

# 一种四倍体鲤科鱼 *Sox* 基因的克隆和序列分析

郭宝成<sup>\*</sup>, 李俊兵<sup>\*</sup>, 童超波<sup>\*</sup>, 何舜平<sup>\*</sup>

中国科学院水生生物研究所, 武汉 430072;

中国科学院研究生院, 北京 100049

<sup>\*</sup> 联系人, E-mail: [clad@ihb.ac.cn](mailto:clad@ihb.ac.cn)

2008-01-28 收稿, 2008-04-12 接受

国家自然科学基金重点资助项目(批准号: 30530120)

**摘要** 为了研究重要功能基因在多倍体基因组中的演化情况, 通过分子克隆的方法从四倍体大鳞结鱼(*Tor douronensis*)中获得 13 条 *Sox* 基因序列, 分别为 *SoxB*, *SoxC* 和 *SoxE* 组成员. 对大鳞结鱼 *Sox* 基因系统进化分析的结果支持鱼类特异的基因组重复的假说, 同时表明, 基因 *Sox19* 可能为真骨鱼特有的 *Sox* 基因家族成员. 序列分析显示, 与其他物种相比大鳞结鱼中 *Sox* 基因的碱基替代, 绝大多数为同义替代. 结果表明, 所得到的 *Sox* 基因序列均受到净化选择, 同时由多倍化所造成的重复基因对 *Sox4a*, *Sox9a* 和 *Sox9b* 受到相同的净化选择压力. 根据基因 *Sox9a* 和 *Sox9b* 重复基因对间内含子序列长度和相似性的差异推测大鳞结鱼可能为异源四倍体.

**关键词**  
大鳞结鱼  
*Sox* 基因  
四倍体  
基因组重复  
基因演化

*Sox* 基因是一类与动物发育相关的重要转录因子, 其序列特征是含有一个高度保守的由 79 个氨基酸残基组成的 DNA 结合域, 即 HMG 结构域. *Sox* 基因存在于整个动物界中, 由 Gubbay 等人<sup>[1]</sup>于 1990 年最先在研究哺乳动物性别决定因子 *SRY* 基因时发现. Bowles 等人<sup>[2]</sup>根据 *Sox* 基因家族成员的序列相似性、结构特征及功能, 在染色体上的位置系统地分析了 *Sox* 基因家族在后生动物中的进化, 并认为 *Sox* 基因家族成员分属于 *SoxA*~*J* 10 个组别. Schepers 等人<sup>[3]</sup>对小鼠和人的 *Sox* 基因进行了系统的命名和进化关系分析, 认为小鼠与人中分别存在 20 个直系同源的 *Sox* 基因成员, 代表了 *SoxA*~*H* 8 个不同组别. 有趣的是, 对斑马鱼(*Danio rerio*)<sup>[4]</sup>和红鳍东方鲀(*Takifugu rubripes*)<sup>[5]</sup>等真骨鱼 *Sox* 基因的研究发现, 很多哺乳动物的 *Sox* 基因在真骨鱼中有两个直系同源基因, 并认为这一现象是真骨鱼类特异的基因组重复所导致的结果.

由多倍化所导致的全基因组重复被认为是产生大量遗传物质的最佳和最快捷的途径, 并在脊椎动物的进化过程中起到了非常重要的作用<sup>[6-8]</sup>. 因此, Ohno<sup>[6]</sup>提出了脊椎动物进化过程中至少经历过两次

全基因组重复的假设, 即“2R”假说(two-round genome duplication), 一次发生在所有的脊椎动物祖先分化之前, 一次发生在有颌类和无颌类分化之后<sup>[9-11]</sup>. 而近年来, 大量的比较基因组研究表明, 真骨鱼类经历过一次特有的全基因组复制(fish-specific genome duplication, FSGD), 并且可能在整个真骨鱼类群成功的适应性辐射演化过程中起到了重要作用<sup>[12-16]</sup>. 有趣的是, 多倍体依然广泛存在于现生鱼类的不同类群中, 如鲟形目、鲑形目和鲤形目中的很多物种为多倍体, 可能是鱼类进化的重要特征<sup>[17]</sup>. 一些原始种类, 如匙吻鲟(*Polyodon spathula*)、肺鱼(*Protopterus dolloi*)和斑点雀鲷(*Lepisosteus oculatus*)为四倍体, 现存的海水鲑科鱼类在大约 25~100 百万年前发生过同源四倍化(autotetraploidization)<sup>[6,18]</sup>. 而在鲤形目的鲤在科、在在科和在在鱼在科中存在大量不同倍性的物种, 如鲫鱼(*Carassius auratus*)为四倍体( $N = 4, 2n = 100$ ), 齐口在在鱼(*Schizothorax prenanti*)为六倍体( $N = 6, 2n = 148$ ). 同时已有的研究表明, 整个金整在属鱼类全整为四倍体并整整金整在属为四倍体起源<sup>[19]</sup>. 因此多倍化事件被认为在鲤在科、在在科和在在鱼在科的整个在科或者在科中某些

属级水平上不断重复发生<sup>[20,21]</sup>。

多倍化最直接的影响就是造成基因组中大量功能基因的重复,而这些重复基因的分子演化机制已经成为分子进化领域研究的热点之一。但目前对多倍化后重复基因组的分子演化机制研究主要集中在酵母(*Saccharomyces cerevisiae*)<sup>[22]</sup>、拟南芥(*Arabidopsis thaliana*)<sup>[23,24]</sup>、小麦<sup>[25]</sup>以及经历过古老多倍化的真骨鱼中<sup>[26]</sup>,而对年轻多倍化脊椎动物基因组在分子水平上演化的研究还非常有限。例如,在多倍化发生后的早期,发生重复的重要功能基因在功能上是完全冗余的,那么自然选择对这些重复基因的选择压力是否完全消失,重复基因进化速率是否相同,在年轻多倍体脊椎动物中都还有待于进一步的研究。为了研究上述问题,我们克隆和分析了一种四倍体鲤科鱼类大鳞结鱼(*Tor douronensis*)的 *Sox* 基因。细胞生物学研究表明,大鳞结鱼为进化上的四倍体( $N=4$ ),染色体数目为  $2n=100$ <sup>[20]</sup>。同时本实验室最新的研究<sup>[27]</sup>表明,大鳞结鱼为新近形成的四倍体,其物种形成时间大约在2百万年前左右,使其成为研究年轻多倍体基因(组)在分子水平演化的理想研究对象。因此通过克隆和分析四倍体大鳞结鱼 *Sox* 基因,本研究的主要目的旨在分析 *Sox* 基因在四倍体大鳞结鱼演化情况,验证 *Sox* 是否支持鱼类特异基因组重复的假说,探讨大鳞结鱼是同源多倍化还是异源多倍化起源。

