Isolation and sequencing of the HMG domain of ten *Sox* genes from *Odorrana schmackeri* (Amphibia: Anura)

Ning Wang [1]; Jing J. Wang [1]; Rui Jia [1]; Jin W. Xu [1] & Liu W. Nie [1, 2]

[1] *The Provincial Key Laboratory of the Conservation and Exploitation Research of Biological Resources in Anhui, College of Life Sciences, Anhui Normal University, Wuhu, 241000, Anhui, China. E-mail: lwnie@mail.ahnu.edu.cn*
[2] *Corresponding author.*

ABSTRACT. *Sox* (SRY-related HMG-box) genes encode a family of transcriptional regulators, which are characterized by a conserved 79-amino acid domain known as HMG-box. They play essential roles in a diverse range of processes including sex determination and the development of the central nervous system (CNS), neural crest and endoderm. In this paper, the HMG domain of ten distinct *Sox* gene family members (*os-Sox2*, *os-Sox3a*, *os-Sox3b*, *os-Sox4*, *os-Sox11a*, *os-Sox11b*, *os-Sox14a*, *os-Sox14b*, *os-Sox21a*, *os-Sox21b*) were isolated from both male and female *Odorrana schmackeri* (Boettger, 1892) using PCR, and no sexual differences were found. Molecular phylogenetic analysis of the HMG domain suggested that these ten *Sox* genes are members of the *SoxB* and *Sox*C groups. In addition, sequence analysis suggested that four *Sox* genes (*os-Sox3*, *os-Sox11*, *os-Sox14*, *os-Sox21*) were duplicated. The duplication-degeneration-complementation model should be implied to explain the evolution and diversity of the *Sox* gene family in *O. schmackeri*.

KEY WORDS. Duplication-degeneration-complementation; SSCP.

Gene members belonging to the *Sox* family are characterized by a recognisable 240-nucleotide sequence that encodes a 79-amino acid motif known as the HMG-box domain. This domain exhibits an L-shaped structure containing three α-helices and an N-terminal β-strand. It binds DNA in the minor groove, inducing a significant bend of the DNA helix (Lefebvre *et al.* 2007). Many HMG-box genes act as transcription factors regulating gene expression during developmental patterning or cell differentiation. Based on their characteristic primary structures, mammalian HMG proteins are classified into three structurally distinct classes: the HMG-nucleosome binding family (HMGN), the HMG-AT-hook family (HMGA), and the HMG-box family (HMGB) (Maruyama *et al.* 2005). Furthermore, Stros *et al.* (2007) subdivided the mammalian HMGB proteins into two major groups. One is the HMGB-type non-sequence-specific DNA binding proteins with two HMG-box domains and a long highly acidic C-tail, whereas the other is a more diverse group composed of proteins having mostly a single HMG-box and no acidic C-tails, although some of them may have even up to six copies of the HMG-box domain (e.g.UBF). SOX family proteins, as they have only one HMG-box and bind DNA at the common core consensus sequence AACAAT, can be grouped into the last group of HMGB (Mertin *et al.* 1999).

Sox genes seem to be restricted to animals and have been found in various species. In mammals, more than 30 *Sox* genes have been identified. Together with those metazoan orthologues, the *Sox* genes have been divided into ten subgroups: A, *Sry*; B, *Sox1*, *-2*,*-3*, *-14*, *-21*, *-25*; C, *Sox4*, *-11*, *-24*, *-22*; D, *Sox5*, *-6*, *-12*, *-13*, *-23*; E, *Sox8*, *-9*, *-10*; F, *Sox7*, *-17*, *-18*; G, *Sox15*, *-16*, *-20*; H, *Sox30*; I, *Sox31* and J, *Sox32*, *-33* (Bowles *et al.* 2000).

The SOX family of transcription factors plays key roles during development, including cell-fate determination of pluripotent cells, cell proliferation, differentiation, maturation and maintenance of stem cells during organogenesis (Lefebvre *et al.* 2007). Remarkably, *Sox* genes are involved in stemness and in the control of embryonic stem (ES) cells differentiation into tissue-specific cells, which are two important fields of research. To date, many *Sox* genes have been found to be involved in these two processes. Concerning stemness, mouse *Sox2* is thought to cooperate with *OCT4* (octamer-binding protein) in the early embryogenesis to regulate the gene expression in fertilized eggs (Li *et al.* 2007). In addition, ectopic expression of SOX2/*Sox2* is used to convert human somatic cells or mouse mature B lymphocytes to induce pluripotent stem (iPS) cells (Meissner *et al.* 2007, Park *et al.* 2008, Hanna *et al.* 2008). In contrast to *Sox2*, *Sox15* was found to exhibit different functions in the control of transcriptional processes in mouse ES cells, (Maruyama *et al.* 2005, Béranger *et al.* 2000). Vertebrate *Sox1*, *Sox2*, *Sox3* are required for stem-cell maintenance in the central nervous system (CNS), and their effects are counteracted by *Sox21* (Sandberg *et al.* 2005). With regard to differentiation, the *Sox* genes are involved for instance in mammalian testis determination that is known to be triggered by *SRY*. In addition, SOX9 mutations cause XY sex reversal in human (Barrionuevo *et al.* 2006), whereas *Sox8* reinforces *Sox9* function in testis differentiation of mice (Chaboissier *et al.* 2004).

