

半滑舌鳎(*Cynoglossus semilaevis*)*Sox* 基因家族 生物信息学分析*

高 进^{1,2,3} 马佳璐^{2,3,4} 刘 洋^{1,2,3} 邵长伟^{2,3}
贾晓东² 陈松林^{1,2,3①}

(1. 南京农业大学无锡渔业学院 无锡 214081; 2. 农业部海洋渔业可持续发展重点实验室 中国水产科学研究院黄海水产研究所 青岛 266071; 3. 青岛海洋科学与技术国家实验室 海洋渔业科学与食物产出过程功能实验室 青岛 266200; 4. 中国海洋大学水产学院 青岛 266003)

摘要 *Sox* (*SRY*-related HMG-box) 基因家族是在动物体内发现的一类编码转录因子的基因家族, 广泛参与了动物生长发育、理化反应, 特别是性别决定和分化等过程。本研究采用生物信息学方法, 利用 *Sox* 基因高度保守的 HMG-box 区序列作为种子序列, 检索半滑舌鳎(*Cynoglossus semilaevis*) 基因组注释的蛋白数据库, 共鉴定分离出 23 个 *Sox* 基因。并在全基因组水平对半滑舌鳎 *Sox* 基因家族进行了保守结构域序列、进化、基因结构、染色体定位及基因表达模式的系统分析。结果显示, 通过保守结构域分析发现了半滑舌鳎 *Sox* 基因家族除 *CseSox32* 外, 皆存在一段 9 个氨基酸残基 (RPMNAFMVW) 的高度保守基序; 结合保守结构域及进化分析, 所有半滑舌鳎 *Sox* 基因被分为 B₁、B₂、C、D、E、F 和 K 共 7 个亚族, 且不同的亚族在进化上存在种间趋同性和种内特异性; 基因结构分析将半滑舌鳎 *Sox* 基因分为两大类, 即单外显子类和多外显子类; 而染色体定位分析则显示, *Sox* 基因在染色体上散乱分布, 不存在集簇现象; 不同类型性腺及幼鱼变态前后的 *Sox* 基因家族表达谱显示, 半滑舌鳎 *Sox* 基因具有不同的组织表达模式, 表明其在性别分化、性腺发育及早期幼鱼发育过程中可能发挥了重要作用。本研究对于今后半滑舌鳎 *Sox* 基因家族深入的功能验证及分子作用机制研究具有重要的意义, 也为日益丰富的水产基因组资源的挖掘利用提供了参考。

关键词 *Sox* 基因; HMG 结构域; 半滑舌鳎; 生物信息学

中图分类号 S91 **文献标识码** A **文章编号** 2095-9869(2016)02-0041-08

Sox (*SRY*-related HMG-box) 基因家族是一类 *SRY* (Sex determination region of Y chromosome) 相关基因构成的基因家族, 属于 HMG-box 超家族成员 (Laudet *et al*, 1993)。最早发现的 *Sox* 基因是在哺乳动物中鉴定出来的睾丸决定因子 *SRY* (Gubbay *et al*, 1990), 根据一段长约 79 个氨基酸的 DNA 结合基序 HMG-box 的保守性, 将在 HMG-box 区与 *SRY* 产物具有 50% 以上的氨基酸序列相似性的编码基因定义为 *Sox* 基因家

族成员 (Chardard *et al*, 1993)。随着种内及种间越来越多的 *Sox* 基因被发现, Bowles 等 (2000) 提出了新的 *Sox* 基因定义标准, 即以 HMG 结构域的第 5-10 个氨基酸残基 (RPMNAF) 作为识别 *Sox* 基因的标志性基序, 其对除 *SRY* 外的所有 *Sox* 基因都是保守的。目前, 根据 HMG-box 序列的系统发生分析、蛋白结构特征及功能的不同, *Sox* 基因分属 A-J 共 10 个亚族 (Schepers *et al*, 2002)。此外, 青鳉中 *Sox* 基因全基因组范围的

* 国家自然科学基金项目 (31130057)、山东省自然科学基金 (ZR2014CP011)、山东省农业良种工程重大课题和泰山学者攀登计划专项共同资助。高 进, E-mail: gaojin427@126.com

① 通讯作者: 陈松林, 研究员, E-mail: chensl@ysfri.ac.cn

收稿日期: 2016-01-15, 收修改稿日期: 2016-02-21

分析将 *Olsox32* 定义为 K 亚族(Cui *et al.*, 2011)。广泛存在的 *Sox* 基因在软骨形成、血细胞生成、神经系统发育、晶状体形成以及性别决定和分化等早期胚胎发育过程中发挥着重要的作用(常重杰等, 2002)。