## 1 材料与方法

( ) 实验材料。本研究中所采用的实验样品(雌性)采集于云南勐腊。标本的鉴定依据于陈宜瑜等人<sup>[28]</sup>1998 的分类系统。实验所用肌肉样品用 95%酒精固定,标本保存于中国科学院水生生物研究所淡水鱼类博物馆,标本号为 IHBCY0405871。

( ) DNA 的提取以及基因的扩增、克隆和整序。从 95%酒精固定的肌肉样品中提取基因组 DNA,采用标准的酚/氯仿抽提法<sup>[29]</sup>。分别采用两对位于 *Sox* 基因 HMG 结构域两端的简并引物 *SoxN* 和 *Sox9*<sup>[30]</sup> 用于 *Sox* 基因的扩增。引物 *SoxN* 的序列为 5'-ATG AAY GCN TTY ATG GTN TGG-3'和 5'-GGN CGR TAY TTR TAR TCN GG-3',引物 *Sox9* 的序列为 5'-ATG AAY GCS TTY ATG GTI TGG-3'和 5'-GTC IGG GTG RTC YTT CTT RTG YTG-3'。PCR 反应的总体积为 50  $\mu$ L,含有 1 $\times$ 缓冲液(Invitrogen), 0.25

mmol/L dNTPs (Promega), 2.5 U *Taq* DNA 聚合酶(Invitrogen), 引物各 0.2 pmol/L, 40~50 ng 基因组 DNA, 最后补足灭菌双蒸水至终体积。PCR 反应条件为: 94 预变性 3 min; 94 变性 30 s, 52 退火 60 s, 72 延伸 1 min, 35 个循环; 72 终延伸 5 min。每次反应设立不含 DNA 模板的空白对照组。扩增产物经 1%的琼脂糖凝胶电泳后,用 AxyPrep DNA 凝胶回收试剂盒(Axygen)进行回收。回收产物用 TA cloning 试剂盒(Invitrogen)进行连接、转化与克隆。对所得到的克隆进行整序,为保证所整序列的可靠性,每个克隆分别进行正反两向整序核对。

( ) 数据分析。

(1) *Sox* 基因的鉴定。用 Clustal X<sup>[31]</sup>软件对整序所得到的同一克隆的正反向序列进行排序,并结合整序所得的峰图进行核对保证序列的正确性。在 GenBank 数据库上进行 BLASTN 和 TBLASTX 搜索,对所得序列进行初步鉴定,并根据已有的 *Sox* 基因序列识别在 HMG 保守区有内含子的大鳞结鱼 *Sox* 基因的结构。将所得到的大鳞结鱼 *Sox* 基因核苷酸序列在 MEGA 4.0<sup>[32]</sup>中转换为其相应的假定氨基酸序列,并根据 Bowles 等人<sup>[2]</sup>的研究结果,以不同 *Sox* 基因所具有的特征氨基酸对所得到的大鳞结鱼的 *Sox* 基因进行进一步的鉴定。

(2) 系统进化分析。分别采用最小进化法(minimum evolution, ME)和贝叶斯法(Bayesian inference)对所得到的大鳞结鱼 *Sox* 基因进行系统进化关系分析。用所得到的大鳞结鱼 *Sox* 基因的假定氨基酸序列在 MEGA 4.0 中进行 ME 分析,模型为 Equal Input。对所得到的大鳞结鱼 *Sox* 基因的核苷酸序列运用 MrBayes 3.1.2<sup>[33]</sup>进行贝叶斯分析(Bayesian Inference, BI),替代模型的选择采用 Modeltest 3.7<sup>[34]</sup>程序进行分析,按 hLRTs 标准,最适合的替代模型为 TrN + I + G。马可夫蒙特卡洛(MCMC)参数如下: starting tree = random, number of chains = 4, number of generations = 4000000, sample frequency = 100, temperature = 0.5, burn-in = 20000。在系统进化分析中,分别采用已知的人(*Homo sapiens*), 斑马鱼和红鳍东方鲀的 *Sox* 基因序列与所得到的大鳞结鱼 *Sox* 基因序列一起进行分析,并以人的 *TCF7* 基因(GenBank 登录号为: NM\_201632.1)为外类群。

(3) 分子进化分析。序列的相似性分析(包括核苷酸水平与氨基酸水平)在 MatGET v2.02<sup>[35]</sup>(BLOSUM50

scoring matrix)中进行. 在 MEGA 4.0 中对所得到的整分大鳞结鱼 *Sox* 基因序列的同义替代率(synonymous substitution rates, *Ks*)和非同义替代率(nonsynonymous substitution rates, *Ka*)以改进的 Nei-Gojobori<sup>[36]</sup>法为模型进行计算, 对所分析的 *Sox* 基因对间的碱基替代数(*d*)的计算则采用 Kimura 双参数法. *Z* 检验同样在 MEGA 4.0 中进行, 以检验大鳞结鱼中由多倍化导致的具有两个拷贝的 *Sox* 基因进化中所受到的选择压力.

