku.edu.cn/). Introns, exons and untranslated regions (UTRs) are represented by lines, yellow and blue bars, respectively

https://doi.org/10.17221/4/2022-CJAS

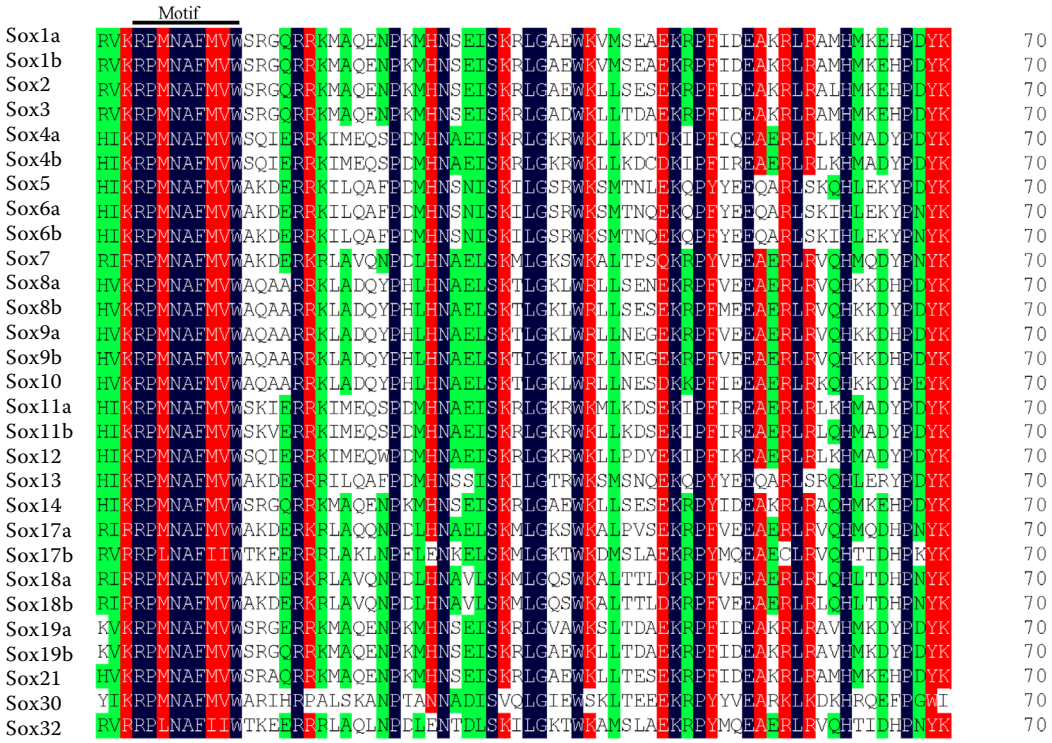

Figure 4. Multiple alignments of HMG box domains of rainbow trout Sox proteins

Motifs are marked on the sequences. The HMG box domain of each Sox protein was predicted using the SMART online program (http://smart.embl-heidelberg.de/). DNAMAN program was used to perform multiple alignments of amino acid sequences of the HMG-box domain of all the Sox proteins

