RESEARCH ARTICLE

cDNA cloning and expression analysis of two distinct *Sox8* genes in *Paramisgurnus dabryanus* (Cypriniformes)

XIAOHUA XIA, JIE ZHAO, QIYAN DU and ZHONGJIE CHANG*

Molecular and Genetic Laboratory, College of Life Sciences, Henan Normal University, 46 East of Construction Road, Xinxiang, Henan 453007, People's Republic of China

Abstract

The *Sox9* gene attracts a lot of attention because of its connection with gonadal development and differentiation. However, *Sox8*, belonging to the same subgroup *SoxE*, has rarely been studied. To investigate the function as well as the evolutionary origin of *SOXE* subgroup, we amplified the genomic DNA of *Paramisgurnus dabryanu* using a pair of degenerate primers. Using rapid amplification of the cDNA ends (RACE), it was discovered that *P. dabryanu* has two duplicates: *Sox8a* and *Sox8b*. Each has an intron of different length in the conserved HMG-box region. The overall sequence similarity of the deduced amino acid of *PdSox8a* and *PdSox8b* was 46.26%, and only two amino acids changed in the HMG-box. This is the first evidence showing that there are two distinct duplications of *Sox8* genes in Cypriniformes. Southern blot analysis showed only one hybrid band, with lengths 7.4 or 9.2 kb. Both semi-quantitative RT-PCR and real-time quantitative PCR assay displayed that both *PdSox8a* and *PdSox8b* are downregulated during early embryonic development. In adult tissues, the two *Sox8* genes expressed ubiquitously, and expression levels are particularly high in the gonads and brain. In gonads, both *PdSox8a* and *PdSox8b* are expressed at a higher level in the tesis than in the ovary. *PdSox8a* and *PdSox8b* may have functional overlaps and are essential for the neuronal development and differentiation of gonads.

[Xia X., Zhao J., Du Q. and Chang Z. 2010 cDNA cloning and expression analysis of two distinct *Sox8* genes in *Paramisgurnus dabryanus* (Cypriniformes). *J. Genet.* **89**, 183–192]

Introduction

An important advance in the research of mammalian sex determination mechanism was made in 1990 when the sex-determining region of the Y chromosome (*Sry*) was first discovered and regarded as the mammalian Y-linked testis determining gene (Gubbay *et al*. 1990). The SRY protein contains a DNA binding domain known as the HMG-box, which binds to the sequence (A/T)AACAAT (Harley *et al*. 1994). All SOX protein family members contain the SRY-type conservative HMG-box and have diverse functions ranging from a role in early embryogenesis to functions in lineage specification and terminal differentiation events (Kiefer 2007). The SOX protein family constitutes a large group of transcription factors. Based on the differing levels of structural and organizational similarities, and likely evolutionary relatedness, the *Sox* family has been subdivided into 10 subgroups: A–J (Bowles *et al*. 2000).

The *SoxE* group is composed of three members: *Sox8, Sox9* and *Sox10*. Gain-of-function studies in mice and humans indicate that *Sox9* is suf cient for male sex determination even in the absence of *Sry*, and *Sox9* is a critical component of the male sex-determining pathway (Canning and Lovell-Badge 2002). *Sox10* is clearly an important transcriptional regulator in neural crest cell development (Kelsh 2006), and its expression is similar to *Sox8* and *Sox9* during mouse embryonic testis development (Cory *et al*. 2007). Recently, *Sox8* was identified as the third member of *SoxE*. Initially, *Sox8* was proposed as a regulator of male sex determination, testicular differentiation or germ cell development (Schepers *et al*. 2002, 2003). However, $Sox8^{-/-}$ mice showed decreased adiposity and premature osteoblast differentiation that resulted in poor tarsal development and low bone density (Sock *et al*. 2001; Schmidt *et al*. 2005). Till today, the exact functions of *Sox8* have not been identified. Expression patterns of *Sox8* are diverse, but strong *Sox8* expression was apparent in brain and testes of adult tissues (Schepers *et al*. 2000).

*For correspondence. E-mail: changzhongjie@tom.com.

Keywords. RACE; real-time quantitative PCR; semi-quantitative RT-PCR; *Sox8a*; *Sox8b*; *Paramisgurnus dabryanus.*

More than 24,000 species of fishes inhabit a wide range of aquatic habitats worldwide (Kirpichnikov 1981), with wide variety of biological characteristics and great economic value. *Paramisgurnus dabryanu* (Cypriniformes; Cobitidae) is an emerging aquaculture species which has two duplicates *Sox8a* and *Sox8b*, and there are very few available reports of two orthologues of *Sox8* in vertebrates. To facilitate the study of *PdSox8a* and *PdSox8b* genes, we investigated their expression patterns in embryogenesis and various adult tissues in *P. dabryanu* using semi-quantitative RT-PCR and real-time quantitative RT-PCR. The results are interpreted in terms of the evolution and origin of *Sox* genes.

Materials and methods

Fish stocks

Adult *P. dabryanu* was collected from wetlands in the old course of Yellow River, Yanjin (Henan, P. R. China). Females and males were induced to mate by intramuscular injection of human chorionic gonadotropin hormone (HCG). Fertilized eggs were allowed to develop further until use.

Preparation of genomic DNA and PCR amplification of HMG-box

Genomic DNA was isolated from blood samples using the protease K digestion and phenol–chloroform extraction method (Sambrook *et al*. 1989). The purified DNA was quantified and stored at −20°C. Based on the conserved HMG-box of *SoxE*, a pair of degenerate primers were designed.