## 2 结果

### 2.1 大鳞结鱼 *Sox* 基因的鉴定和系统演化关系分析

使用两对简并引物 *SoxN* 和 *Sox9* 进行 PCR 扩增, 将所得到的 PCR 产物用于克隆, 并对所得到的 36 个克隆进行了整序分析. 对所得到的大鳞结鱼 *Sox* 基因

的核苷酸序列(不含内含子序列)及其假定的氨基酸序列进行排序分析(图 1). 将所得到的大鳞结鱼 *Sox* 基因的核苷酸序列转换为其假定的氨基酸序列后, 根据 Bowles 等人<sup>[2]</sup>2000 年的研究结果, 结合对 *Sox* 基因所含有的特征氨基酸的分析以及 BLAST 搜索的结果, 发现共得到的 13 条不同的大鳞结鱼 *Sox* 基因, 且分别属于不同的 *Sox* 基因组别, 所得到的详细信息见表 1.

氨基酸序列的 ME 分析(图 2(a))显示, 所得到的大鳞结鱼 *Sox* 基因分别属于不同 *Sox* 基因组别 B、C 和 E<sup>[2]</sup>. 用引物 *SoxN* 进行 PCR 扩增所得到的 *Sox* 基因属于 *SoxB* 和 *SoxC*, 并且所得到的 *Sox* 基因在其 HMG 结构域没有内含子序列, 与 Galay-Burgos 等人<sup>[30]</sup>的研究结果相同. 所得到的 *Sox* 基因相对应的人的 *Sox* 基因分别为 *Sox4*, *Sox11*, *Sox21*, 其中一条 *Sox*

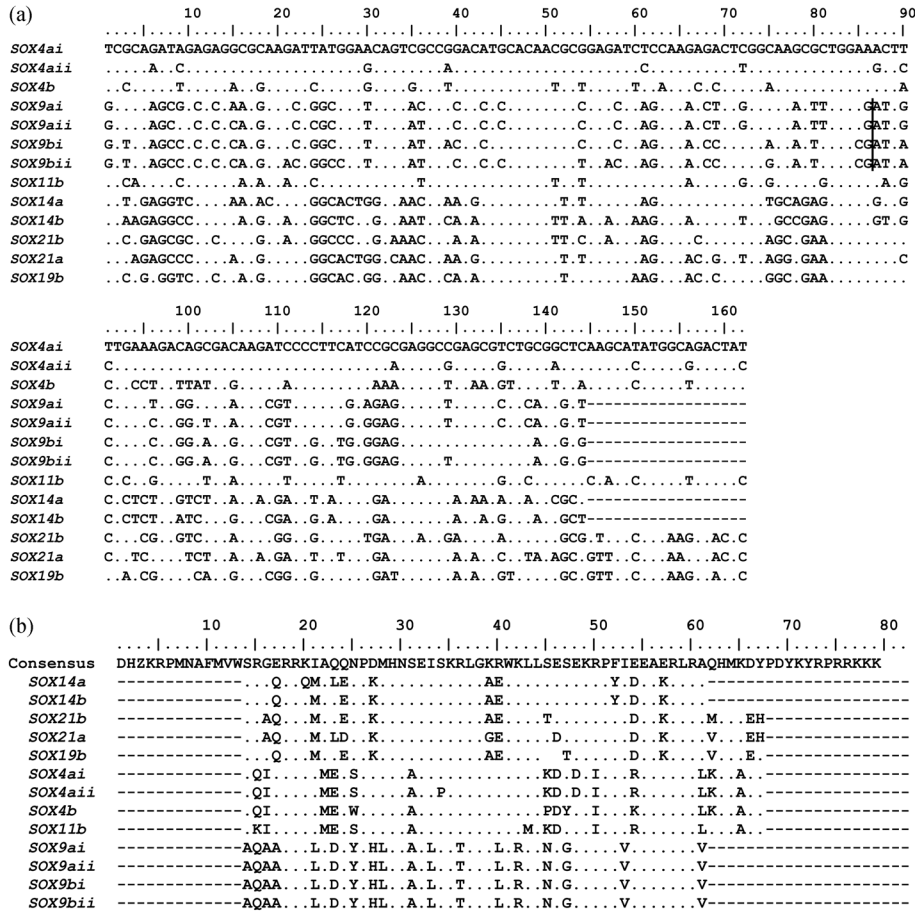


图 1 大鳞结鱼 *Sox* 基因 HMG 结构域核苷酸(a)和氨基酸序列(b)的排列  
圆点(·)表示一致性的核苷酸或氨基酸残基. (a) 中(|)表示大鳞结鱼 *Sox* 基因内含子所在位置

表1 *Tor douronensis* Sox 基因及相关信息(以及其与斑马鱼直系同源基因)

| 大鳞结鱼 Sox 基因                 |     |                       |                |             | 组别 | 斑马鱼 Sox 基因名称              |
|-----------------------------|-----|-----------------------|----------------|-------------|----|---------------------------|
| 克隆 ID                       | 拷贝数 | 长度/bp                 | 基因名称           | GenBank 登录号 |    |                           |
| TDN15, TDN <sup>a</sup> 20  | 2   | 162                   | <i>Sox4ai</i>  | EU399799    | C  | <i>Sox4a</i>              |
| TDN16, TDN24                | 2   | 162                   | <i>Sox4aai</i> | EU399800    |    | <i>Sox4b</i>              |
| TDN12, TDN14, TDN18         | 3   | 162                   | <i>Sox4b</i>   | EU399801    |    | <i>Sox11b</i>             |
| TDN25, TD9 <sup>b</sup> 112 | 2   | 162                   | <i>Sox11b</i>  | EU399802    |    | <i>Sox9a</i>              |
| TD924~TD928, TD936          | 6   | 335(191) <sup>c</sup> | <i>Sox9ai</i>  | EU399808    | E  | <i>Sox9a</i>              |
| TD941~TD944                 | 4   | 627(483)              | <i>Sox9aai</i> | EU399809    |    | <i>Sox9b</i>              |
| TD9518~TD920                | 3   | 805(661)              | <i>Sox9bi</i>  | EU399810    |    | <i>Sox9b</i>              |
| TD953, TD958                | 2   | 823(679)              | <i>Sox9bii</i> | EU399811    |    |                           |
| TDN5, TDN10, TD9110         | 3   | 144                   | <i>Sox14a</i>  | EU399803    |    | <i>Sox14a<sup>d</sup></i> |
| TD9113, TD9117              | 2   | 144                   | <i>Sox14b</i>  | EU399804    |    | <i>Sox14b<sup>d</sup></i> |
| TDN13, TDN19, TDN21         | 3   | 162                   | <i>Sox19b</i>  | EU399807    | B  | <i>Sox19b</i>             |
| TDN7, TDN9                  | 2   | 162                   | <i>Sox21a</i>  | EU399806    |    | <i>Sox21a</i>             |
| TDN4, TDN8                  | 2   | 162                   | <i>Sox21b</i>  | EU399805    |    | <i>Sox21b</i>             |