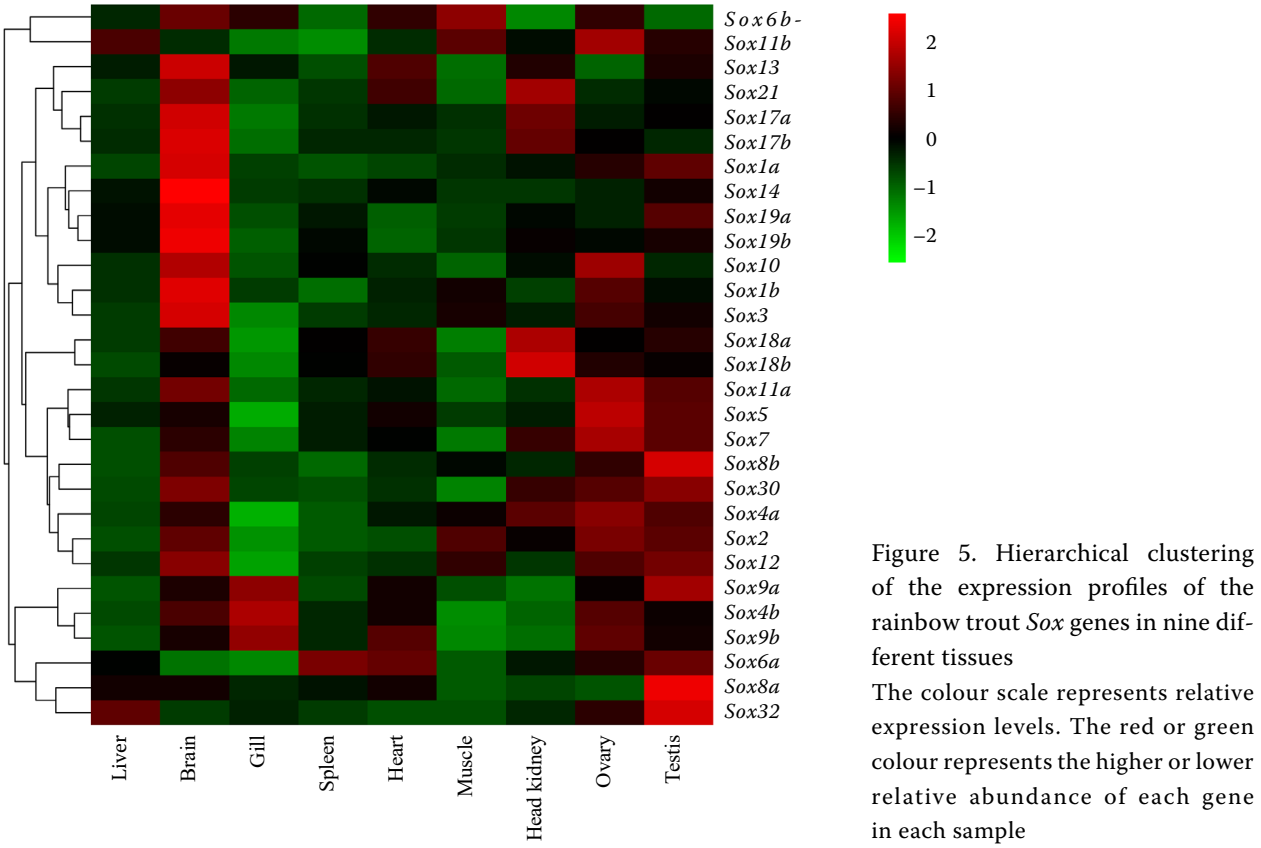

Figure 5. Hierarchical clustering of the expression profiles of the rainbow trout *Sox* genes in nine different tissues

The colour scale represents relative expression levels. The red or green colour represents the higher or lower relative abundance of each gene in each sample

https://doi.org/10.17221/4/2022-CJAS

genes were identified. The gene repertoire varies significantly between humans (20) and mouse (20) and rainbow trout (29), which may be due to whole-genome replication in teleosts. Gene duplication not only increases the genome size but also facilitates the diversification of gene functions, which ensures optimal adaptability and evolution. After the TSGD event, some genes often repeat and lose during the evolution process, which provides useful information for studying the evolution of genes (Niimura et al. 2014). In rainbow trout we found additional gene copies in the branches *Sox1*, *Sox4*, *Sox*6, *Sox8*, *Sox9*, *Sox11*, *Sox17, Sox18* and *Sox19*. Gene duplication is a key mechanism for generating new genes that promote the evolution of a species (Fortna et al. 2004).