HMG-forward: 5′-T(C\G)AA(G\A)(A\C)G(G\A)CC (C\T)ATGAA(C\T)GC-3′; HMG-reverse: 5′-G(G\A)TG (G\A)TC(C\T) TT(C\T)TTGTGCTG-3′. PCR reaction was performed in a total volume 20 μL containing 0.2 mM dNTP, 2.0 mM $MgCl_2$, 20 μM each primer, 1 U *Taq* DNA polymerase and 1 μg genomic DNA template. Amplification conditions were: 95°C for 5 min, followed by 35 cycles at 94°C (40 s), 56°C (60 s) and 72°C (60 s), ending with an extension at 72°C for 10 min. PCR products were separated by electrophoresis on 1.5% agarose gels and visualized by scanning with a UV imaging system (Gucun, Shanghai, China).

Preparation of total RNA and RT-PCR

Total RNA extraction kit from RNAiso reagent (TaKaRa, Shiga, Japan) was used to extract RNA. Approximately 25 mg of mature adult tissues at different stages of developing embryo were homogenized using a glass tissue homogenizer. The concentration and purity were determined by the BioPhotometer (Eppendorf, Hamburg, Germany). For the expression profile study, the first-strand cDNA synthesis was carried out with the PrimeScript Reverse Transcriptase (TaKaRa, Shiga, Japan), using 1 μg of total RNA isolated from various stages of embryos, and different adult tissues.

Cloning the full-length cDNA of PdSox8a and PdSox8b

The RACE Core Set (TaKaRa, Shiga, Japan) was used to produce 5′ and 3′ RACE cDNA, with the gene-specific primers (table 1) from HMG-box sequence of *Sox8*. The reverse transcription reactions were performed with 3–5 μg total RNA isolated from adult brain. The RACE products were gel-purified, cloned into pGEM-T vector (Promega, Madison, USA), and then sequenced.

Southern blot analysis

About 5 μg of genomic DNA extracted from individual fish was digested with restriction enzymes *Bam*HI (6 U/μg of DNA), separated on 0.8% agarose gels, and then blotted onto nylon membranes. Membranes were hybridized under high stringency (hybridization in 35% formamide, 0.1% Na-pyrophosphate, 50 mM Tris-Cl (pH 7.5), 5× SSC, 1% SDS, 5× Denhardt's, 100 μg/mL calf thymus DNA at 40°C, washing in 1× SSC 1% SDS at 60°C) with random primer-labelled *Sox8a* or *Sox8b* probes. The blot was exposed to X-ray film for 3–7 days at −20°C with an intensifying screen.

Data analysis

The resulting sequences were confirmed by the BLASTx program on the NCBI Blast Server (http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi). The deduced amino acid sequences of *PdSox8a* and *PdSox8b* were aligned with those of other vertebrate *SoxE* members taken from GenBank, using ClustalW multiple alignment program software (http://www.ebi.ac.uk/Tools/clustalw2/index.html). A phylogenetic tree was constructed by the neighbour-joining algorithms in the MEGA3 program (Kumar *et al*. 2004).

Semi-quantitative reverse transcription (RT)-PCR

The semi-quantitative RT-PCR reactions were performed to analyse the expression profiles in different tissues and at various stages of embryo. Two pairs of specific primers were designed on the basis of the 3′ non-conserved regions of *PdSox8a* and *PdSox8b* gene, respectively (table 2). *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase) gene was used as the referential gene. The relative optical densities of the bands were determined by Band Leader software (Magnitec, Tel Aviv, Israel).

Fluorescent real-time quantitative reverse transcription (RT)-PCR

To explore a particular expression profile of *Sox8* genes, the real-time RT-PCR assay was carried out in an ABI 7500 real-time PCR system (Applied Biosystems, Foster City, USA) with SYBR green fluorescent label. *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase) gene was also used to normalize data for differences among samples. The primers used are shown in table 3. The amplifications were performed in a 50 μL reaction volume containing 12.5 μL of 2× SYBR Premix ExTaq (TaKaRa, Shiga, Japan), 3–5 pm

Table 1. Primers for RACE-PCR amplification.

Names	Oligonucleotide sequence ($5' \rightarrow 3'$)	Length (bp)
3′GSP1 Sox8a	GCAAACTCTGGAGACTGTTGAC	22
3′GSP2 Sox8a	AAGAAGCCGAGAGACTGAGAG	21
3′GSP1 Sox8b	TCTGAGAATGAGAAGAGACCGTT	23
3′GSP2 Sox8b	TAGAAGAGGCAGAAAGACTGAG	22
3′RACE outer	TACCGTCGTTCCACTAGTGATTT	23
3′RACE inner	CGCGGATCCTCCACTAGTGATTTCACTATAGG	32
5′GSP1 SOX8a	AAGAGACCGTTTGTGGAAGAAGC	23
5′GSP2 SOX8a	GCAAAACCCTCGGCAAACTCTGGA	24
5′GSP1 SOX8b	CATCAGAAAATGTGAGGCAGAACA	24
5′GSP2 SOX8b	GAGGGTCCAGCACAAGAAAGACT	23
5′RACE outer	CATGGCTACATGCTGACAGCCTA	23
5′RACE inner	CGCGGATCCACAGCCTACTGATGATCAGTCGATG	34

Table 2. Primers for semi-quantitative RT-PCR analysis.