a) TDN 表示克隆来自引物 *Sox*N 的扩增产物, b) TD9 表示克隆来自引物 *Sox*9 的扩增产物; c) 括号内数字为内含子长度; d) 红鳍东方鲀的 *Sox* 基因

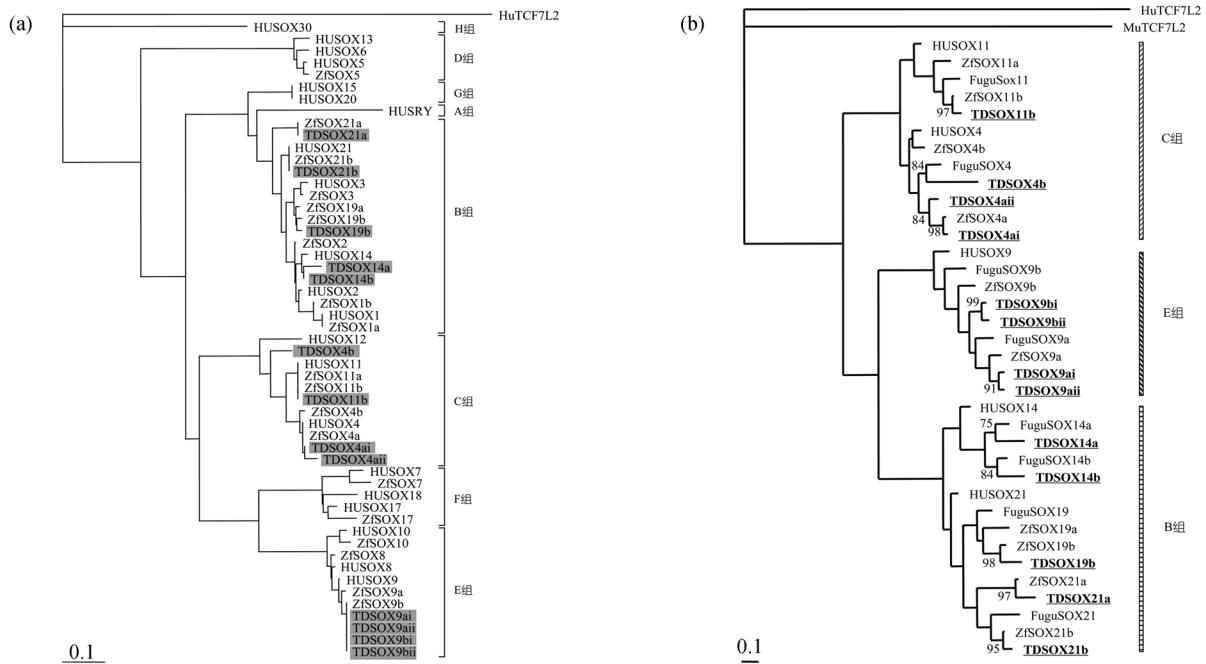


图2 以序列 Hu-TCF7L2 为外类群的大鳞结鱼 Sox 基因 HMG 结构域序列系统进化分析

(a) 基于氨基酸序列的 ME 分析; (b) 基于核苷酸序列的贝叶斯分析. 图中给出了大鳞结鱼 Sox 基因所属组别. (b) 中各节点数字为贝叶斯后验概率(Bayesian posterior probabilities, BPP). HU, 人; TD, 大鳞结鱼; Zf, 斑马鱼; Fugu, 红鳍东方在(*Takifugu rubripes*)

基因序列在人中没有对应的基因. 使用引物 *Sox*9 产生的 PCR 产物长度从 200~1300 bp 不等, 得到的基因属于 *SoxB* 和 *SoxE* 组, 相对应的人的 *Sox* 基因分别为 *Sox9* 和 *Sox14*. 值得注意的是, 这些 *Sox* 基因在人中只有一个拷贝, 而在大鳞结鱼中多数有一个以上的

拷贝. 对核苷酸序列的 Bayesian 分析结果(图 2(b))同样显示, 所得 *Sox* 基因分别属于 *SoxB*, *SoxC* 和 *SoxE* 三个不同组别, 表明所得的大鳞结鱼 *Sox* 基因在斑马鱼或红鳍东方在中的直系同源基因分别为 *Sox4b*, *Sox11b*, *Sox14a*, *Sox14b*, *Sox19b*, *Sox21a*, *Sox21b* 以及

分别具有 2 个不同拷贝的 *Sox4a*, *Sox9a* 和 *Sox9b*. 其中在人中没有直系同源基因的序列, 其在斑马鱼中的直系同源基因为 *Sox19b*.