During the evolution of fish genomes, family-specific gene duplication and gene loss events frequently occur; these events represent one of the major drivers of genome evolution and provide useful information for the study of gene evolution (Han et al. 2019). In this study, we found that the *Sox32* gene was lost in mammals, but it was identified in teleosts. *Sox32* has a tight genetic link with F subgroups, as evidenced by phylogenetic research and various alignment findings of the HMG-box domain sequence. Similar results were found in channel catfish (*Ictalurus punctatus*) (Zhang et al. 2018). Furthermore, studies in zebrafish have shown that *Sox32* is involved in endoderm formation and is an important component of endoderm differentiation in teleosts (Kikuchi et al. 2001). *Sox30* existing only in some teleosts was identified in rainbow trout, which was consistent with that in *I. punctatus* (Zhang et al. 2018). On the other hand, *Sox30* was not found in *Danio rerio*, *Oryzias latipe*, or *Cyprinus carpio*. This shows that the evolution and function of *Sox30* in teleosts and other vertebrates are complex.

Some studies have shown that the *Sox* gene is involved in the process of cell development (Abdelalim et al. 2014; Jeng et al. 2018). In this study, all *Sox* genes were expressed in all tissues, indicating that the function of this gene family is conserved. Some *Sox* genes were more abundant in different tissues, indicating their tissue-specific and functional differences. The study has found that multiple *Sox* genes are involved in the development of mouse gonads and central nervous system (Jeng et al. 2018). *Sox1*, *Sox2*, and *Sox3* are key determinants of neurogenesis, and they keep nerve cells undifferentiated by interacting with neuroproteins (Abdelalim et al. 2014). *Sox1* is expressed during the development of neural plate formation, and *Sox2* and *Sox3* maintain the characteristics of progenitor cells by inhibiting neurogenesis (Kan et al. 2004). Interestingly, in our study, *Sox1a*, *Sox1b*, *Sox2* and *Sox3* were highly expressed in rainbow trout brain, suggesting that *Sox1* plays a crucial role in brain development. In addition, *Sox13*, *Sox14*, *Sox19a* and *Sox19b* were also highly expressed in the brains of rainbow trout. Study on mouse and zebrafish found that *Sox13* and *Sox19* are expressed in the developing central nervous system (CNS) and are considered to be the earliest CNS molecular markers (Wang et al. 2006). Therefore, our results can confirm that *Sox* genes of subfamily B (B1 and B2) are involved in the neurodevelopment of rainbow trout brain.

Sox protein has a highly conserved SRY-related high-mobility group (HMG) box. SRY is involved in the testicular development, and inhibiting the expression of SRY can lead to defective testicular development (Larney et al. 2014). SRY is the target gene of *Sox9*, and it was found that *Sox9b* plays a key role in the development of medaka gonads (Nakamura et al. 2012). Study in mice found that *Sox9* mutation can lead to reversal or severe infertility, while *Sox8* mutation can lead to reduced fertility (Barrionuevo and Scherer 2010). In addition, *Sox8*, *Sox9* and *Sox10* are important regulators of mammalian gonadal development and differentiation (Georg et al. 2012). Similarly, we found that *Sox8a*, *Sox8b* and *Sox9a* were highly expressed in the testis of rainbow trout, and *Sox9b* and *Sox10* were highly expressed in the ovary. All of the above mentioned further proved the importance of the E subfamily in animal gonadal development.

CONCLUSION

We have performed genome-wide identification and expression profile of the *Sox* gene family in rainbow trout and a total of 29 *Sox* genes were identified. The structure, motif and the phylogenetic relationship of *Sox* genes were analysed. We also performed expression analyses of the *Sox* genes in different tissues, and found that *Sox* genes may play an important role in the development of rainbow trout gonads and central nervous system. Our research provides new insights for further study of the function of the *Sox* genes.

https://doi.org/10.17221/4/2022-CJAS

Conflict of interest

The authors declare no conflict of interest.

REFERENCES

Abdelalim EM, Emara MM, Kolatkar PR. The SOX transcription factors as key players in pluripotent stem cells. Stem Cells Dev. 2014 Aug 15;23(22):2687-99.

Argenton F, Giudici S, Deflorian G, Cimbro S, Cotelli F, Beltrame M. Ectopic expression and knockdown of a zebrafish sox21 reveal its role as a transcriptional repressor in early development. Mech Dev. 2004 Feb 1;121 (2):131-42.