Name	Oligonucleotide sequence ($5' \rightarrow 3'$)	product size (bp)
Sox8a SQ-for	CGAGAGAAGACGCCTGCT	
		427
Sox8a SQ-rev	CGTGTTGGAGAATGAGGG	
Sox8b SQ-for	TCAGCATCAAGGAGCGAAAC	
		451
Sox8b SQ-rev	GTAGGTCAGAACAGTCAGTC	
GAPDH SQ-for	ACCAACTGCTTGGCTCCCC	254
GAPDH SQ-rev	GGAATGATTTGCCCACG	

Table 3. Primers for fluorescent quantitative RT-PCR analysis.

Name	Oligonucleotide sequence ($5' \rightarrow 3'$)	product size (bp)
Sox8a FQ-for	ATTCCTCACGGAGACCCATACT	
		98
Sox8a FQ-rev	CCTGGATAATGTGGTGTAAACCG	
Sox8b FQ-for	CAAATGAGTCAAGCCAACAAAG	
		145
Sox8b FQ-rev	GTGTAGGTCAGAACAGTCAGTCAG	
GAPDH FQ-for	ACCAACTGCTTGGCTCCCC	254
GAPDH FQ-rev	GGAATGCTTTGCCCACG	

of each primer, 2 μL of diluted cDNA, 0.5 μL 50× ROX Reference DyeII (TaKaRa, Shiga, Japan). The PCR amplification was carried out as follows: 95°C for 10 s followed by 40 cycles of 95°C for 5 s and 60°C for 34 s. Data were analysed with 7500 system SDS Software v 1.4 (Applied Biosystems, Foster City, USA). The $2^{-\triangle\triangle Ct}$ method was used to analyse the expression level of *Sox8a* and *Sox8b* (Livak and Schmittgen 2001).

Results

Cloning and structural analysis of the SoxE HMG-box

With the degenerate primers set specific to the HMG-box motif of *SoxE*, four fragments of length 369 bp, 436 bp, 487 bp and 538 bp were amplified from genomic DNA of *P. dabryanus* (figure 1). No sex-specific bands were observed. The amplified fragments were purified and cloned.

Using PCGENE program (Intelligenetics, Mountainview, USA), the four sequences were analysed and predicted. It was found that the fragments 436 bp and 538 bp encoded 64 amino acids but only two were different, with the possibility that the two fragments encode same gene differing in possessing introns of 241 bp and 343 bp, respectively. Both of the splicing sites of the introns were consistent with the 'GT–AG' rule (figures 2 and 3). Homologous search with the tBLASTx program on the NCBI Blast server (http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi) showed that the sequences were similar to *SoxE* subgroup genes in different animals. In order to obtain complete information of the two fragments 5′ and 3′ , RACE was carried out.

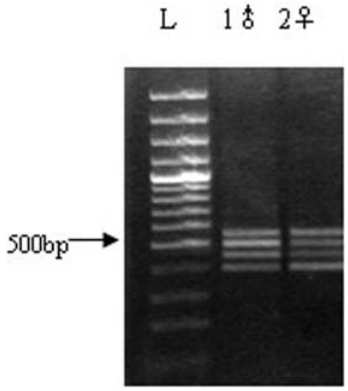

Figure 1. DNA fragments of HMG-box conserved regions amplified by degenerate primers. L, 100-bp DNA ladder; 1, male *P. dabryanus*; 2, female *P. dabryanus*.

Cloning of the full-length cDNA and sequence alignments

Employing the RACE strategy, the full-length cDNA of two fragments 436 bp and 538 bp were cloned. The 5′-RACE and 3′-RACE results were sequenced and spliced to obtain the full-length cDNA (figures 4 and 5). The 436-bp fragment cDNA spans 2467 bp with a 1287-bp open reading frame (ORF), a 53-bp 5′ untranslated region and a 1127-bp 3′ untranslated region including poly (A), encodes a putative protein of 428 amino acids with a characteristic HMG box DNA binding domain of 79 amino acids (aa: 86-164) (figure 6A). While another fragment cDNA, spans 2171 bp containing a 1053-bp ORF, a 201-bp 5′ untranslated region and a 917-bp 3′ untranslated region including poly (A), encodes a putative protein of 350 amino acids with a characteristic HMG box DNA binding domain of 79 amino acids (aa:

```
1       TGAAGCGACCGATGAACGCCTTTATGGTTTGGGCTCAAGCCGCGCGCAGGAAACTGGCAG
1          K  R  P  M  N  A  F  M  V  W  A  Q  A  A  R  R  K  L  A
61      ACCAATATCCTCACCTGCACAACGCCGAGCTGAGCAAAACCCTCGGCAAACTCTGGAGAC
21      D  Q  Y  P  H  L  H  N  A  E  L  S  K  T  L  G  K  L  W  R
        gtaattcatatatcaacaactcgtttattggttgatgtatagttttaaagcattaaagtaccacttaaatct
        ttgtagtgaaactatggcatttggcagtccatctgcagccaatatccatggtactttgactgtaataaaacc
        ctaattaacatttcgttattgttattataataataaatggattattagcttactctaaaataaacaaggttg
        atgatgatcagaactctcttatcag
121     TGTTGACAGAGAACGAAAAGAGACCGTTTGTGGAAGAAGCCGAGAGACTGAGAGTCCAGC
41       L  L  T  E  N  E  K  R  P  F  V  E  E  A  E  R  L  R  V  Q
181     ACAAGAAAGATTATC
61      H  K  K  D  Y
```