Sox3 is important for normal oocyte development, for male testis differentiation, and for gametogenesis in mouse (Weiss *et al.* 2003). Other differentiation processes are also regulated by *Sox* genes. In mouse, *Sox2* regulates the differentiation of endodermal progenitor cells of the tongue into taste bud sensory cells versus keratinocytes (Suzuki 2008, Okubo *et al.* 2006) and *Sox4* facilitates thymocyte differentiation (Schilham *et al.* 2007).

Odorrana schmackeri (Boettger, 1892) (2n = 26), the piebald odorous frog (Amphibia: Anura: Ranidae), is endemic to China (Lau *et al.* 2004). Amphibians have evolved a large diversity of morphological changes that are different from aquatic vertebrate, including the tetrapod limb. They are a transitional group from aquatic to terrestrial in vertebrate evolution. Therefore, they play a key role in the analysis of the genetic basis of the morphological and lifestyle transition and the evolution of genes that function well in different animals (Mannaert *et al.* 2006). Given the importance of the *Sox* gene family and the function of growth regulation of *Sox* genes in different animals, we isolated and sequenced the HMG domain of ten *Sox* genes from *O. schmackeri*. Based on our results, we discuss the evolution and diversity of the *Sox* gene family.

MATERIAL AND METHODS

Isolation of the HMG domain of the *Sox* genes

To isolate the HMG domain of the *Sox* genes, two male and female *O. schmackeri* were captured from Huangshan, Anhui Province, China. Total genomic DNA was obtained from muscle tissues with the Genomic DNA Extraction Kit (Axygen). A pair of degenerate primers were designed according to the sequence of the HMG-box in multiple *Sox*/*SRY* genes (L1:5'-AGCGACCCA TGAAYGCNTTYATNG-3';L2:5'-ACGAGGTCGATAYTTRTARTYN GG-3'). The PCR was carried out in a 25 µl reaction mixture containing 16µl ddH_2O, 100 ng of genomic DNA, 1.5 mM Mg^{2+}, 200 µM of each dNTP, 0.2 µM of each primer and 1 unit of Taq DNA polymerase. The cycling conditions were 4 min at 95°C, followed by 5 cycles of 40s at 94°C, 40s at 48°C, 1 min 20 sec at 72°C then 30 cycles of 40s at 94°C, 40s at 52°C, 1 min 20 sec at 72°C. The final extension was done during 10 min at 72°C.

Screening and sequencing

The PCR products were detected on 1.8% agarose gels and cloned into a pMD18-T Vector. The positive clones were identified using colony PCR technique, with primers and reaction conditions as above (Shen *et al.* 2000). In order to identify different positive clones, the individual samples were further screened by SSCP (single-strand conformation polymorphism) analysis (Nie *et al.* 1999). The sequencing was done with universal sequencing primers on an ABI377 auto-sequencer.

Sequence and phylogenetic analysis

Except for *O. schmackeri*, all the sequences of *Sox* genes were obtained from GenBank. The consensus sequence was cited from Bowles *et al.* (2000). DNA sequences were analyzed using the basic local alignment search tool (BLAST) and CLUSTAL X1.8 programs. Boostrapping values were calculated using the modules SEQBOOT (1000 replicates), PROTDIST (distance estimation: Kimura-two parameter; analysis of 1000 data sets), NEIGHBOR (Neighbor-Joining method; outgroup: ye-MATA1; analysis of 1000 data sets) and CONSENSE (outgroup: ye-MATA1) of the PHYLIP (version 3.68) software package. The phylogenetic tree was computed with the same parameters as above. TreeView (version 1.6.6) was used for visualization and printing of the trees,.

RESULTS AND DISCUSSION

Isolation, nomenclature and analysis of the HMG domain of *Sox* genes

A 215 bp fragment was obtained from both male and female *O. schmackeri* genomic DNA using PCR technique. This fragment was gel purified and subcloned into pMD18-T Vector. After PCR screening of colonies, 150 positive clones were further screened with SSCP. Subsequently, 33 clones were sequenced and ten distinct sequences corresponding to the HMG domain of different *Sox* genes were obtained from both male and female O. s*chmackeri*. No sexual difference was found between them. After database searches and phylogenetic analysis,they were found to belong to members of the *SoxB* and *SoxC* subgroups that were named *os-Sox2*, *os-Sox3a*, *os-Sox3b*, *os-Sox4*, *os-Sox14a*, *os-Sox14b*, *os-Sox11a*, *os-Sox11b*, *os-Sox21a* and *os-Sox21b* (*Sox* of *O. schmackeri*, *os-Sox*), individually. These genes have been submitted to GenBank under the accession numbers EU873071, EU873072, EU873073, EU873074, EU873075, EU873076, EU873077, EU873078, EU873079 and EU873080. The predicted amino acid sequences of these genes had between 90% and 98% sequence identity to the corresponding SOX genes in human.