鉴定基因家族的方法主要有分子生物学和生物信息学两种, 随着越来越多物种的基因组测序和基因注释的完成, 利用生物信息学方法对不同基因家族成员进行鉴定和分析, 这对于筛选重要性状相关功能基因具有重要意义。目前, 基于这种生物信息学分析的方法对感兴趣的基因家族进行分析已在完成基因组测序的动植物中应用(于雪等, 2014; 梁小娟等, 2015; 丁安明等, 2014), 从而成为在全基因组层面鉴定和研究基因家族的重要手段。*Sox* 基因家族分析不仅有整体上的结构和胚胎发育时期的表达研究(Bowles *et al.*, 2000; Schepers *et al.*, 2002; Cui *et al.*, 2011), 同时也有基因家族成员或亚族在生理及生化层面上的功能分析(Barrionuevo *et al.*, 2010; Azim *et al.*, 2009; Hosking *et al.*, 2009)。杨超(2013)¹⁾进行罗非鱼 *Sox* 基因家族分离鉴定及生物信息学分析, 并综合分析了性腺转录组数据。实验不仅对 *Sox* 基因在不同性别、不同性腺发育时期、关键阶段的表达模式和动态变化规律进行了全面分析, 也为整个基因家族在性别决定和性腺发育过程中潜在作用的研究奠定了基础。而家蚕 *Sox* 基因家族的鉴定和研究(魏玲, 2011)²⁾不仅揭示了 *BmsoxE* 在家蚕生殖腺发育中的重要作用及调控机理, 并且有助于家蚕性别决定与分化及生殖腺发育的分子机制研究。

半滑舌鳎不仅是重要的海水养殖鱼类, 也是性别决定机制研究的理想鱼类模型, 具有很高的经济价值和理论研究价值。而基因组测序的完成(Chen *et al.*, 2014)也为半滑舌鳎 *Sox* 基因家族的全基因组生物信息学分析奠定了基础。性别决定与分化一直是 *Sox* 基因功能研究的经典领域, 同时也是半滑舌鳎研究中亟待解析的分子机制之一。迄今, 有关半滑舌鳎 *Sox* 基因的研究主要有 *Sox10* 的 cDNA 克隆和组织、胚胎与幼鱼的表达分析以及 *Sox9a* 的克隆与表达分析等(Deng *et al.*, 2008; Dong *et al.*, 2011), 尚未见有关半滑舌鳎 *Sox* 基因家族的全基因组分析的研究报道。鉴于 *Sox* 基因家族在半滑舌鳎早期发育及性别调控中的重要功能, 运用生物信息学的方法在基因组中进行分离、鉴定及分析, 可为 *Sox* 基因在半滑舌鳎性别调控和胚胎发育中的功能研究提供参考。

1 材料与amp;方法

1.1 基因组及功能注释信息

从 NCBI(<http://www.ncbi.nlm.nih.gov/>)下载半滑舌鳎基因组数据及基因注释文件并进行数据转化, 用于构建本地 BLAST 基因组数据库。构建进化树的人(*Homo sapiens*)、斑马鱼(*Danio rerio*)、青鳉(*Oryzias latipes*)、罗非鱼(*Oreochromis niloticus*)、河豚(*Takifugu rubripes*)和三棘刺鱼(*Gasterosteus aculeatus*)的基因组和蛋白编码基因序列则来自 Ensemble (<http://asia.ensembl.org/index.html>)。

1.2 Sox 蛋白序列鉴定

以 *Sox* 基因家族 HMG-box 一致序列(DHVKR-PMNAFMVWSRGERRKIAQQNPDMHNSEISKRLGKRWKLLSESEKRPFIEEAERLRAQHMKDYPDYKYRPRRKKK)为种子序列(汪锐等, 2002), 将其编译为 Fasta 格式并在半滑舌鳎本地 BLAST 基因组数据库中进行氨基酸的同源性比对, 根据比对得到的结果筛选出 *Sox* 基因氨基酸序列的候选片段。同时, 利用保守域预测的在线软件 Pfam (<http://pfam.xfam.org/>)和 SMART (<http://smart.embl-heidelberg.de/>)确保上述候选氨基酸片段含有 HMG-box 保守域(Finn *et al.*, 2015; Letunic *et al.*, 2015), 完成半滑舌鳎 *Sox* 基因家族成员的分离鉴定。

1.3 蛋白保守结构域分析

将种子序列与半滑舌鳎 *Sox* 基因蛋白序列 BLAST 后得到的一段 79 aa 序列定义为 *Sox* 基因蛋白保守结构域(HMG 结构域), 全部保守结构域序列经 ClustalX 软件比对后, 利用 GeneDoc 软件实现比对结果的可视化。

