Analyses of Conservative Sequences of Five SOX Genes in *Panthera Tigris*[1]

SHAO Hong-wei

Zhejiang Institute of Mechanical & Electrical Engineering

ABSTRACT The SOX genes of *Panthera tigris altaica* Temminck were detected by a pair of degenerate primers, which were special to the SOX gene family within the HMG-box. With the genomic DNA as template, one specific band was observed with the length of 220 bp. Five fragments of DNA were found and their identities to the SRY of human were 75%, 56%, 51%, 67% and 48%, respectively; to the Sry of mouse were 73%, 54%, 57%, 66% and 48%, respectively. The genes with the maximum similarity to these five fragments were H-SOX3, H-SOX4, H-SOX11, H-SOX21 and H-SOX22. So, these five DNA fragments are part of Sox genes of *Panthera tigris altaica*, and were named as PtSox3, PtSox4, PtSox11, PtSox21 and PtSox22 according to the general nomenclature.

Keywords: *Panthera tigris altaica*, PtSox3, PtSox4, PtSox11, PtSox21, PtSox22

In mammals, male sex determination is controlled by the presence of the Y chromosome, which carries a gene encoding the testis-determining factor (TDF) on its short arm. The product of this gene induces the testis-determining pathway, resulting ultimately in a male phenotype. In 1990, the testis-determining factor gene has been isolated and named SRY (sex-determining Region, Y chromosome). SRY was found to contain a conservative domain with similarity to the DNA-binding domain of the chromosomal proteins HMG-1 and HMG-2 (High Mobility Group)[1]. A novel gene family – – SOX (SRY-related box) family was identified based on the conservation of the HMG-box. By convention HMG domains of SOX proteins are at least 50% identical to the HMG domain of SRY. The HMG domains of SOX genes, a DNA binding motif of approximately 79 amino acids, bind to the minor groove of the specific DNA sequences and induce target DNA to bend. At least 30 members have now been identified in birds, reptiles, amphibians, fish, insects and nematodes[2~6]. SOX genes are involved in a diverse range of developmental processes in early embryogenesis. They were found to take part in sex determination, chondrogenesis, haemopoisis, neural development, lens development and so on of early embryogenesis[7].

Tiger (*Panthera tigris*), one of the most endangered species, is protected by international organization and governments. Tigers are distributed in the eastern and southern parts of Asia historically. Only one species but eight subspecies presents are in the world, including *Panthera tigris altaica*, which is distributed in the northeast of china, the far east of Russian and north of Korea. The trace of wandering tigers can be found occasionally in the eastern mountains of Heilongjiang province and Jilin province. *Panthera tigris altaica* had been widely distributed in the mountains of northeast China historically. But now their number is so small that it is classified as rare animal. Ecological and genetical studies have been of wild animals carried out by many institutions. This paper reports the SOX genes of it —— genes involved in developmental processes.

1. MATERIALS AND METHODS

1. Materials

The material was taken from muscle tissue of a female tiger in The People's Park of Xinxiang city,

which was three-month-old, and died by accident. The degenerate primer set was designed within the HMG-box motif according to the sequence conservation of Sox gene family. Primer I was 5' GGG GAA TTC ATG GA(T/C) GC(G/A/T/C/) TT(T/C) AT(G/A/T/C) GT(G/A/T/C)TGG3'; and primer II was 5'GGG AAG CTT (G/A/T/C)GG (G/A/T/C)CG (A/G)TA (C/T)TT (G/A)TA (G/A)T(T/C) (G/A/T/C)GG 3'.

1.2. Preparation of genomic DNA

DNA was extracted from muscle sample. 0.5~1g tissue was sheared quickly and grounded into fine powder in liquid nitrogen. 2 ml STE were added and mixed throughly, then centrifuged at 3000 rpm for 10 min. The supernatant fluid was removed. 2ml STE and 2ml lye's buffer (0.5M EDTA, pH8.0; 0.5% SDS). Proteinase K was added at a final concentration of 150 μg/mL and the solution was incubated overnight at 50℃. This was followed by three phenol-chloroform and chloroform extractions. After precipitation with ethanol, the pellet was rinsed in 70% ethanol, moderately dried and dissolved in TE.