Sequence alignments

The alignments of the nucleotide and putative amino acid sequences of the *O. schmackeri* HMG domains of *Sox* genes are shown in figures 1 and 2, respectively. These ten amino acid sequences were aligned with 39 *Sox* gene sequences from GenBank, including mammalian, reptilian and invertebrates (Fig. 3). From the alignment one can see many highly conserved residues among all the analyzed sequences (about 22 in 69). Sequences in the same subgroup are known to share high similarity and even characteristic sequences in the HMG domain. Zhang *et al.* (2008) suggested that residues at positions 15-19 were characteristic sequences of different subgroups. Similar to that "MAQE(D)N" in group B (except for hu-SOX3, mo-SOX3, ce-SOXB1, dr-SOXB2.1 and dr-SOXB2.2), "IMEQS" in group C (except for ce-SOXC and dr-SOXC) were group specific. However, sequences of "MKE(D)H(Y)" in group B and "MADY" in group C (except for ce-SOXC and dr-SOXC) at position 57-61 seem to be characteristic sequences as well. There were differences in one or two amino acid residues between SOXs of *O. schmackeri* and *Xenopus laevis* (Daudin, 1802). This seems to reflect the genetic

```
                5         15         25         35         45         55
os-Sox2     AGCAACCCAT GAACGCATTT ATCGTATGGT CCAGAGGTCA GAGGAGGAAG ATGGCCCAGG
os-Sox3a    ...G...... ...T..T... .......... ....G..G.. .......... ..........
os-Sox3b    ...G...... ...T..C..C .....G.... ......CC.. .......... ..........
os-Sox14a   ...G...... ......T... .....G.... ....G..G.. .......... ..........
os-Sox14b   ...G...... ......C..C .....G.... ....G..C.. .......... ..........
os-Sox21a   ...G...... ......T... .....G.... ......CC.. .......... ..........
os-Sox21b   ...G...... ......C..C .....G.... ......CC.. .......... ..........
os-Sox4     ...G...... ......T... .....G.... .ACAGATCG. .C........ ..CATGG..C
os-Sox11a   ...G...... .........C .....G.... .T.AGATCG. .C..C..... ..CATGG..C
os-Sox11b   ...G...... ......G... .....G.... .T.AGATCG. .C..C..... ..CATGG..C

                65         75         85         95        105        115
os-Sox2     AGAACCCTAA GATGCATAAC TCTGAGATCA GTAAGAGGCT TGGAGCTGAA TGGAAGCTCC
os-Sox3a    .......C.. ......C... .......... .C........ G..G..C..T ..........
os-Sox3b    .......... ......C... .......... .C........ G..G..C..T ..........
os-Sox14a   .C..T..C.. ......C..T ..G....... ....A..A.. ...G.....G .....A..T.
os-Sox14b   .......C.. ......C..T ..G....... ....A..A.. ...G.....G .....A..T.
os-Sox21a   .......... ......C... ..G....... .C...C.C.. G..T..C..G ......T.A.
os-Sox21b   .......... ......C... ..G....... .C...C.C.. G..T..C..G ......T.A.
os-Sox4     ...G...GG. C.....C..T G........T C....C.... G..GAAGCGC .......AG.
os-Sox11a   .ATC...GG. C.....C... G.C......T CC...C.... G..CAAGCGG ......A.G.
os-Sox11b   .ATC...GG. C.....C.G. G.C......T CC...C.... G..CAGGCGG ......A.G.

               125        135        145        155        165        175
os-Sox2     TCTCCGAGGC CGAGAAGAGA CCCTTCATTG ATGAGGCCAA GAGGCTCCGA GCCCTCCACA
os-Sox3a    .GAG...CT. T.....A... ..T.....A. .C.....T.. ...A..GA.. ..TG......
os-Sox3b    .GAG...CT. T.....A... ..T.....A. .C.....T.. ...A..GA.. ..TG...G..
os-Sox14a   .G..T..A.. ......A... ....A..... .C..A..... A...T.GA.G ..T.AA....
os-Sox14b   .G..T..A.. ......A... ....A..... .C..A..... A...T.GA.G ..T.AA....
os-Sox21a   .GA.....T. G..A...... ..G.....A. ....A..... ...A..G..G ..TA.G....
os-Sox21b   .GA.....T. G..A...... ..G.....A. ....A..... ...A..G..G ..TA.G.G..
os-Sox4     .AAAG..CAG T..T....TC ........CA GG......G. .C....G..C CT.AAG....
os-Sox11a   ..AAT..CAG ..T.....TC ........CC GG..A...G. .C.C.....G CT.AAA....
os-Sox11b   ..AAT..CAG ........TC ........CC GG..A...G. .C.C.....G CT.AAA....

               185        195        205        215
os-Sox2     TGAAGGAACA CCCAAACTAC AAATATCGAC CTCGT
os-Sox3a    ........T. ...C...... .......... .....
os-Sox3b    ........T. ...G...... .......... .....
os-Sox14a   .......... ...CG.T..T ..G....... .....
os-Sox14b   .......... ...CG..... ..G....... .....
os-Sox21a   .......G.. ...GG..... ..G....... .....
os-Sox21b   .......G.. ...CG..... .......... .....
os-Sox4     ..GCC..CT. ...GG..... ..G....... .....
os-Sox11a   ..GC...CT. .......... ..G....... .....
os-Sox11b   ..GC...CT. ...CG..... .......... .....
```

Figure 1. Alignment of nucleotide sequences of the HMG domain of *os-Sox* genes. Nucleotide residues identical to *os-Sox2* are indicated by a dash.