Figure 2. The DNA sequence and possible amino acid sequence of fragment 436 bp in *P. dabryanus*. Underlined sequences were primers used for PCR reaction. Lower case letters indicate introns.

```
1       TAAAGAGACCCATGAACGCTTTTATGGTGTGGGCGCAGGCTGCGCGCAGAAAACTGGCAG
1          K  R  P  M  N  A  F  M  V  W  A  Q  A  A  R  R  K  L  A
61      ATCAGTACCCCCATCTTCATAACGCAGAGCTCAGCAAGACCTTGGGCAAGTTGTGGAGGC
21      D  Q  Y  P  H  L  H  N  A  E  L  S  K  T  L  G  K  L  W  R
        gtaagtgggtacatttttgaaattctatttttgtaaaactatttaatatgatgttatgaatgtatttgcatgtaacatgaacaatcccc
        tgcaagaacttttgtcttactacatgaagtctaactgtgatatttgtagttggtaatataaatgataatcagtctgcatatctgcca
        tggttactgcactttactaaggctgtaaattgtaatgtgcagtaatctcatcaaacaaaacttgttttggctcttatttttcaaatca
        aatatttttatttactgcacaacatttcaattggataataaaaactacattcaatcatgtaatttgtaaatgtccag
121     TGCTCTCTGAGAATGAGAAGAGACCGTTTATAGAAGAGGCAGAAAGACTGAGGGTCCAGC
41       L  L  S  E  N  E  K  R  P  F  I  E  E  A  E  R  L  R  V  Q
181     ACAAGAAAGACTATC
61      H  K  K  D  Y
```

Figure 3. The DNA sequence and possible amino acid sequence of fragment 538 bp in *P. dabryanus*. Underlined sequences are primers used for PCR reaction. Lower case letters indicate introns.

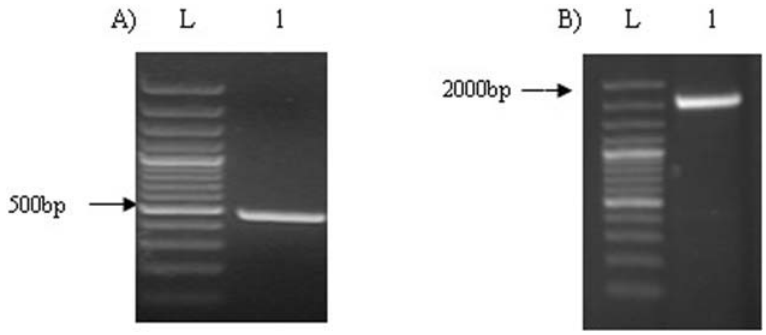

Figure 4. The Results of RACE-PCR on *PdSox8a* (436-bp fragment). A, the result of 5′-RACE; B, the result of 3′-RACE; L, 100-bp DNA ladder.

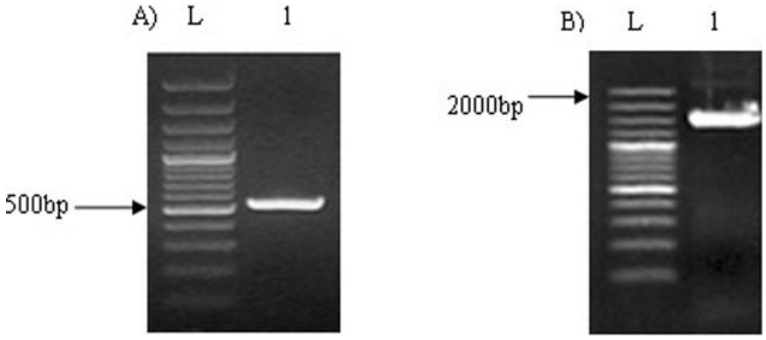

Figure 5. The results of RACE-PCR on *PdSox8b* (538-bp fragment). A, the result of 5′-RACE; B, the result of 3′-RACE; L, 100-bp DNA ladder.

89-167) (figure 6B). Comparing the deduced full-length amino acid sequences with each other, two proteins were highly conserved in HMG-box, only two amino acids have been changed (figure 7). The overall sequence identity of deduced amino acid sequences encoded by them was 46.26%. BLAST analysis showed that these two sequences were highly similar to the *Sox8* in the GenBank database. Consequently, these two fragments were named *PdSox8a* and *PdSox8b*.

Southern blot results

To detect the differences of these two genes in genome, four individual genomic DNA, two male and two female were analysed by Southern blot. In the throughly digested-samples with *Bam*HI, the two genes showed one hybrid band, respectively, with length 7.4 kb or 9.2 kb. No sex-specific band was observed. These results suggested the existence of two *Sox8* genes, *Sox8a* and *Sox8b* on genomic level (figure 8).

Homology and phylogenetic analysis of vertebrate Sox8 genes

To examine the phylogenetic relationships of *Sox8* genes between *P. dabryanus* and other vertebrates, a phylogenetic tree for *Sox8* gene was constructed using the neighbour-joining method by MEGA3 and ClustalW (figure 9). The result revealed that *Sox8* gene is quite highly conserved across species, and *Pdsox8a* and *PdSox8b* were clustered into same class with all the other *Sox8* genes. So, these two genes were indeed *Sox8* genes.

PdSox8a and PdSox8b tissue expression and developmental ontogeny

Transcriptional activities of *PdSox8a* and *PdSox8b* were obtained by semi-quantitative RT-PCR and fluorescent real-time RT-PCR with the primers outside the HMG-box motif. Both of them showed that *PdSox8a* are noticeably downregulated during the early period of embryonic development, with the highest level in gastrulae and lowest in yolk-sac absorption phase. The expression level of *PdSox8b* was lower than the former, except in the neurula period, with slight downtrend in other periods. In adult tissues, two *Sox8* genes were expressed ubiquitously but it is noteworthy that two *Sox8* genes have high expression levels in gonads and brain. In the brain, the expression levels of both *PdSox8a* and *PdSox8b* are similar, but in the gonads both have a higher level expression in the tesis than the ovary (figures 10 and 11).