### 2.2 大鳞结鱼 *Tor douronensis* 中 *Sox* 基因序列特征及演化分析

表 2 显示了大鳞结鱼 *Sox* 基因序列与人及斑马鱼或红鳍东方在的直系同源(ortholog)基因及并系同源(Paralog)基因之间在核苷酸和氨基酸序列的相似性. 与人相比, 在核苷酸水平, 大鳞结鱼 *Sox* 基因序列与其斑马鱼的直系同源基因之间具有更高的相似性; 在氨基酸水平则没有差异, 表明突变主要为同义突变. 基因 *Sox4a* 的两个不同拷贝 *Sox4ai* 和 *Sox4aii* 与斑马鱼 *Sox4a* 基因之间的核苷酸相似性分别为 96.9% 和 90.7%, 而它们之间的核苷酸相似性为 90.1%; *Sox4ai* 与斑马鱼 *Sox4a* 的氨基酸序列完全一致, *Sox4aii* 与 *Sox4ai* 和斑马鱼 *Sox4a* 的氨基酸序列相似性则同为 98.1%. 相对于斑马鱼的 *Sox9a* 和 *Sox9b*, 大鳞结鱼 *Sox9a* 和 *Sox9b* 的两个拷贝间具有更高的核苷酸和氨基酸相似性. 有趣的是, *Sox9a* 两个不同拷贝所具有的内含子序列在长度和相似性上的差异非常大, 长度分别为 191 和 483 bp, 相似性仅为 35%;

*Sox9b* 两个不同拷贝所具有的内含子长度分别为 661 和 679 bp, 相似性为 68.8%.

对于大鳞结鱼中基因 *Sox4a*, *Sox9a* 和 *Sox9b* 两个不同拷贝的核苷酸替代速率的分析见表 3. 分析结果显示, 对于由多倍化形成的基因 *Sox4a*, *Sox9a* 和 *Sox9b* 的两个不同拷贝, 它们之间核苷酸的替代速率, 同义替代速率远远大于非同义替代率, *Z* 检验的结果也显著地拒绝了中性替代的假设.

## 3 讨论

### 3.1 大鳞结鱼中 *Sox* 基因家族成员及其命名

通过对所得到的 *Sox* 基因序列特征及系统演化关系分析, 本研究最终在年轻四倍体大鳞结鱼中共得到 13 条 *Sox* 基因的 HMG 功能结构域的分序列, 分别属于 *SoxB*, *SoxC* 和 *SoxE* 组<sup>[2]</sup>. 对于所得到的每一条 *Sox* 基因序列, 都是至少两个不同克隆整序的结果, 因此保证了所获得的 *Sox* 基因序列的可信性. 根据系统进化分析的结果及其在斑马鱼或红鳍东方鲀中的直系同源基因, 所得到的大鳞结鱼 *Sox* 基因分别为 *Sox4b*, *Sox11b*, *Sox14a*, *Sox14b*, *Sox19b*, *Sox21a*, *Sox21b*. 对于具有两个拷贝的基因 *Sox4a*, *Sox9a* 和

表 2 大鳞结鱼 *Sox* 基因与其直系同源基因及并系同源基因在核苷酸及氨基酸水平上序列之间的相似性(百分比)

| 编号 | 大鳞结鱼                       | 重复基因对   | 斑马鱼        |            | 人          |
|----|----------------------------|---|------------|------------|------------|
|    |                            |   | a          | b          |            |
| 1  | <i>Sox4ai</i>              | 90.1 <sup>a</sup> /98.1 <sup>b</sup>                            | 96.9/100.0 | 84.6/98.1  | 85.8/100.0 |
| 2  | <i>Sox4aii</i>             |   | 90.7/98.1  | 87.7/96.3  | 90.1/98.1  |
| 3  | <i>Sox4b</i>               | Vs. <i>Sox4ai</i> : 76.5/94.4<br>Vs. <i>Sox4aii</i> : 76.5/92.6 | 77.8/94.4  | 74.1/94.4  | 76.5/94.4  |
| 4  | <i>Sox9ai</i>              | 84.6/100.0  | 78.4/87.0  | 75.3/88.9  | 74.1/87.0  |
| 5  | <i>Sox9aii</i>             |   | 78.4/87.0  | 74.7/88.9  | 74.7/87.0  |
| 6  | <i>Sox9bi</i>              | 84.0/100.0  | 74.1/87.0  | 78.4/88.9  | 72.8/87.0  |
| 7  | <i>Sox9bii</i>             |   | 75.3/87.0  | 77.8/88.9  | 70.4/87.0  |
| 8  | <i>Sox11b</i>              | NC <sup>c</sup>   | 85.2/100.0 | 95.1/100.0 | 84.6/100   |
| 9  | <i>Sox14a</i> <sup>d</sup> | Vs. <i>Sox14b</i> : 76.4/97.9                                   | 72.8/87.0  | 72.2/87.0  | 67.9/87.0  |
| 10 | <i>Sox14b</i> <sup>d</sup> |   | 73.5/88.9  | 75.3/88.9  | 66.0/88.9  |
| 11 | <i>Sox19b</i>              | NC  | 79.6/98.1  | 88.9/98.1  | NC         |
| 12 | <i>Sox21a</i>              | Vs. <i>Sox21b</i> : 75.3/96.3                                   | 88.9/100.0 | 76.5/96.3  | 75.9/96.3  |
| 13 | <i>Sox21b</i>              |   | 78.0/96.3  | 95.1/100.0 | 80.2/100.0 |

a) 核苷酸序列相似性; b) 氨基酸序列相似性; c) NC 表示无法计算; d) 红鳍东方在(*Takifugu rubripes*)的 *Sox* 基因

表 3 大鳞结鱼 *Sox* 基因整合重复基因对替代比较

| 重复基因对                            | 长度/bp | 同义替代数 | 非同义替代数 | 非同义替代速率/ <i>Ka</i> | 同义替代数速率/ <i>Ks</i> | <i>Ka/Ks</i> | <i>Z</i> 检验的 <i>P</i> 值 |
|----------------------------------|-------|-------|--------|--------------------|--------------------|--------------|-------------------------|
| <i>Sox4ai</i> vs. <i>Sox4aii</i> | 162   | 15    | 1      | 0.0087             | 0.4276             | 0.0203       | 0.0009 <sup>a</sup>     |
| <i>Sox9ai</i> vs. <i>Sox9aii</i> | 144   | 7     | 0      | 0.0000             | 0.1748             | 0.0000       | 0.0118                  |
| <i>Sox9bi</i> vs. <i>Sox9bi</i>  | 144   | 8     | 0      | 0.0000             | 0.1984             | 0.0000       | 0.0075                  |

a) 显著性水平,  $P < 0.05$

*Sox9b*, 根据 Moghadam 等人<sup>[37]</sup>对虹鳟(*Oncorhynchus mykiss*) *Hox* 基因的命名原则, 随机地将这些基因的两个不同拷贝命名为 *Sox4ai*, *Sox4aai*, *Sox9ai*, *Sox9aai*, *Sox9bi* 和 *Sox9bii*.