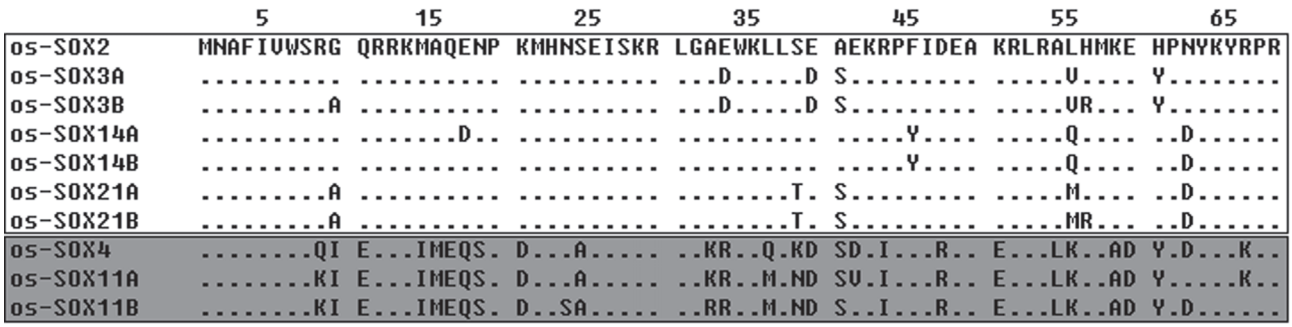

Figure 2. Alignment of the putative amino acid sequence of the HMG domain of os-SOX proteins. (□) SOXB subgroup, (■) SOXC subgroup. Amino acid residues identical to os-SOX2 are indicated by a dash.

basis of adaptive evolution under different environmental conditions. As *X. laevis* is found throughout much of Africa, and in isolated, introduced populations in North America, South America, and Europe (Lau *et al*. 2004), it lives far from *O. schmackeri* which is unique in china. As a result of geographical variation they may have obtained different mutations during the evolutionary process, leading to the differences in gene sequences. Further research is needed to address this issue. Species and the sequence accession numbers of the *Sox* genes used in figure 3 were listed in table I.

Molecular phylogenetic analysis

The sequences in table I were used in the molecular phylogenetic analysis of the *Sox* gene family. ye-MATA1 (P36981) was chosen to serve as the outgroup (Fig. 4).

According to the NJ tree, all sequences used in the phylogenetic analysis segregated into nine groups (A-H and J). SOXE and SOXF clustered together, so did SOXB, SOXG and SOXA. SOXC and SOXD were in monophyletic clades. SOXH consisted of mammalian SOX30s, was distantly related to all the other SOX groups. SOXB group was subdividied into two subgroups (B1 and B2). The human SRY and SOX15 are closely related to SOX2 and SOX3, so there may be some evolutionary relationship among them. It is likely that SOXB, SOXC, SOXD, SOXE SOXF and SOXJ are ancient, because they all contain invertebrate sequences. However, SOXA, SOXG and SOXH might have evolved recently.

In the phylogenetic tree (Fig. 4), the ten *Sox* genes isolated in our research gathered with their human orthologues in group B or C. So the fine topology of the tree supports the *Sox* gene name and group assignments of these *Sox* genes. Moreover, the clustering of invertebrate *Sox* genes can further confirm the results.

Every SOX group like SOXB has many members in mammals, except for group A, G and H (Bowles *et al*. 2000). However, most of the SOX groups are represented by a single SOX sequence in invertebrates. For instance, in *C.elegans* and *Drosophila*, the group C and D are each represented by a single gene, whereas in *Drosophila* the groups E and F have one member each (Fig. 4).

Table I. Sox gene sequences used in this paper.

Sequence	Accession number	Sequence	Accession number
A		D	
hu-SRY	AAT37462	ce-SOXD	AF097319
B		mo-SOX5	BAA32567
ce-SOXB1	U38377	mo-SOX6	CAA09270
xe-SOX1	BAE72677	hu-SOX5	CAG32994
xe-SOX2	AF005476	hu-SOX6	AAK26243
mo-SOX3	AAH52024	E	
ch-SOX2	NP_990519	dr-SOXE	AJ251580
ch-SOX3	BAA77266	hu-SOX8	NM_014587
hu-SOX1	NP_005977	hu-SOX9	CAA86598
hu-SOX2	CAA83435	F	
hu-SOX3	CAA50465	dr-SOXF	AJ250955
dr-SOXB2.1	AC015146	hu-SOX7	NP_113627
dr-SOXB2.2	AC015146	hu-SOX17	NM_022454
xe-SOX14	ABY90181	hu-SOX18	AB033888
xe-SOX21	ABY90180	G	
hu-SOX14	AAI06731	hu-SOX15	NP_008873
hu-SOX21	NP_009015	H	
C		mo-SOX30	AAF99391
ce-SOXC	U80032	hu-SOX30	NP_848511
dr-SOXC	AJ252125	J	
zf-SOX4A	BC065354	ce-SOXJ	U51998
mo-SOX4	NP_033264	Outgroup	
hu-SOX4	NP_003098	ye-MATA1	P36981
hu-SOX11	BAA88122		

This expansion of gene family during evolution suggests that there is a single ancestral gene of each group which gave rise to the multiple genes of the vertebrate lineage, by rounds of duplication (Koopman *et al*. 2004). More especially, in the formation