1.3. Degenerate primer PCR

The degenerate PCR was performed in a volume of 20μl with about 200ng genomic DNA, 1 μmol/L of each primer, 200 μmol/L dNTP and 2u of Taq DNA polymerase. Thirty-five amplification cycles were performed with annealing temperature 52℃ for 40 sec and extension at 72℃ for 1 min. The amplification products were electrophoresed on 1.5% agarose gels to check the amplification bands.

1.4. Cloning and nucleotide sequencing

PCR-amplified products were purified and ligated into the T-vector. Sangon Com. was entrusted to sequence the clones with dideoxy-mediated chain-termination method using T7 promoter sequence as general primers. Genbank searching determined the similarity and possible amino acids sequences of the DNA sequences. The cloned DNA fragments were named after the sequence that has the highest similarity with them.

2. RESULTS

```
FI   ATGGATGCCTTTATTGTATGGTCCCGTGGGCAGCGGCGCAAGATGGCCCTGGAGAACCCCAAGATGCAC
FII  ATGGATGCATTCATGGTGTGGTCGCAGATCGAGCGGCGCAAGATCATGGAGCAGTCGCCCGACATGCAC
FIII ATGGATGCATTCATCGTGTGGTCCAAGATCGAGCGCAGGAAGATCATGGAGCAGTCCCCGGACATGCAC
FIV  ATGGACGCGTTTATCGTATGGTCGCGGGCTCAGCGGCGCAAGATGGCCCAGGAAAACCCCAAGATGCAC
FV   ATGAATGCGTTCATGGTATGGTCGCAGCACGAGCGTTGGAAGATCATGGACCAGTGGCCCGAGATGCAC

FI   AACTCCGAGATCAGCAAGCGCTTGGGCGCCGACTGGAAACTGCTGACCGACGCCGAGAAGCGGCCGTTC
FII  AACGCCGAGATCTCCAAGCGGCTGGGCAAACGCTGGAAGCTGCTCAAAGACAGCGACAAGATCCCTTTC
FIII AACGCCGAGATCTCCAAGCGGCTGGGCAAGCGGTGGAAAATGCTGAAGGACAGCGAGAAGATCCCGTTC
FIV  AACTCGGAGATCAGCAAGCGCCTGGGCGCCGAGTGGAAGCTGCTCACCGAGTCGGAGAAGCGGCCGTTC
FV   AACGCCGAGATCTCCAAGCGCCTGGGCCGCCGCTGGCAGCTGCTGCAGGACTCGGAGAAGATCCCGTTT

FI    ATCGACGAGGCCAAGCGACTGCGCGCCGTGCACATGAAAGAGTACCCCAACTACAAATACCGGCCA
FII   ATTCGGGAGGCGGAGCGGCTGCGCCTCAAGCACATGGCTGACTACCCAGACTACAAATATCGCCCA
FIII  ATCCGGGAGGCGGAGCGGCTGCGGCTCAAGCACATGGCCGACTACCCTAATTATAAGTACCGACCC
FIV   ATCGACGAGGCGAAGCGCCTGCGCGCCATGCACATGAAGGAGCACCCAAACTACAAGTACCGCCCA
FV    GTGCGGCCCCCGGAGCGGCTGCGCCTCAAGCACATTAACTGACTACCCCAACTACAAGTACCGGCC
```