Discussion

In this study, it was found that there were two distinct duplications of *Sox8* in *P. dabryanus: PdSox8a* and *PdSox8b*,

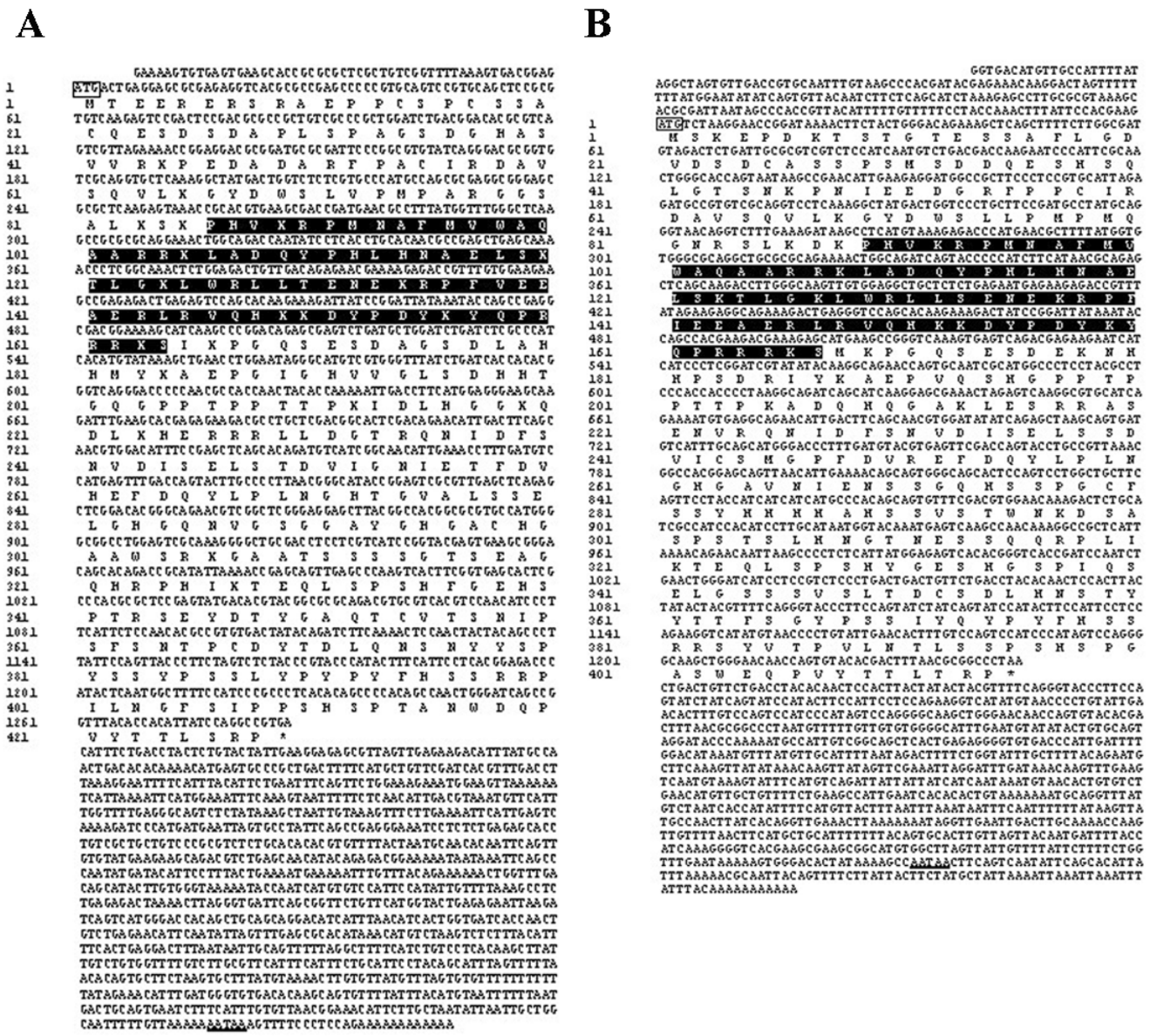

Figure 6. Nucleotide sequence and deduced amino acid sequence of *PdSox8* cDNAs. A, nucleotide sequence and deduced amino acid sequence of *PdSox8a* cDNA. B, nucleotide sequence and deduced amino acid sequence of *PdSox8b* cDNA. The HMG-box conserved region is indicated by white letters in black boxes. ATG shows initiation codon. An asterisk indicates the stop codon. The putative polyadenylation signal AATAA is underlined.

by RACE and Southern blotting. Sequence alignments indicated that the overall sequence similarity of deduced amino acid encoded by *PdSox8a* and *PdSox8b* was 46.26%, and there were only two amino acids changes in the HMG-box. The phylogenetic analysis of *SoxE* showed that (i) *Sox8* gene is highly conserved across species, implying that perhaps the SOX8 proteins are conserved in the course of evolution and play some fundamental role in vertebrate development; (ii) *Sox8a/Sox8b* in *P. dabryanus* are all orthologues of mammalian *Sox8* gene. Both of them were in a separate clade from Sox9 or Sox10 proteins. It can thus be inferred that *PdSox8a* and *PdSox8b* diverged from a common ancestor after the origin of *SoxE*.