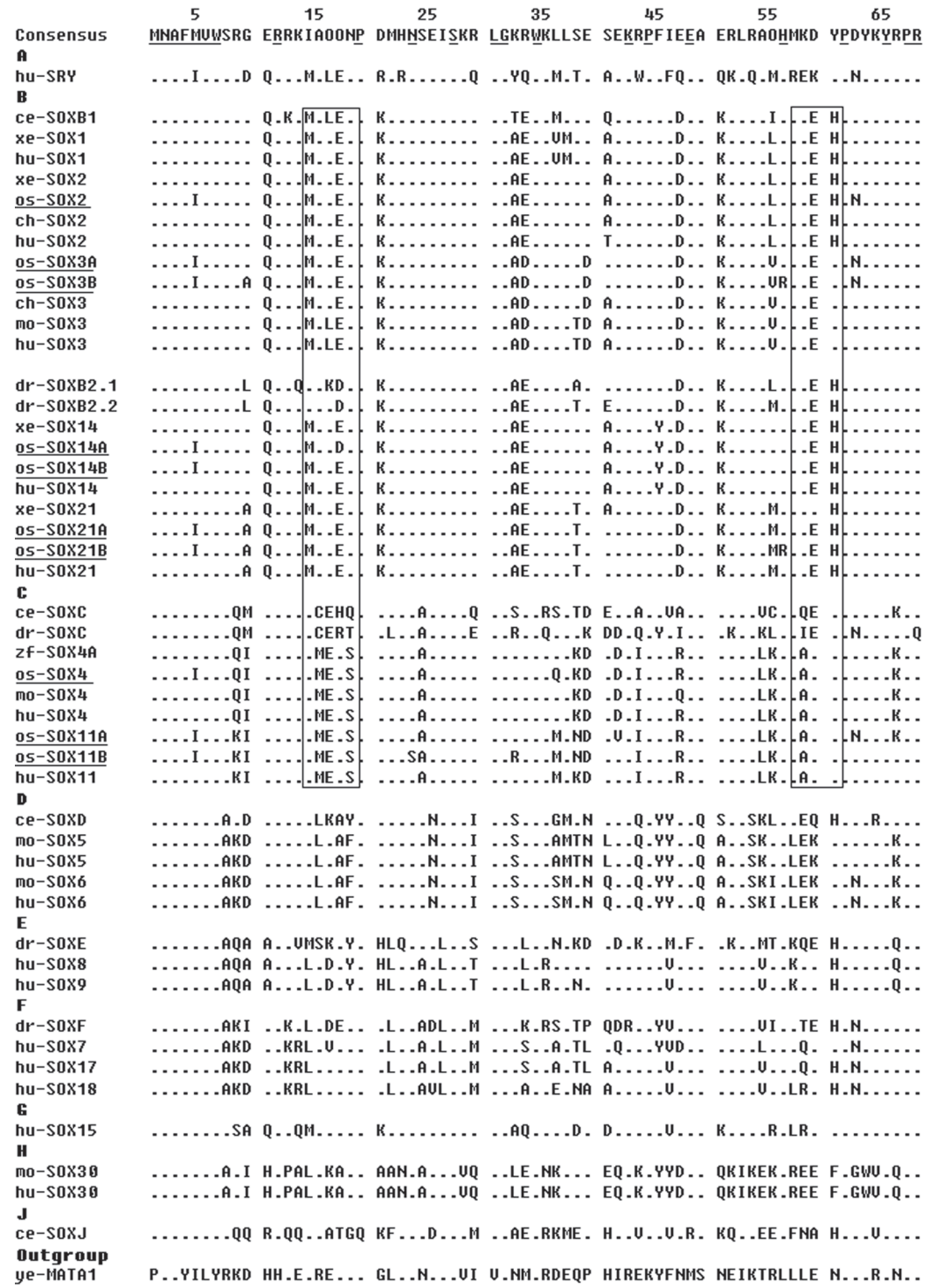

Figure 3. Alignment of the HMG domain of the *Sox*/SOX genes at amino acid level. Sequences are arranged into groups as defined by Bowles *et al.* (2000). The consensus sequence was cited from Bowles *et al.* (2000) too. Amino acid residues identical to the consensus sequence are indicated by a dash. Residues highly conserved are lined under the consensus sequence. The SOXs obtained from *O. schmackeri* are lined below. The characteristic sequences of group B and C are boxed.