Fig.1 The DNA sequence of five Sex gene fragments in *Panthera tigris altaica*

```
FI      MDAFIVWSRGQRRKMALENPKMHNSEISKRLGADWKLLTDAEKRPFIDEAKRLRAVHMKEYPNYKYRP
FII     MDAFMVWSQIERRKIMEQSPDMHNAEISKRLGKRWKLLKDSDKIPFIREAERLRLKHMADYPDYKYRP
FIII    MDAFIVWSKIERRKIMEQSPDMHNAEISKRLGKRWKMLKDSEKIPFIREAERLRLKHMADYPNYKYRP
FIV     MDAFIVWSRAQRRKMAQENPKMHNSEISKRLGAEWKLLTESEKRPFIDEAKRLRAMHMKEHPNYKYRP
FV      MXAFMVWSQHERXKIMDQWXEMHNAEISKRLGRRWQLLQDSEKIPFVRPPERLRLKHDYEHPNYKYRP
```

Fig.2 The possible amino acid sequence of these five gene fragments from similarity search.

Degenerate primer set was designed on the basis of published Sox gene sequences from different species, which was specific to the HMG-box motif. With genomic DNA of *Panthera tigris altaica* as template, one band was observed with a size of 220 bp. To determine the structure feature of the amplification product, the 220 bp band was cloned and sequenced. Five gene fragments with different DNA sequences were obtained. The DNA sequences and possible amino acids sequences of them were shown in **Fig.1** and **Fig.2**. By similarity comparison, the former 100 sequences with high score were Sox genes from those of fruit fly to human. The sequence identities of possible amino acid sequence of these five gene fragments to the SRY/ Sry of human and mouse and the highest identity of possible amino acid sequence to the five novel gene fragments were showed in **Table 1**. Therefore, the cloned fragment was part of one Sox gene.

Table 1 The sequence identities of possible amino acid sequence of these five gene fragments to the SRY/ Sry of human and mouse and the highest identity of possible amino acid sequence to the five novel gene fragments.

Gene	Fragment I	Fragment II	Fragment III	Fragment IV	Fragment V
H-SRY	75%	56%	51%	67%	48%
M-Sry	73%	54%	57%	66%	48%
With the gene	H-SOX 3	H-SOX 4	H-SOX 11	H-SOX 21	H-SOX 22
The maximum similarity	95%	98%	94%	96%	90%
Name	PtSox 3	PtSox 4	PtSox 11	PtSox 21	PtSox 22

By convention, HMG domains of SOX proteins are at least 50% identical to the HMG domain of SRY/Sry and are named according to the maximum similarity to the corresponding genes. So, the five fragments were named as PtSox3, PtSox4, PtSox11, PtSox21 and PtSox22. The sequence similarities among the five Sox gene fragments were showed in **Table 2**.

Table 2 The sequence similarities between among the five Sox gene fragments in *Panthera tigris altaica*

Gene	PtSox 3	PtSox 4	PtSox 11	PtSox 21	PtSox 22
PtSox 3		64.5%	69%	89.7%	56%
PtSox 4			92.6%	63%	72%
PtSox 11				66%	70.5%
PtSox 21					57%

By convention, HMG domains of SOX proteins are at least 50% identical to the HMG domain of SRY/Sry and are named according to the maximum similarity to the corresponding genes. So, the five fragments were named as PtSox3, PtSox4, PtSox11, PtSox21 and PtSox22. The sequence similarities among the five Sox gene fragments were showed in **Table 2**.

3. DISCUSSION

SOX gene family was identified on the basis of sex-determining gene-SRY in mammals. The main characteristic of this family is the presence of the conservative HMG domain. The genes having HMG domain are composed of two subfamilies – – representatives of the HMGI group