1.4 基因结构分析及染色体定位

为分析半滑舌鳎 *Sox* 基因家族各个成员的基因结构, 利用在线 GSDS(Gene Structure Display Server) 2.0 软件(<http://gsds.cbi.pku.edu.cn/index.php>)显示外显子、内含子等结构特征(Hu *et al.*, 2015)。同时, 实现基因家族成员间的结构比对。

结合半滑舌鳎基因组数据及 *Sox* 基因编码信息对 *Sox* 家族基因进行染色体定位。

1.5 基因进化分析

分析 *Sox* 基因家族成员的进化关系, 进一步将半

1) 杨超. 尼罗罗非鱼 *Sox* 家族生物信息学分析及 *Sox9a/b* 在性腺发育中的可能作用. 西南大学博士研究生学位论文, 2013, 2-23

2) 魏玲. 家蚕 *Sox* 基因的鉴定和功能研究. 西南大学博士研究生学位论文, 2011, 19-50

滑舌鳎 *Sox* 基因家族成员与人、斑马鱼、青鳉、尼罗罗非鱼、河豚及三棘刺鱼等模式动物的 *Sox* 基因构建系统发育树。使用 ClustalX 软件对 *Sox* 基因氨基酸序列进行多序列联配, 将联配的结果导入 MEGA 6.06 软件 (Tamura *et al.*, 2013) 中, 采用邻接树法 (Neighbor-Joining) 以 MEGA 软件构建进化树, 设置 Bootstrap 为 10000 次重复检测其置信度。将进化树导入 FigTree 软件中优化, 获取最终的系统进化树图。

1.6 基因表达分析

半滑舌鳎各个时期及组织的转录组数据来源于 NCBI(National Center for Biotechnology Information) 的 SRA 数据库, 主要分析雌雄鱼及伪雄鱼性腺组织和变态前后幼鱼的基因表达谱。半滑舌鳎发育过程中基因表达谱数据来源组织包括精巢(ZZ, SRX106096)、卵巢(ZW, SRX106097)、伪雄鱼性腺(ZW, SRX106098)、变态前幼鱼全身(ZW, SRX106100)及变态后幼鱼全身(ZW, SRX106101)。其中, 精巢、卵巢和伪雄鱼性腺为 1 龄左右, 精巢和伪雄鱼性腺较小, 黑色, 外覆一

层白膜, 成熟的精巢里有白色液体; 卵巢较大, 透明, 外覆一层无色膜; 幼鱼变态前后, 形态差别较大。变态前幼鱼为左右对称状, 浮游生活; 而变态后幼鱼眼部发生偏转, 呈左右不对称的扁平状, 且开始底栖生活。参考 Chen 等(2014)的基因表达值计算方法, 用 RPKM(Reads Per Kilobase of exon model per Million mapped reads)表示基因表达值大小。利用 R 语言的 pheatmap 程辑包对半滑舌鳎 *Sox* 基因在不同性腺组织及变态前后的基因表达数据进行数据聚类分析和作图。

2 结果与分析

2.1 半滑舌鳎 *Sox* 基因家族成员的分离鉴定

基于半滑舌鳎基因组数据, 以 HMG-box 区种子序列为探针进行同源比对搜索, 结合基因注释信息以及在线软件的保守结构域验证结果, 本研究从半滑舌鳎基因组中共分离鉴定出 23 个 *Sox* 基因。所有的 *Sox* 基因家族成员分属 7 个亚族(表 1), 分别为 B1 亚族

表 1 半滑舌鳎 *Sox* 基因家族成员
Tab.1 The members of *Sox* family in *C. semilaevis*

基因名称 Gene name	亚族 Subfamily	染色体 Chr	定位 Location	内含子 Introns	氨基酸长度 Length of amino acid (aa)
<i>CseSox1a</i>	B ₁	16	3513987:3515030	0	347
<i>CseSox1b</i>	B ₁	16	3511437:3513767	1	234
<i>CseSox2</i>	B ₁	2	11745338:11746300	0	320
<i>CseSox3</i>	B ₁	15	7919138:7920040	0	300
<i>CseSox19</i>	B ₁	1	34055459:34059837	0	273
<i>CseSox14a</i>	B ₂	2	19688673:19689437	0	254
<i>CseSox14b</i>	B ₂	20	4440262:4440990	0	242
<i>CseSox21</i>	B ₂	16	6061385:6062131	0	248
<i>CseSox4a</i>	C	13	7152957:7154195	0	412
<i>CseSox4b</i>	C	18	4387800:4388855	0	351
<i>CseSox11</i>	C	1	8543232:8544329	0	365
<i>CseSox5</i>	D	8	27203273:27269289	13	765
<i>CseSox6a</i>	D	5	4896103:4982230	13	758
<i>CseSox6b</i>	D	6	5526510:5582721	12	782
<i>CseSox13</i>	D	11	10848352:10861197	12	636
<i>CseSox8a</i>	E	17	8377577:8379370	2	467
<i>CseSox8b</i>	E	8	11421914:11423988	2	465
<i>CseSox9a</i>	E	8	12667918:12669974	2	591
<i>CseSox9b</i>	E	17	8571305:8574670	3	477
<i>CseSox10a</i>	E	17	12429291:12432070	2	497
<i>CseSox10b</i>	E	9	13240812:13242979	2	500
<i>CseSox18</i>	F	10	13797087:13800724	1	590
<i>CseSox32</i>	K	3	13775043:13818974	2	264