考虑到大鳞结鱼为四倍体( $N = 4$ ,  $2n = 100$ )<sup>[20]</sup>, 并且 *Sox* 基因 HMG 结构域序列含有足够的信息可用作鉴定不同 *Sox* 基因<sup>[2]</sup>, 所以基因 *Sox4a*, *Sox9a* 和 *Sox9b* 具有的两个拷贝应该由大鳞结鱼的四倍化引起的, 而不是等位基因的异质性或整合基因组重复(segmental duplication)的结果<sup>[37,38]</sup>. 当然如果彻底地解决这一问题, 一种有效的方法就是将这些基因的不同拷贝准确地定位到大鳞结鱼的不同染色体上. 对于本研究中获得的大多数 *Sox* 基因, 如 *Sox4b*, *Sox11b*, *Sox14a* 和 *Sox14b* 等, 均为单一拷贝. 对此一种可能的解释就是因为大鳞结鱼发生四倍化的时间较短<sup>[27]</sup>, 导致这些基因的两个拷贝在序列上是相同的; 另外一种可能则是 PCR 扩增过程中因引物或者退火温度等原因造成这些基因不同拷贝扩增的不平衡, 导致最后所得到的含有这些基因不同拷贝的克隆数量上偏差. 对此, 要得到这些 *Sox* 基因的其他拷贝可能需要更多克隆的整序. 此外, *Sox* 基因家族的一些成员在哺乳动物中为性别决定因子, 因此通过对雄性大鳞结鱼的研究, 可能会得到更多的或者不同的 *Sox* 基因家族成员, 并对大鳞结鱼或者真骨鱼类的性别决定机制中 *Sox* 基因所起到的作用有一定的启示.

### 3.2 *Sox* 基因在大鳞结鱼中的演化

多倍化所导致的全基因组重复, 最直接的影响就是造成基因组中功能基因的重复. 在本研究中, 大鳞结鱼的基因 *Sox4a*, *Sox9a* 和 *Sox9b* 具有两个不同的拷贝. 有关这些重复基因在分子水平上所受到的选择压力及进化命运, 在理论上, Lynch 等人<sup>[39]</sup>将其总结为以下3种情况: ( ) 重复基因的一个拷贝因突变而失去功能成为假基因, 称之为无功能化(nonfunctionalization); ( ) 重复基因的一个拷贝因自然选择而获得新的有益的功能, 而另外一个拷贝则保留原来的功能, 称之为新功能化(neofunctionalization); ( ) 重复基因的两个拷贝因突变的积累而发生功能分化, 共同承担原来基因的整合功能, 称之为在功能化(subfunctionalization). 而 Lynch 等人<sup>[39]</sup>通过对人、小鼠(*M. musculus*)、鸡(*G. gallus*)、整虫(*C. elegans*)、果蝇(*D. melanogaster*)、拟

南芥(*A. thaliana*)和水稻(*O. sativa*)以及啤酒酵母(*S. cerevisiae*)基因组的分析结果显示, 这几种真核生物所具有的重复基因在其形成的初期都经历过一段选择压力很小的时期, 这段时期内, 重复基因的进化接近于中性进化. Brunet 等人<sup>[26]</sup>在真骨鱼中的研究也发现发生重复的基因对的进化速率是不同的. Luo 等人<sup>[40]</sup>在研究四倍体金鱼(*Carassius auratus auratus*)中 *Hox* 基因的进化时发现, 金鱼的 *Hox* 基因由碱基缺失而形成假基因的频率很高, 并认为是金鱼多倍化后因剂量效应(dosage effects)所导致的结果. 在本研究所分析的大鳞结鱼 *Sox* 基因中, 没有呈现类似于金鱼中 *Hox* 基因高频率碱基缺失的情况. 因多倍化而产生的重复基因对 *Sox4a*, *Sox9a* 和 *Sox9b*, 其核苷酸替代绝大多数为同义替代, 说明这些基因序列的进化是处于自然选择压力之下的,  $Z$  检验也显著地拒绝这些基因的进化为中性进化的假设, 并不支持 Lynch 等人<sup>[39]</sup>的研究结果, 而与 Hughes 等人<sup>[41]</sup>在非洲爪蟾(*Xenopus laevis*)中以及 Tiffin 等人<sup>[38]</sup>在玉米(*Zea mays*)的研究结果相类似. 有趣的是, 大鳞结鱼 *Sox9a* 和 *Sox9b* 基因不同拷贝的内含子长度及序列相似性差异非常大, 而在它们内含子中未发现微卫星或其它移动元件的插入, 而且序列分析表明内含子序列间无法很好地进行排序比对. 考虑到核苷酸序列的变异速率和大鳞结鱼的物种分化时间, 造成这一现象的可能原因是大鳞结鱼是通过异源多倍化(allopolyploidization)所形成的异源四倍体(allotetraploidy), 基因 *Sox9a* 和 *Sox9b* 不同拷贝的内含子长度及相似性差异在其两个祖先种中就已经存在. 当然, 要得到这一问题的明确答案, 还需要对大鳞结鱼中更多基因及其结鱼属中的近缘物种进行研究.

