Isolation and sequencing of the HMG domain of ten Sox genes from Odorrrana 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 Odorrrana 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 SoxC 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, BEANGER 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 (Schelham et al. 2007). Odorana 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 TGAAYGCNTTATNG-3'; L2:5'-ACGAGGTCGATAYTTRTARTYN GG-3'). The PCR was carried out in a 25 µl reaction mixture containing 16µl ddH2O, 100 ng of genomic DNA, 1.5 mM Mg2+, 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 with 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. Boosting 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. schmackeri. 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) 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
Table 1. Nucleotide sequences of the HMG domain of os-Sox genes from *Odorrana schmackeri*.

<table>
<thead>
<tr>
<th></th>
<th>5</th>
<th>15</th>
<th>25</th>
<th>35</th>
<th>45</th>
<th>55</th>
</tr>
</thead>
<tbody>
<tr>
<td>os-Sox2</td>
<td>AGCAACCAT</td>
<td>GAACGCATT</td>
<td>ATCGTATGTT</td>
<td>CAGAGGTC</td>
<td>GAGAGGAG</td>
<td>ATGACCAGG</td>
</tr>
<tr>
<td>os-Sox3a</td>
<td>......G......</td>
<td>......T......</td>
<td>......G......</td>
<td>......C......</td>
<td>......G......</td>
<td>......G......</td>
</tr>
<tr>
<td>os-Sox3b</td>
<td>......G......</td>
<td>......T......</td>
<td>......G......</td>
<td>......C......</td>
<td>......G......</td>
<td>......G......</td>
</tr>
<tr>
<td>os-Sox14a</td>
<td>......G......</td>
<td>......T......</td>
<td>......G......</td>
<td>......C......</td>
<td>......G......</td>
<td>......G......</td>
</tr>
<tr>
<td>os-Sox14b</td>
<td>......G......</td>
<td>......T......</td>
<td>......G......</td>
<td>......C......</td>
<td>......G......</td>
<td>......G......</td>
</tr>
<tr>
<td>os-Sox21a</td>
<td>......G......</td>
<td>......T......</td>
<td>......G......</td>
<td>......C......</td>
<td>......G......</td>
<td>......G......</td>
</tr>
<tr>
<td>os-Sox21b</td>
<td>......G......</td>
<td>......T......</td>
<td>......G......</td>
<td>......C......</td>
<td>......G......</td>
<td>......G......</td>
</tr>
<tr>
<td>os-Sox4</td>
<td>......G......</td>
<td>......T......</td>
<td>......G......</td>
<td>......C......</td>
<td>......G......</td>
<td>......G......</td>
</tr>
<tr>
<td>os-Sox11a</td>
<td>......G......</td>
<td>......T......</td>
<td>......G......</td>
<td>......C......</td>
<td>......G......</td>
<td>......G......</td>
</tr>
<tr>
<td>os-Sox11b</td>
<td>......G......</td>
<td>......T......</td>
<td>......G......</td>
<td>......C......</td>
<td>......G......</td>
<td>......G......</td>
</tr>
</tbody>
</table>

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.
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 subdivided 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).

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...
Isolation and sequencing of the HMG domain of ten Sox genes from *Odorrana schmackeri*

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 previous 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 McKimme 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 overlapping 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, they are not pseudogenes (Galav-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 (Galav-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 fundamental 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 occurred in Odorana 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