Some studies reported that there are two duplicates of *Sox9* and *Sox11* in the zebra fish (Martino *et al*. 2000; Chiang *et al*. 2001). Later, it was found that two duplicates of *Sox8* were discovered in *Fugu rubripes* (Koopman *et al*. 2004). The identity of overall deduced amino acid encoded by *TrSox8a* and *TrSox8b* was 42.55%. At present, there are few available reports of two orthologues of *Sox8* in other teleosts. In this study, we found that there are also two distinct classes of *Sox8* in *P. dabryanus*: *Sox8a* and *Sox8b*. This is the first

```
PdSox8a     MTEERERSRAEPPCSPCSSACQESDSDAPLSPAGSDGHAS-----VVRKP-EDADARFPA 54
PdSox8b     MSKEPDKNSTG---TESSAFLGDVDSDCASSPSMSDDQESHSQLGTSNKPNIEEDGRFPP 57
            *::* ::. :    : .*:   : ***.. **: **.: *     . .**  : *.***.
PdSox8a     CIRDAVSQVLKGYDWSLVPMPARGGSALKSKPHVKRPMNAFMVWAQAARRKLADQYPHLH 114
PdSox8b     CIRDAVSQVLKGYDWSLLPMPMQGNRSLKDKPHVKRPMNAFMVWAQAARRKLADQYPHLH 117
            ******************:*** :*. :**.******************************
PdSox8a     NAELSKTLGKLWRLLTENEKRPFVEEAERLRVQHKKDYPDYKYQPRRRKSIKPGQSESD- 173
PdSox8b     NAELSKTLGKLWRLLSENEKRPFVEEAERLRVQHKKDHPDYKYQPRRRKSMKPDQSEPDE 177
            ***************:*********************:***********:**.***.*
PdSox8a     AGSDLAHHMYKAEPGIGHVVGLSDHHTGQGPPTPPTTPKIDLHGGKQDLKHERRRLLDGT 233
PdSox8b     KNHHPTDHIYKAEP---------VQSHGPPTPPTTPKADQHQG---AKLESRRASENV 223
             . . :.*:*****            .:********** * * *    * * **  :..
PdSox8a     RQNIDFSNVDISELSTDVIGNIETFDVHEFDQYLPLNGHTGVALSSELGHGQNVGSGGAY 293
PdSox8b     RQNIDFSNVDISELSSDVICSMGPFDVREFDQYLPLNGHGAVNIENGGGQHSSPGCFSTY 283
            ***************:*** .: .***:*********** .* :..  *: .. *. .:*
PdSox8a     GHGACH--GAAWSRKGAATSSS--SGTSEAGQHRPHIKTEQLSPSHFGEHSPTRSEYDTY 349
PdSox8b     HHHHAHSGGSTWNKDSGSPSTSLHNGTNESSQQRPLIKTEQLSPSHYGESHGSP----- 337
             * .*  *::*.:...:.*:*  .**.*:.*:** **********:**   :
PdSox8a     GAQTCVTSNIPSFSNTPCDYTDLQNSNYYSPYSSYPSSLYPYPYFHSSRR---PILNGF 405
PdSox8b     -----IQSELGSSSNSLTDCSDLHNSTYYTTFPGYPSSIYQYPYFHSSRRSYVAPVLNTL 392
                 : *:: * **:  * :**:**.**:.:..****:* *********    *:** :
PdSox8a     SIPPSHSPTANWDQPVYTTLSRP 428
PdSox8b     SSP-SHSPGASWEQPVYTTLTRP 414
            * * **** *.*:*******:**
```

Figure 7. Alignment of the deduced amino acid sequences of *PdSox8a* and *PdSox8b*. The sequences in dark shadings are the HMG-box motif. Stars indicate identical amino acids.

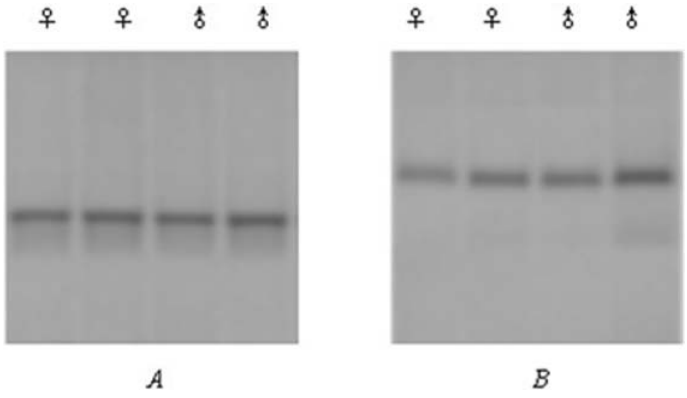

Figure 8. Southern blot for detection of *PdSox8a* and *PdSox8b* in genomic DNA. (A) *PdSox8a*; (B) *PdSox8b*.

evidence that there are two distinct duplications of *Sox8* genes in order Cypriniformes, the second report in Osteichthyes. All these research suggested that the duplications occurred in an ancestral fish linage

There are two classical theories of gene duplication. The subfunctionalization model predicts that the *Sox8* gene was parental to *Sox8a* and *Sox8b*; the functions of the single *Sox8* gene in the reduplication ancestor were the sum of the functions of the current *Sox8a* and *Sox8b* genes. The original *Sox8* would have had at least two essential functions (or essential expression domains) that are now split between the two derivative genes (Force *et al*. 1999). On the other hand, the neofunctionalization model predicts that at least one of the copies of the *Sox8* gene must have acquired a novel function induced by a beneficial mutation i.e., one or both of the new *Sox8* genes will have acquired a new function not