containing multiple HMG domains which are less sequence-specific in their binding, while members of the SOX group have single sequence-specific HMG domain with the SRY as representative,and regulates the expression of the target genes. By convention, SOX proteins are more than 50% identical to SRY in the HMG domain. This definition is artificial and arbitrary, not suitable to all cases. The fifth fragment of the tiger SOX gene does not conformed to this rule in this report. Similar instances have been found in a few of the SOX genes in invertebrate and mammals[7]. For example, the H-SOX30 is 48% identical to H-SRY in the HMG domain. Therefore, this definition is now inaccurate. Bowles's results provided an alternative criterion to define the SOX genes using the conservation of key motifs within the domain. The HMG domain sequence RPMNAFMVW (positions 5~13) appears to be conserved for all SOX sequences[8]. The reason that the SOX genes to SRY/Sry is low may be: first, the SRY/Sry gene is on the Y chromosome which evolves more rapidly than other chromosomes for mammals. Second, for other animals there is no SRY/Sry gene of their own but only the SRY/Sry of human and mouse to compare with. So, the similarity is low for them.

The SOX genes are dispersed throughout the mouse and human genomes[7,9], arguing against a purely tandem duplication model of SOX family expansion. One theory is that the family has arisen from a common ancestor *via* ancient duplication, dispersion, mutation, and acquisition mechanisms. We presume that at various times throughout metazoan evolution, HMG box-containing sequences duplicated, in each case leaving one redundant copy which was then free to evolve a new function or else be lost from the genome[10,11]. Rather than relying on slow genetic drift, it is likely that spare HMG-containing fragments recruited preexisting functional domains and hence formed mosaic proteins capable of rapidly taking on novel functions[12]. Since the HMG domain is a DNA binding domain with a recognition site that is largely conserved throughout the family, it is likely to have diverged by gradual drift. Hence, the HMG domain may be considered an independent evolutionary unit, and HMG domain variation will be an accurate marker of the pattern of evolution of the family, provided that sufficient information is present and that back mutations have not obscured its history.

Between 15 and 20 different SOX proteins have already been identified in both mouse and human. If partially cloned SOX proteins from other species are taken into account, it has to be assumed that the number of SOX proteins in any given vertebrate species will be >20. It can not be surprising that most tissue and cell types express a SOX protein during at least one stage of their development. In fact, a couple of studies have shown that a number of tissues or cell types express more than one SOX protein at certain times. So, identifying more SOX genes and extensive studies on their structures and functions is significant to reveal the mechanism of early embryogenesis.

REFERENCES

1. Gubbay J, Collignon J, Koopman P, *et al*.: 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, 1990.
2. Wright EM, Spoke, and P Koopman: Seven new members of the SOX gene family expressed during mouse development. *Nucleic Acids Research*, **21**(3): 744, 1993.
3. Argentaro AJ, Olsson R, Critcher SG, *et al*.: Genomic characterization and fine mapping of the human SOX13 gene. *Gene*, **250** (1~2): 181~189, 2000.
4. Chang ZJ, Zhou R, Yu QX: The Conservative Region Sequence Analysis of PdSOX9 in ParamisgurnusDabryanus. *Acta Genetica Sinica* , **27** (2):121~126, 2000.
5. Chang ZJ, Zhou R, Yu QX: Detection of SOX genes in two kinds of loaches. *Developmental and reproductive biology*, **9** (2):11~17, 1999.

6. Chang ZJ, Zhou R, Yu QX: Progress in SOX gene family. *Hereditas* (Beijing), **22**(1): 51~53, 2000.
7. Wegner, M: From head to toes: The multiple facets of Sox proteins. *Nucleic Acids Res*., **27**(6):1409~1420, 1999.
8. Bowles J, Schepers G, Koopman P: Phylogeny of the SOX family of developmental transcription factors based on sequence and structural Indicators. *Dev. Biol.*, **227**(2): 239~255, 2000.
9. Holland PWH, Garcia-Fernandez J, Williams N A, *et al*.: Gene duplications and the origins of vertebrate development. *Development*, Suppl: 125~133, 1994.
10. Force A, Lynch M, Pickett FB, *et al*.: Preservation of duplicate genes by complementary degenerative mutations. *Genetics*, **151**: 1531~1545, 1999.
11. Ohno S: The one-to-four rule and paralogues of sex-determining genes. *EXS*, (91):1~10, 2001.