### 3.3 *Sox* 基因在脊椎动物基因组中的进化及鱼类特异的全基因组重复

本研究所得到的大鳞结鱼 *Sox* 基因分别属于 *SoxB*, *SoxC* 和 *SoxE* 组. 有趣的是, 大鳞结鱼中 *SoxB* 组的一个成员 *Sox19b*, 在哺乳动物中没有相应的直系同源基因, 而在斑马鱼<sup>[42]</sup>和红鳍东方<sup>[5]</sup>中则存在于其直系同源基因. 可能的原因是, *Sox19* 基因存在于真骨鱼与哺乳动物的共同祖先中, 而真骨鱼与哺乳动物分化之后, 哺乳动物丢失了基因 *Sox19*; 也可能是 *Sox19* 基因不存在于真骨鱼与哺乳动物的共同

祖先中,真骨鱼与哺乳动物分化之后,真骨鱼的祖先通过基因重复获得了 *Sox19*,并保留在现生真骨鱼的基因组中.而要解决这一基因的起源问题,需要研究一些原始脊椎动物以检整 *Sox19* 的同源直系基因是否存在他们的基因组中.

本研究结果显示,与哺乳动物相比较,大鳞结鱼中 *Sox* 基因(*Sox4*, *Sox9*, *Sox14* 和 *Sox21*)有 2~4 个不同的拷贝,同样的现象也存在于斑马鱼和红鳍东方在其他真骨鱼中<sup>[4,5,43]</sup>.有趣的是,大量在哺乳动物中为单拷贝的基因在真骨鱼中都有两个不同的拷贝.例如,在哺乳动物中只存在 4 个 *Hox* 基因簇,而在几种真骨鱼中则存在至少 7 个 *Hox* 基因簇<sup>[12,13,44,45]</sup>.因此,比较基因组学和系统进化研究的结果提出,在真骨鱼的祖先同哺乳类分离之后而尚未大量分化之前,其可能经历了一次特异的全基因组重复,所以导致现在哺乳动物中为单拷贝的基因在真骨鱼中都有两个不同的拷贝.在本研究中,综合所得到的大鳞结鱼 *Sox* 基因的数目及系统进化分析的结果,同样支持真

骨鱼特异的全基因组重复的假说(图 3). Meyer 等人<sup>[46]</sup>则提出脊椎动物基因/基因组进化的“one-to-four-to-eight”的观点,大鳞结鱼中 *Sox9* 相对于哺乳动物有四个不同的拷贝,因此相对于不同倍性的多倍体鱼类其基因/基因组进化模式可能为“one-to-four-to-eight-to-sixteen-to-more”,是研究多倍体脊椎动物进化的理想对象.考虑到 *Sox* 基因家族成员与 *Hox* 基因簇相比其在基因组中存在于不同的染色体上<sup>[38]</sup>,因此 *Sox* 基因与 *Hox* 基因簇一起成为研究真骨鱼是否经历过全基因组重复的理想模型.

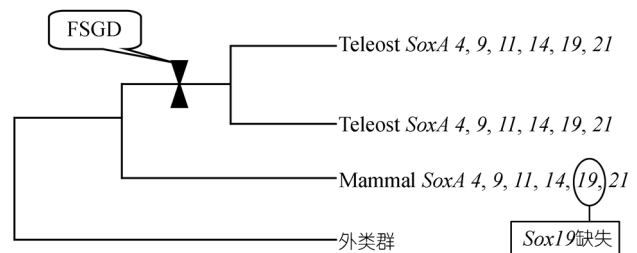


图 3 真骨鱼 *Sox* 基因支持鱼类特异的全基因组重复

致谢 非常感谢库喜英在论文撰写过程中给予的帮助和建议.

参考文献

- 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*, 1990, 346(6281): 245—250
- 2 Bowles J, Schepers G, Koopman P. Phylogeny of the *Sox* family of developmental transcription factors based on sequence and structural indicators. *Dev Biol*, 2000, 227(2): 239—255
- 3 Schepers G E, Teasdale R D, Koopman P. Twenty pairs of *Sox*: Extent, homology, and nomenclature of the mouse and human *Sox* transcription factor gene families. *Dev Cell*, 2002, 3(2): 167—170
- 4 de Martino S, Yan Y L, Jowett T, et al. Expression of *Sox11* gene duplicates in zebrafish suggests the reciprocal loss of ancestral gene expression patterns in development. *Dev Dyn*, 2000, 217(3): 279—292
- 5 Koopman P, Schepers G, Brenner S, et al. Origin and diversity of the *Sox* transcription factor gene family: Genome-wide analysis in *Fugu Rubripes*. *Gene*, 2004, 328: 177—186
- 6 Ohno S. *Evolution by Gene Duplication*. New York: Springer-Verlag, 1970
- 7 Wolfe K H. Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet*, 2001, 2(5): 333—341
- 8 Comai L. The advantages and disadvantages of being polyploidy. *Nat Rev Genet*, 2005, 6(11): 836—846
- 9 Holland P W, Garcia-Fernandez J, Williams N A, et al. Gene duplications and the origins of vertebrate development. *Development*, 1994, Suppl: 125—133
- 10 Skrabanek L, Wolfe K H. Eukaryote genome duplication - Where's the evidence? *Curr Opin Genet Dev*, 1998, 8(6): 694—700
- 11 Hughes A L, Friedman R. 2r or Not 2r: Testing hypotheses of genome duplication in early vertebrates. *J Struct Funct Genom*, 2003, 3(1-4): 85—93
- 12 Amores A, Force A, Yan Y L, et al. Zebrafish *Hox* clusters and vertebrate genome evolution. *Science*, 1998, 282(5394): 1711—1714
- 13 Christoffels A, Koh E G, Chia J M, et al. *Fugu* genome analysis provides evidence for a whole-Genome duplication early during the evolution of ray-finned fishes. *Mol Biol Evol*, 2004, 21(6): 1146—1151
- 14 Hoegg S, Brinkmann H, Taylor J S, et al. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost Fish. *J Mol Evol*, 2004, 59(2): 190—203
- 15 Vandepoele K, De Vos W, Taylor J S, et al. Major events in the genome evolution of vertebrates: Paraneome age and size differ