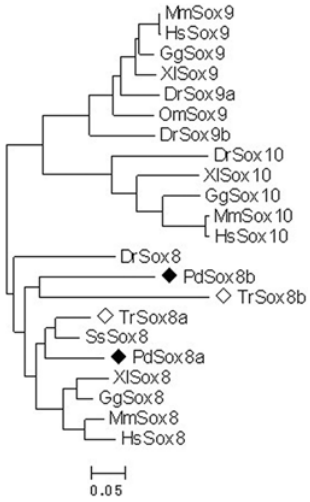

Figure 9. Phylogenetic study neighbour-joining phylogenetic tree of vertbrate SOX8 proteins. The accession numbers for the sequences are as follows: Pd, *Paramisgurnus dabryanus*; DrSox8, *Daniorerio* (AAX73357); TrSox8a, *Takifugu rubripes* (AAQ18505); TrSox8b, *Takifugu rubripes* (AAQ18506); SsSox8, *Salmo salar* (ABC24688); XlSox8, *Xenopus laevis* (AAI69525); GgSox8, *Gallus gallus* (AAF73917); MmSox8, *Mus musculus* (AAH85619); HsSox8, *Homo sapiens* (AAH31797); DrSox9a, *Danio rerio* (NP571718); DrSox9b, *Danio rerio* (NP 571719); OmSox9, *Oncorhynchus mykis* (NP001117651); XlSox9, *Xenopus laevis* (NP001084276); GgSox9, *Gallus gallus* (NP989612); MmSox9, *Mus musculus* (NP035578); HsSox9, *Homo sapiens* (NP000337); DrSox10, *Danio rerio* (AAI63883); XlSox10, *Xenopus laevis* (NP001082358); GgSox10, *Gallus gallus* (NP990123); MmSox10, *Mus musculus* (NP035567); HsSox10, *Homo sapiens (NP 008872)*.

present in the *Sox8* gene of the reduplication ancestor (Ohno 1970). Two orthologues of mammalian *Sox8* gene, *PdSox8a* and *PdSox8b*, not only lend strong support to a 'fish-specific whole-genome duplication' theory, but also infer further duplication of some genes.

The high expression level of *Sox8a* and lower level of *Sox8b* but the same trend in the early embryonic development stages may imply that their functions have already diverged to some degree. These results are consistent with a shared evolutionary origin; *SoxE* genes frequently display overlapping expression patterns of activity during mammalian embryonic development and have been shown to compensate for, or cooperate with, each other in models of gene mutation (O'Bryan *et al*. 2008). For example, a knockout of *Sox8* in mice has no obvious developmental phenotype and only idiopathic weight loss and reduced bone density (Sock *et al*. 2001), suggesting that *SoxE* genes are able to act redundantly in a variety of developmental contexts. It has also been hypothesized that *Sox8* may substitute for *Sox9* in species where *Sox9* is expressed too late to be involved in sex determination or regulation of *Amh* expression (Schepers *et al*. 2003; Takada and Koopman 2003). In adult tissues, *Sox8a* and *Sox8b* genes are expressed ubiquitously in various tissues, and their expression patterns in some tissues are very similar. This implies a possible functional redundancy. Abundant expression of *Sox8a* and *Sox8b* observed in the tesis, ovary and brain suggests that may be both of them are essential for the differentiation of gonads and for neuronal development. However, the exact functions of these two *Sox8* genes will need to be proved by further studies.

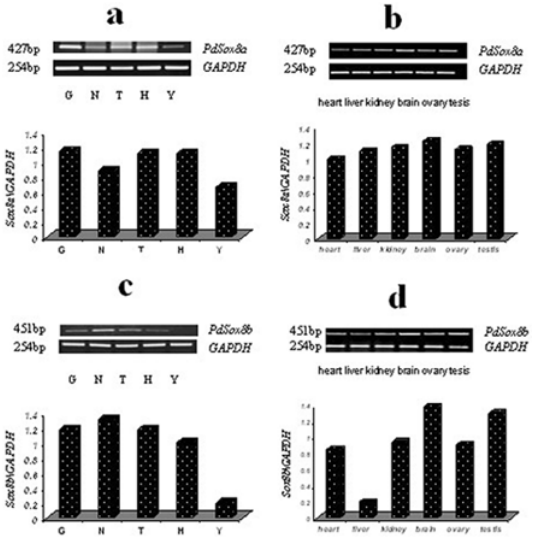

Figure 10. Expression analysis by semi-quantitative RT-PCR of *PdSox8a* and *PdSox8b* gene. a, Expression of *PdSox8a* in embryonic development stages; b, expression of *PdSox8a* in six adult tissues; c, expression of *PdSox8b* in embryonic development stages; d, expression of *PdSox8b* in six adult tissues; G, gastrulae; N, neurula; T, tail-bud formed; H, hatched larva, Y, yolk-sac absorption phase.