Barrionuevo F, Scherer G. SOX E genes: SOX9 and SOX8 in mammalian testis development. Int J Biochem Cell Biol. 2010 Mar 3;42(3):433-6.

Cui J, Shen X, Zhao H, Nagahama Y. Genome-wide analysis of Sox genes in Medaka (Oryzias latipes) and their expression pattern in embryonic development. Cytogenet Genome Res. 2011 Aug;134(4):283-94.

Fortna A, Kim Y, MacLaren E, Marshall K, Hahn G, Meltesen L, Brenton M, Hink R, Burgers S, Hernandez-Boussard T, Karimpour-Fard A, Glueck D, McGavran L, Berry R, Pollack J, Sikela JM. Lineage-specific gene duplication and loss in human and great ape evolution. PLoS Biol. 2004 Jul 13;2(7): 18 p.

Georg I, Barrionuevo F, Wiech T, Scherer G. Sox9 and Sox8 are required for basal lamina integrity of testis cords and for suppression of FOXL2 during embryonic testis development in mice. Biol Reprod. 2012 Oct 1;87(4):99-101.

Han F, Wang Z, Wu F, Liu Z, Huang B, Wang D. Characterization, phylogeny, alternative splicing and expression of Sox30 gene. BMC Mol Biol. 2010 Dec 11;11(1): 11 p.

Han Z, Liu Y, Deng X, Liu D, Liu Y, Hu Y, Yan Y. Genome-wide identification and expression analysis of expansin gene family in common wheat (Triticum aestivum L.). BMC Genomics. 2019 Feb 1;20: 19 p.

Hett AK, Ludwig A. SRY-related (Sox) genes in the genome of European Atlantic sturgeon (Acipenser sturio). Genome. 2005 Apr;48(2):181-6.

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L. The zebrafish reference genome sequence and its relationship to the human genome. Nature. 2013 Apr;496(7446):498-503.

Jeng S, Wu G, Yueh W, Kuo S, Dufour S, Chang C. Gonadal development and expression of sex-specific genes during sex differentiation in the Japanese eel. Gen Comp Endocrinol. 2018 Feb 1;257:74-85.

Jiang T, Hou CC, She ZY, Yang WX. The SOX gene family: Function and regulation in testis determination and male fertility maintenance. Mol Biol Rep. 2012 Nov 27;40 (3):2187-94.

Kan L, Israsena N, Zhang Z, Hu M, Zhao LR, Jalali A, Sahni V, Kessler JA. Sox1 acts through multiple independent pathways to promote neurogenesis. Dev Biol. 2004 May 15; 269(2):580-94.

Kanda H, Kojima M, Miyamoto N, Ito M, Takamatsu N, Yamashita S, Shiba T. Rainbow trout Sox24, a novel member of the Sox family, is a transcriptional regulator during oogenesis. Gene. 1998 May 12;211(2):251-7.

Kikuchi Y, Agathon A, Alexander J, Thisse C, Waldron S, Yelon D, Thisse B, Stainie DY. Casanova encodes a novel Sox-related protein necessary and sufficient for early endoderm formation in zebrafish. Genes Dev. 2001 Jun 15;15(12):1493-505.

Koopman P, Schepers G, Brenner S, Venkatesh B. Origin and diversity of the Sox transcription factor gene family: Genome-wide analysis in Fugu rubripes. Gene. 2004 Mar 17;328:177-86.

Larney C, Bailey TL, Koopman P. Switching on sex: Transcriptional regulation of the testis-determining gene Sry. Development. 2014 Jun;141(11):2195-205.

Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A, Grammes F, Grove H, Gjuvsland A, Walenz B, Hermansen RA, von Schalburg K, Rondeau EB, Di Genova A, Samy JK, Olav Vik J, Vigeland MD, Caler L, Grimholt U, Jentoft S, Vage DI, de Jong P, Moen T, Baranski M, Palti Y, Smith DR, Yorke JA, Nederbragt AJ, Tooming-Klunderud A, Jakobsen KS, Jiang X, Fan D, Hu Y, Liberles DA, Vidal R, Iturra P, Jones SJ, Jonassen I, Maass A, Omholt SW, Davidson WS. The Atlantic salmon genome provides insights into rediploidization. Nature. 2016 Apr;533(7602):200-5.

Ma F, Liu Z, Kang Y, Quan J. Genome-wide identification of hsp90 gene in rainbow trout (Oncorhynchus mykiss) and their regulated expression in response to heat stress. DNA Cell Biol. 2020 Mar 5;39(3):428-40.

Mukherjee A, Shan X, Mutsuddi M, Ma Y, Nambu JR. The Drosophila sox gene, fish-hook, is required for postembryonic development. Dev Biol. 2000 Jan 1;217(1):91-106.

Nakamura S, Watakabe I, Nishimura T, Toyoda A, Taniguchi Y, Tanaka M. Analysis of medaka sox9 orthologue reveals a conserved role in germ cell maintenance. PLoS One. 2012 Jan 12;7(1): 12 p.

Niimura Y, Matsui A, Touhara K. Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. Genome Res. 2014 Jul 22; 24(9):1485-96.

https://doi.org/10.17221/4/2022-CJAS

Pan Q, Anderson J, Bertho S, Herpin A, Wilson C, Postlethwait JH, Schartl M, Guiguen Y. Vertebrate sex-determining genes play musical chairs. C R Biol. 2016 Jul-Aug 10; 339(7-8):258-62.

Schepers GE, Teasdale RD, Koopman P. Twenty pairs of Sox: Extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. Dev Cell. 2002 Aug 1;3(2):167-70.

Takamatsu N, Kanda H, Ito M, Yamashita A, Yamashita S, Shiba T. Rainbow trout SOX9: cDNA cloning, gene structure and expression. Gene. 1997 Nov 20;202(1-2):167-70.

Wan H, Han K, Jiang Y, Zou P, Zhang Z, Wang Y. Genome-wide identification and expression profile of the Sox gene family during embryo development in large yellow croaker. DNA Cell Biol. 2019 Aug 16;38(10):1100-11.

Wang Y, Ristevski S, Harley VR. SOX13 exhibits a distinct spatial and temporal expression pattern during chondrogenesis, neurogenesis, and limb development. J Histochem Cytochem. 2006 Jul 11;54(12):1327-33.

Wei L, Yang C, Tao WJ, Wang DS. Genome-wide identification and transcriptome-based expression profiling of the Sox gene family in the Nile tilapia (Oreochromis niloticus). Int J Mol Sci. 2016 Feb 23;17(3): 15 p.

Yamashita A, Suzuki S, Fujitani K, Kojima M, Kanda H, Ito M, Takamatsu N, Yamashita S, Shiba T. cDNA cloning of a novel rainbow trout SRY-type HMG box protein, rtSox23, and its functional analysis. Gene. 1998 Mar 16; 209(1-2):193-200.

Yu H, Du X, Li X, Qu J, Zhu H, Zhang Q, Wang X. Genome-wide identification and transcriptome-based expression analysis of sox gene family in the Japanese flounder Paralichthys olivaceus. J Oceanol Limnol. 2018 Sep 20;36 (5):1731-45.

Zafar I, Iftikhar R, Ahmad SU, Rather MA. Genome wide identification, phylogeny, and synteny analysis of sox gene family in common carp (Cyprinus carpio). Biotechnol Rep. 2021 Mar 16;30: 17 p.

Zhang S, Chen X, Wang M, Zhang W, Pan J, Qin Q, Zhong L, Shao J, Sun M, Jiang H, Bian W. Genome-wide identification, phylogeny and expressional profile of the Sox gene family in channel catfish (Ictalurus punctatus). Comp Biochem Physiol Part D Genom Proteom. 2018 Dec 1; 28:17-26.

Received: January 12, 2022
Accepted: April 19, 2022
Published online: April 30, 2022