- considerably between ray-finned fishes and land vertebrates. *Proc Natl Acad Sci USA*, 2004, 101(6): 1638—1643
- 16 Crow K D, Stadler P F, Lynch V J, et al. The “fish-specific” Hox cluster duplication is coincident with the origin of teleosts. *Mol Biol Evol*, 2006, 23(1): 121—136
- 17 Le Comber S C, Smith C. Polyploidy in fishes: Patterns and processes. *Biol J Linnean Soc*, 2004, 82(4): 431—442
- 18 Allendorf F W, Thorgaard G H. Tetraploidy and the evolution of salmonid fish. In: Turner J B, ed. *Evolutionary Genetics of Fish*. New York: Plenum Press, 1984. 1—53
- 19 肖霁, 张仁东, 冯建国, 等. 17种金线鲃核DNA含量及倍性的研究. *动物学研究*, 2002, 23(3): 195—199
- 20 Zan R, Song Z, Liu W. Studies on karyotypes and DNA contents of some Cyprinoid fishes, with notes on fish polyploids in China. In: Uyeno T, Arai R, Taniuchi T, et al, eds. *Indo Pacific Fish Biology*. Tokyo: Ichthyol Soc Japan, 1986. 877—885
- 21 余先觉, 周曦, 李渝成, 等. *中国淡水鱼类染色体*. 北京: 科学出版社, 1989
- 22 Kellis M, Birren B W, Lander E S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*, 2004, 428(6983): 617—624
- 23 Maere S, de Bodt S, Raes J, et al. Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci USA*, 2005, 102(15): 5454—5459
- 24 Moore R C, Purugganan M D. The early stages of duplicate gene evolution. *Proc Natl Acad Sci USA*, 2003, 100(26): 15682—15687
- 25 Akhunov E D, Akhunova A R, Dvorak J. Mechanisms and rates of birth and death of dispersed duplicated genes during the evolution of a multigene family in diploid and tetraploid wheats. *Mol Biol Evol*, 2007, 24(2): 539—550
- 26 Brunet F G, Crollius H R, Paris M, et al. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Mol Biol Evol*, 2006, 23(9): 1808—1816
- 27 Wang X Z, Li J B, He S P. Molecular evidence for the monophyly of East Asian groups of Cyprinidae (Teleostei: Cypriniformes) derived from the nuclear recombination activating gene 2 sequences. *Mol Phylogenet Evol*, 2007, 42(1): 157—170
- 28 陈宜瑜. *中国动物志. 鱼纲鲤形目 II*. 北京: 科学出版社, 1998
- 29 Sambrook J, Fritsch E F, Maniatis T. *Molecular Cloning: A Laboratory Manual*, 2nd ed. New York: Cold Spring Harbor Laboratory Press, 1989
- 30 Galay-Burgos M, Llewellyn L, Mylonas C C, et al. Analysis of the *Sox* Gene family in the european sea bass (*Dicentrarchus Labrax*). *Comp Biochem Physiol B Biochem Mol Biol*, 2004, 137(2): 279—284
- 31 Thompson J D, Gibson T J, Plewniak F, et al. The Clustal\_X Windows interface: Flexible strategies for multiple sequences alignment aided by quality analysis Tools. *Nucleic Acids Res*, 1997, 25(24): 4876—4882
- 32 Tamura K, Dudley J, Nei M, et al. Mega 4: Molecular evolutionary genetics analysis (Mega) software version 4.0. *Mol Biol Evol*, 2007, 24(8): 1596—1599
- 33 Ronquist F, Huelsenbeck J P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 2003, 19(12): 1572—1574
- 34 Posada D, Crandall K A. Modeltest: Testing the model of DNA substitution. *Bioinformatics*, 1998, 14(9): 817—818
- 35 Campanella J, Bitincka L, Smalley J. Matgat: An application that generates similarity/identity matrices using protein or DNA sequences. *BMC Bioinformatics*, 2003, 4(1): 29
- 36 Zhang J, Rosenberg H F, Nei M. Positive darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci USA*, 1998, 95(7): 3708—3713
- 37 Moghadam H K, Ferguson M M, Danzmann R G. Evidence for *Hox* gene duplication in rainbow trout (*Oncorhynchus Mykiss*): A tetraploid model species. *J Mol Evol*, 2005, 61(6): 804—818
- 38 Lynch M, Conery J S. The evolutionary fate and consequences of duplicate genes. *Science*, 2000, 290(5494): 1151—1155
- 39 Luo J, Stadler P F, He S P, et al. PCR survey of *Hox* genes in the goldfish *Carassius auratus auratus*. *J Exp Zool B Mol Dev Evol*, 2007, 308(3): 250—258
- 40 Hughes M K, Hughes A L. Evolution of duplicate genes in a tetraploid animal, *Xenopus Laevis*. *Mol Biol Evol*, 1993, 10(6): 1360—1369
- 41 Tiffin P, Gaut B S. Sequence diversity in the tetraploid *Zea Perennis* and the closely related diploid *Z. diploperennis*: insights from four nuclear loci. *Genetics*, 2001, 158(1): 401—412
- 42 Vriza S, Lovell-Badge R. The zebrafish Zf-Sox 19 protein: A novel member of the Sox family which reveals highly conserved motifs outside of the DNA-binding domain. *Gene*, 1995, 153(2): 275—276
- 43 Chiang E F, Pai C I, Wyatt M, et al. Two *Sox9* genes on duplicated zebrafish chromosomes: expression of similar transcription activators in distinct sites. *Dev Biol*, 2001, 231(1): 149—163
- 44 Kurosawa G, Yamada K, Ishiguro H, et al. *Hox* gene complexity in medaka fish may be similar to that in pufferfish rather than zebrafish. *Biochem Biophys Res Commun*, 1999, 260(1): 66—70
- 45 Kurosawa G, Takamatsu N, Takahashi M, et al. Organization and structure of *Hox* gene loci in medaka genome and comparison with those of pufferfish and zebrafish genomes. *Gene*, 2006, 370: 75—82
- 46 Meyer A, Schartl M. Gene and genome duplications in vertebrates: The one-to-four(-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol*, 1999, 11(6): 699—704