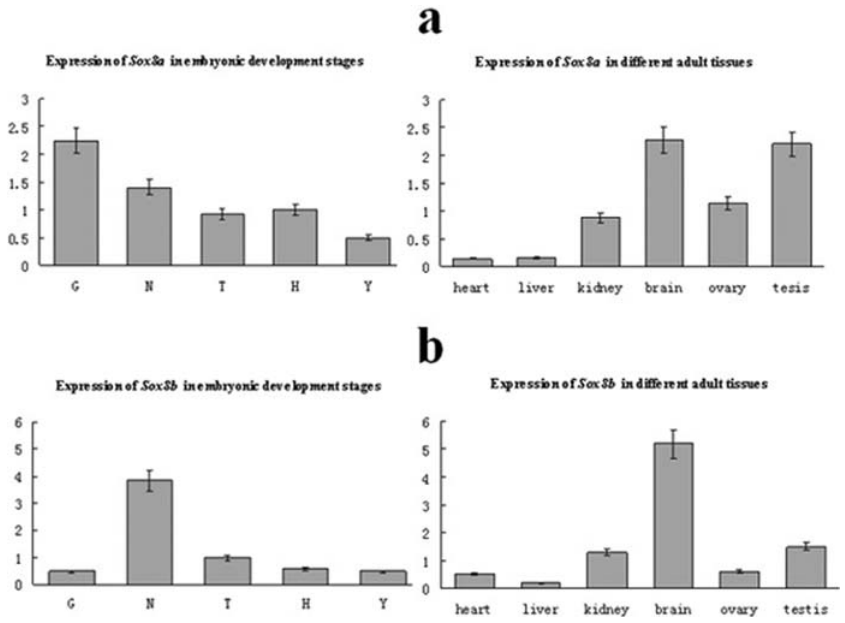

Figure 11. The expression analysis by real-time PCR of *PdSox8a* and *PdSox8b*. a, Expression of *Sox8a* in embryonic development stages and six adult tissues of *P. dabryanus*; b, expression of *Sox8b* in embryonic development stages and six adult tissues of *P. dabryanus*; G, gastrulae; N, neurula; T, tail-bud formed; H, hatched larva; Y, yolk-sac absorption phase.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (30771666).

References

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.

Canning C. A. and Lovell-Badge R. 2002 Sry and sex determination: how lazy can it be? *Trends Genet*. **18**, 111–113.

Chiang E. F., Pai, C. I., Wyatt M., Yan Y. L., Postlethwait J. and Chung B. 2001 Two *Sox9* genes on duplicated zebrafish chromosomes: expression of similar transcription activators in distinct sites. *Dev. Biol*. **231**, 149–163.

Cory A. T., Boyer A., Pilon N., Lussier J. G. and Silversides D. W. 2007 Presumptive pre-Sertoli cells express genes involved in cell proliferation and cell signalling during a critical window in early testis differentiation. *Mol. Reprod. Dev*. **74**, 1491–1504.

Force A., Lynch M., Pickett F. B., Amores A., Yan Y. L. and Postlethwait J. H. 1999 Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545.

Gubbay J., Collignon J., Koopman P., Capel B., Economou A., Münsterberg A. *et al*. 1990 A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346**, 245–250.

Harley V. R., Lovell-Badge R. and Goodfellow P. N. 1994 Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res*. **22**, 1500–1501.

Kelsh R. N. 2006 Sorting out *Sox10* functions in neural crest development. *Bioessays* **28**, 788–798.

Kiefer J. C. 2007 Back to basics: *Sox* genes. *Dev. Dyn*. **236**, 2356–2366.

Kirpichnikov V. S. 1981 *Genetic bases of fish selection*. Springer, New York, USA.

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

Kumar S., Tamura K. and Nei M. 2004 MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* **5**, 150–163.

Livak K. J. and Schmittgen T. D. 2001 Analysis of relative gene expression data using real-time quantitative PCR and the 2 (-Delta Delta C (T)) method. *Methods* **25**, 402–408.

Martino S. D., Yan Y. L., Jowett T., Postlethwait J. H., Varga Z. M., Ashworth A. *et al*. 2000 Expression of *Sox11* gene duplicates in zebra fish suggests the reciprocal loss of ancestral gene expression patterns in development. *Dev. Dyn*. **217**, 279–292.

O'Bryan M. K., Takada S., Kennedy C. L., Scott G., Harada S., Ray M. K. *et al*. 2008 *Sox8* is a critical regulator of adult sertoli cell function and male fertility. *Dev. Biol*. **316**, 359–370.

Ohno S. 1970 *Evolution by gene duplication*. Springer, Berlin, Germany.

Sambrook J., Fritsch E. F. and Maniatis T. 1989 *Molecular cloning: a laboratory manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, USA.

Schepers G. E., Bullejos M., Hosking B. M. and Koopman P. 2000 Cloning and characterisation of the sry-related transcription factor gene *Sox8. Nucleic Acids Res*. **28**, 1473–1480.

Schepers G. E., Teasdale R. D. and Koopman P. 2002 Twenty pairs of *Sox*: extent, homology, and nomenclature of the mouse and human *Sox* transcription factor gene families. *Dev. Cell* **3**, 167–170.

Schepers G. E., Wilson M., Wilhelm D. and Koopman P. 2003 *SOX8* is expressed during testis differentiation in mice and synergizes with SF1 to activate the *Amh* promoter *in vitro*. *J. Biol. Chem*. **278**, 28101–28108.

Schmidt K., Schinke T., Haberland M., Priemel M., Schilling A. F., Mueldner C. *et al*. 2005 The high mobility group transcription factor *Sox8* is a negative regulator of osteoblast differentiation. *J. Cell Biol*. **168**, 899–910.

Sock E., Schmidt K., Hermanns-Borgmeyer I., Bosl M. R. and Wegner M. 2001 Idiopathic weight reduction in mice deficient in the high-mobility-group transcription factor *Sox8. Mol. Cell. Biol*. **21**, 6951–6959.

Takada S. and Koopman P. 2003 Origin and possible roles of the *Sox8* transcription factor gene during sexual development. *Cytogenet. Genome Res*. **101**, 212–218.

Received 16 December 2009, in revised form 2 March 2010; accepted 22 March 2010

Published on the Web: 6 August 2010