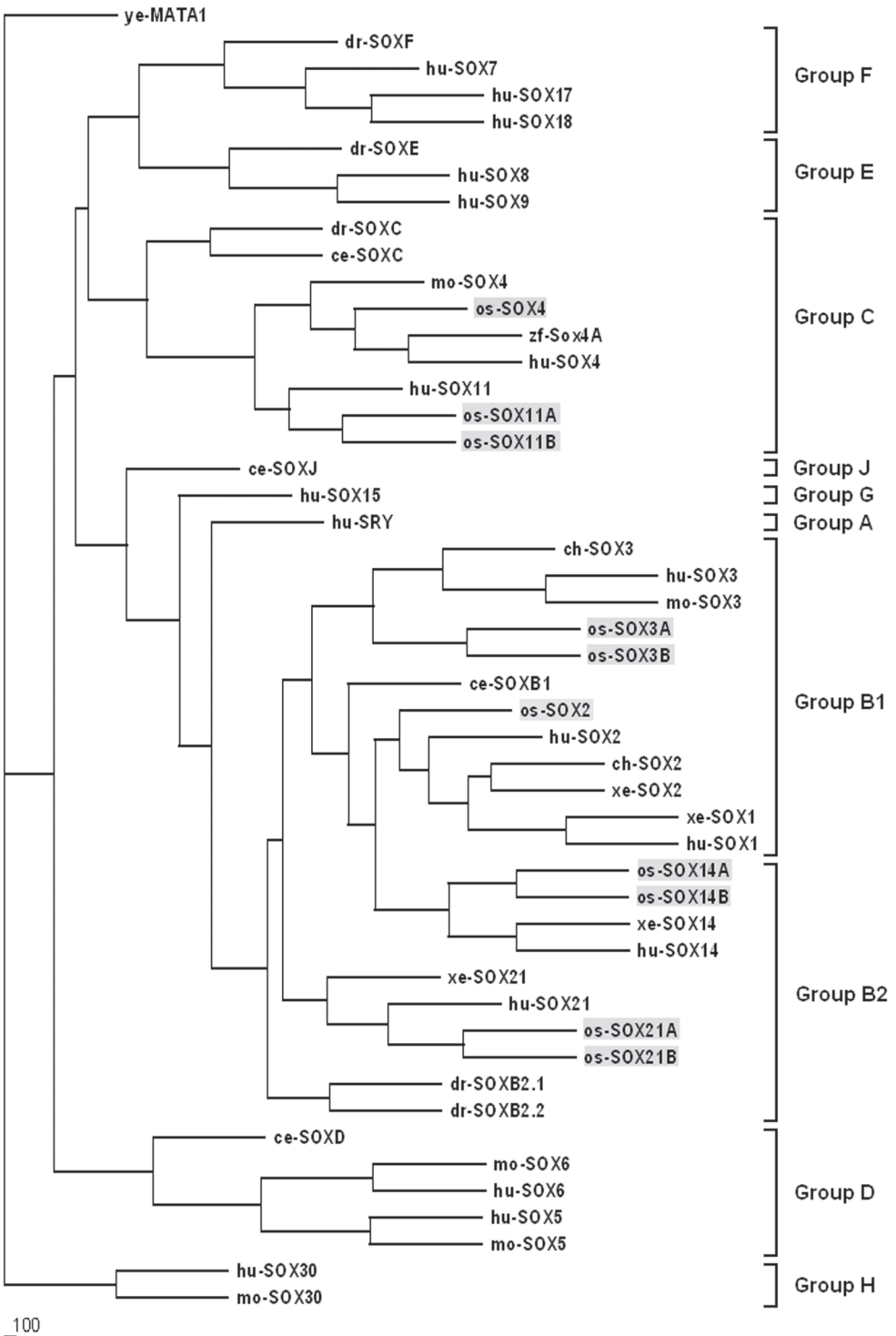

Figure 4. Phylogenetic neighbor-joining tree of *Sox* gene family. An alignment of the HMG domain sequences was made with Clustal X; this was used to derive a phylogenetic tree with PHYLIP software by the neighbor-joining method and the output tree was displayed by TreeView (v1.6.6) without any adjustment. Bootstrapping was carried out on 1000 replicates. Based on the tree and pervious data, genes were ascribed to groups A, B, C, D, E, F, G, H, or J. The *Sox* genes belonging to *O. schmackeri* are shaded.

of vertebrate *SoxB* genes, lineage-specific duplication and diversification were involved as there are one more *Sox* genes in *Drosophila*. The model for the evolution of *SoxB* genes in vertebrate was clearly pictured in Mckimmie *et al.* (2005). As *SoxG*, -A and -H, have only one member (*Sox15*, *Sry* and *Sox30*, respectively) and are restricted to mammals, they can be thought to have arisen recently. It has been suggested that mammalian *Sry* evolve from *Sox3*, its ancestor gene, located on the X chromosome (Foster & Graves 1994). But the origins of *SoxG* and *SoxH* are less clear. However, as function of *Sox15* is related to *Sox2* in some regulation processes in the ES cells (mentioned in the introduction) and *Sox15* is closely related to *SoxB* genes (revealed by the phylogenetic tree) it can be presumed that the origin of *Sox15* is associated with the duplication and variation of *Sox2* during mammalian evolution.

Based on sequence analysis and functional studies, vertebrate *SoxB* have been subdivided into two further groups: B1 (including *Sox1*, *Sox2, Sox3*)and B2 (including *Sox14*, *Sox21*). Although members of group B1 take on additional unique roles, they are all involved in CNS development and regulation of the neuronal phenotype (Collignon *et al.* 1996). They are also coexpressed during lens development, showing an overlapping expression pattern. Similarly, group C proteins SOX4, SOX11 and SOX22 show an overlaping expression in the developing central and peripheral nervous systems (Wegner 1999). All the functional redundancy in SOX groups can be an evidence of the duplication-degeneration-complementation (DDC) model developed by Force *et al.* (1999), which suggests that the partitioning of ancestral subfunctions resulted in the preservation of the duplicate genes. On the whole, during the evolutionary process, the *Sox* gene ancestor in each group were duplicated and their functions were shared by the duplicate genes, which were finally preserved and enriched the gene family.

Several of the *Sox* genes isolated from *O.schmackeri* are duplicates. For example, the genomes of human and mice contain single copy of *Sox3*, *Sox11*, *Sox21* and *Sox14*, whereas each of these genes is duplicated in *O. schmackeri*. As they encode different amino acid sequences, we suggest they are not pseudogenes (Galay-Burgos *et al.* 2004). Similar duplications are common in fish, for instance, in which there are two distinct versions of *Sox9* and *Sox11* in zebrafish (de Martino *et al.* 2000, Chiang *et al.* 2001); two orthologues of *Sox1*, *-4*, *-9*, *-14* in sea bass (Galay-Burgos *et al.* 2004) and two isoforms of *Sox1*, *Sox6*, *Sox8*, *Sox9*, *Sox10*, *Sox14* in Fugu (Koopman *et al.* 2004). An amphibian, *X. laevis*, contains two copies of *xSox17a* and *xSox18* (Hasegawa *et al.* 2002). Intriguingly, several genes except for *Sox* are also doubled in *X. laevis*, such as Estrogen receptors (ER), E-Protein genes, *hairy2* gene and so on (Wu *et al.* 2003, Shain *et al.* 1997, Murato *et al.* 2007). It is thought that gene duplication is a fundanental source of a new gene in the process of evolution (Murato *et al.* 2007). These examples of duplicates can be explained by recent whole-genome duplication in the evolution of tetrapod and teleost lineages. In fish, the 'fish-specific duplication' theory developed by comparative genomics and phylogenetic analyses indicated that a large scale segmental duplication before the radiation of teleosts resulted in the duplicate genes in fishes (Koopman *et al.* 2004). In *X. laevis*, the whole genome was thought to be duplicated by the pseudo-tetra-ploidization in this line of frog that occurred at least 40 million years ago (Hellsten *et al.* 2007). In the case of the *O. schmackeri Sox*, it can be presumed that similar whole-genome duplication may have also occered in *Odorrana* and leaded to the dual copies. The DDC model mentioned above would predict that, these isoforms cooperate to accomplish some functions finished by the single orthologue in mammal species. And this has been confirmed by zebrafish *Sox9* (Chiang *et al.* 2001) and *X. laevis hairy2* gene (Murato *et al.* 2007). Further studies in function of *O. schmackeri Sox* genes are still need to explain the duplicate genes and their evolution.

ACKNOWLEDGEMENTS

This research was supported by National Natural Science Foundation of China (No. 30770296), the Natural and Science Key Project of Anhui Educational Department (KJ2007A022), and the Key Laboratory of Biotic Environment and Ecology Safety in Anhui Province (2008).

LITERATURE CITED

Barrionuevo, F.; S. Bagheri-Fam; J. Klattig; R. Kist; M.M. Taketo; C. Englert & G. Scherer. 2006. Homozygous inactivation of *Sox9* causes complete XY sex reversal in mice. **Biology of Reproduction 74** (1): 195-201.

Béranger, F.; C. Méjean C.; B. Moniot B.; P. Berta & M. Vandromme. 2000. Muscle differentiation is antagonized by SOX15, a new member of the SOX protein family. **The Journal of Biological Chemistry 275** (21): 16103-16109.

Bowles, J.; G. Schepers & P. Koopman. 2000. Phylogeny of the SOX Family of Developmental Transcription Factors Based on Sequence and Structural Indicators. **Developmental Biology 227** (2): 239-255.

Collignon, J.; S. Sockanathan; A. Hacker; M. Cohen-Tannoudji; D. Norris; S. Rastan; M. Stevanovic; P.N. Goodfellow & R. Lovell-Badge. 1996. A comparison of the properties of *Sox3* with *Sry* and two related genes, *Sox1* and *Sox2*. **Development 122** (2): 509-520.

Chaboissier, M.C.; A. Kobayashi; V.I. Vidal; S. Lützkendorf; H.J. van de Kant; M. Wegner; D.G. de Rooij; R.R. Behringer & A. Schedl. 2004. Functional analysis of *Sox8* and *Sox9* during sex determination in the mouse. **Development 131** (9): 1891-1901

Chiang, E.F.; C.I. Pai; M. Wyatt; Y.L. Yan; J. Postlethwait & B. Chung. 2001. Two *Sox9* genes on duplicated zebrafish chromosomes: expression of similar transcription activators in distinct sites. **Developmental Biology 231** (1): 149-163.

de Martino, S.; Y.L. Yan; T. Jowett; J.H. Postlethwait; Z.M. Varga; A. Ashworth & C.A. Austin. 2000. Expression of *Sox11* gene

duplicates in zebrafish suggests the reciprocal loss of ancestral gene expression patterns in development. **Developmental Dynamics 217** (3): 279-292.

Force, A.; M. Lynch; F.B. Pickett; A. Amores; Y.L. Yan & J. Postlethwait. 1999. Preservation of dupicate genes by complementary, degenerative mutation. **Genetics 151** (4): 1531-1545.

Foster, J.W. & J.A. Graves. 1994. An SRY-related sequence on the marsupial X chromosome: Implications for the evolution of the mammalian testis-determining gene. **The Proceedings of the National Academy of Sciences 91** (5): 1927-1931.

Galay-Burgos, M.; L. Llewellyn; C.C. Mylonas; A.V. Canario; S. Zanuy & G.E. Sweeney. 2004. Analysis of the *Sox* gene family in the European sea bass. **Comparative Biochemistry and Physiology Part B 137** (2): 279-284.

Hanna, J.; S. Markoulaki; P. Schorderet; B.W. Carey; C. Beard; M. Wernig; M.P. Creyghton; E.J. Steine; J.P. Cassady; R. Foreman; C.J. Lengner; J.A. Dausman & R. Jaenisch. 2008. Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. **Cell 133** (2): 250-264.

Hasegawa, M.; Y. Hiraoka; J. Hagiuda; M. Ogawa & S. Aiso. 2002. Expression and characterization of *Xenopus laevis SRY*-related cDNAs, *xSox17a1*, *xSox17a2*, *xSox18a* and *xSox18b*. **Gene 290** (1-2): 163-172.

Hellsten, U.; M.K. Khokha; T.C. Grammer.; R.M. Harland; P. Richardson.& D.S. Rokhsar. 2007. Accelerated gene evolution and subfunctionalization in the pseudotetraploid frog *Xenopus laevis*. **BMC Biology 5**: 31.

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

Lau, M.W.N.; G. Baorng; G. Huiqing; Y. Zhigang & L. Kuangyang. 2004. *Rana schmackeri*. **2006 IUCN Red list of threatened species.** Available online at: http://www.iucnredlist.org [Accessed: 2004].

Laudet, V.; D. Stehelin & H. Clevers. 1993. Ancestry and diversity of the HMG box superfamily. **Nucleic Acids Research 21** (10): 2493-2501.

Lefebvre, V.; B. Dumitriu; A. Penzo-Méndez; Y. Han & B. Pallavi. 2007. Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. **The International Journal of Biochemistry & Cell Biology 39** (12): 2195-2214.

Li, J.; G. Pan; K. Cui; Y. Liu; S. Xu & D. Pei. 2007. A dominant-negative form of mouse *Sox2* induces trophectoderm differentiation and progressive polyploidy in mouse embryonic stem cells. **The Journal of Biological Chemistry 282** (27): 19481-19492.

Mannaert A.; K. Roelants ; F. Bossuyt & L. Leyns. 2006. A PCR survey for posterior *Hox* genes in amphibians. **Molecular Phylogenetics and Evolution 38** (2): 449-458.

Maruyama, M.; T. Ichisaka; M. Nakagawa & S. Yamanaka. 2005. Differential Roles for *Sox15* and *Sox2* in transcriptional control in mouse embryonic stem cells. **The Journal of Biological Chemistry 280** (26): 24371-24379.

Mckimmie, C.; G. Woerfel. & S. Russell. 2005. Conserved genomic organisation of Group B *Sox* genes in insects. **BMC Genetics 1186** (10): 1471-2156.

Meissner, A.; M. Wernig & R. Jaenisch. 2007. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. **Nature Biotechnology 25** (10): 1177-1178.

Mertin, S.; S.G. McDowall & V.R. Harley. 1999. The DNA-binding specificity of SOX9 and other SOX proteins. **Nucleic Acids Research 27** (5): 1359-1364.

Murato Y.; K. Nagatomo; M.Yamaguti & C. Hashimoto. 2007. Two alloalleles of *Xenopus laevis hairy2* gene—evolution of duplicated gene function from a developmental perspective. **Development Genes and Evolution 217** (9): 665-673.

Nie, L.W.; X.N. Shan & C.W. Guo. 1999. The PCR amplification and SSCP analysis of *Sox* gene in turtles (*Platysternon megacephalum* and *Cistoclemmys flavomarginatus*). **Chinese Journal of Applied & Environmental Biology 5**: 378-381.

Okubo, T.; L.H. Pevny & B.L. Hogan. 2006. *Sox2* is required for development of taste bud sensory cells. **Genes & Development 20** (19): 2654-2659.

Park, I.H.; R. Zhao; J.A. West; A. Yabuuchi A.; H. Huo; T.A. Ince; P.H. Lerou; M.W. Lensch & G.Q. Daley. 2008. Reprogramming of human somatic cells to pluripotency with defined factors. **Nature 451** (7175): 141-146.

Sandberg, M.; M. Källström & J. Muhr. 2005. *Sox21* promotes the progression of vertebrate neurogenesis. **Nature Neuroscience 8** (8): 995-1001.

Schilham, M.W.; P. Moerer; A. Cumano & H.C. Clevers. 2007. *Sox4* facilitates thymocyte differentiation. **European Journal of Immunology 27** (5): 1292-1295.

Shain D.H.; T. Neuman & M.X.Zuber.1997.Embryonic expression and evolution of duplicated E-protein genes in *Xenopus laevis*: parallels with ancestral E-protein genes. **Genetics** 146 (1): 345-353.

Shen, G.X.; H.F. Zhu; Y. Zhang; S. Wang; Z.G. Zhu & C. Zhou. 2000. Screening of transfected bacteria by bacteria colonies PCR and plasmid PCR. **Immunological journal 2**: 149-151.

Stros, M.; D. Launholt & K.D. Grasser. 2007. The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. **Cellular and Molecular Life 64** (19-20): 2590-2606.

Suzuki, Y. 2008. Expression of *Sox2* in mouse taste buds and its relation to innervation. **Cell and Tissue Research 332** (3): 393-401.

Wegner, M. 1999. From head to toes: the multiple facets of SOX proteins. **Nucleic Acids Research 27** (6): 1409-1420.

Weiss, J.; J.J. Meeks; L. Hurley; G. Raverot; A. Frassetto & J.L. Jameson. 2003. *Sox3* Is Required for Gonadal Function, but Not Sex Determination, in males and females. **Molecular and Cellular Biology 23** (22): 8084-8091.

Wu K.H.; M.L. Tobias ; J.W. Thornton & D.B. Kelley . 2003. Estrogen

receptors in Xenopus: duplicate genes, splice variants, and tissue-specific expression. **General and Comparative Endocrinology 133** (1): 38-49.

ZHANG, Y; L.W. NIE; L. YAN; P.P. ZHENG & W. CHEN. 2008. Clone and sequence analysis of *Sox* senes in *Rana tientaiensis*. **Asiatic Herpetological Research 11**: 151-158.

Submitted: 28.VII.2008; Accepted: 11.III.2009.
Editorial responsibility: Marcio Roberto Pie