CseSox2、*CseSox3*、*CseSox19*)、B2 亚族(*CseSox14a*、*CseSox14b*、*CseSox21*)、C 亚族(*CseSox4a*、*CseSox4b*、*CseSox11*)、D 亚族(*CseSox5*、*CseSox6a*、*CseSox6b*、*CseSox13*)、E 亚族(*CseSox8a*、*CseSox8b*、*CseSox9a*、*CseSox9b*、*CseSox10a*、*CseSox10b*)、F 亚族(*CseSox18*) 和 K 亚族(*CseSox32*)。

2.2 半滑舌鲷 *Sox* 基因家族蛋白结构域分析

通过对半滑舌鲷 *Sox* 共 79 aa 的 HMG 结构域进行分析,结果显示,半滑舌鲷 23 个 *Sox* 基因的结构域是高度保守的(图 1)。本研究分离到的半滑舌鲷 *Sox* 基因除 *CseSox32* 外均含有 *Sox* 基因的保守识别基序(黑色线框标注区域),即满足 HMG 结构域第 5–10 个氨基酸残基(RPMNAF)的保守识别基序对除 *SRY* 外的所有基因都保守的原则,而 *CseSox32* 保守识别基序第 3 个为 L。*CseSox32* HMG 结构域上第 11–13 个氨基酸残基为 IIW,这也与其他 *Sox* 基因家族成员高度保守的 MVW 氨基酸残基不同。此外,全部的 23 个 *Sox* 基因在 HMG 结构域其他位置上也存在多个高度保守区域。

2.3 半滑舌鲷 *Sox* 基因结构分析

经基因结构分析,内含子和外显子结构分布显示,半滑舌鲷 *Sox* 基因家族根据基因结构可以大体分为两大类。第一类为单外显子类,无内含子,主要为 *CseSox1b* 和 *CseSox19* 以外的 B、C 亚族的家族成员。第二类为多外显子类,这类基因按照结构又可细分为两种,一种是内含子不多于 3 个的 *CseSox1b*、*CseSox19* 及 E、F 和 G 亚族的家族成员;另一种是内含子数目显

著多于 3 个且都多于 10 个的 D 亚族家族成员(图 2)。

2.4 半滑舌鲷 *Sox* 基因家族染色体定位

结合基因组及编码信息获取半滑舌鲷 *Sox* 基因家族相关染色体定位信息,结果显示,23 个半滑舌鲷 *Sox* 基因分布于 15 条染色体中,呈不均匀分布(表 1)。所有 *Sox* 基因均定位到常染色体上,未在性染色体上发现 *Sox* 基因家族成员的存在。其中,3、5、6、9、10、11、13、15、18 和 20 号染色体上只存在 1 个 *Sox* 基因,1、2 号染色体上存在两个 *Sox* 基因,而 8、16 和 17 号染色体则存在 3 个。

2.5 半滑舌鲷 *Sox* 基因进化分析

选取不同物种 *Sox* 基因编码的蛋白序列,利用 MEGA 6.06 软件,采用邻接法构建系统发育树(图 3)。结果显示,半滑舌鲷 *Sox* 基因与人、斑马鱼、青鳉、尼罗罗非鱼、河豚和三棘刺鱼一起被划分至 7 个主要的亚族(B-K),其具体的亚族分类保持着高度的一致性。

2.6 半滑舌鲷 *Sox* 基因表达分析

利用 R 语言的 *heatmap* 编辑包对半滑舌鲷 *Sox* 家族基因表达模式进行分析,绿-黑-红 3 个颜色表示基因的相对表达强度逐渐增大。本研究对 *Sox* 基因家族在精巢、卵巢、伪雄鱼性腺以及幼鱼变态前后的表达数据进行聚类分析(图 4),结果显示,半滑舌鲷的 *Sox* 基因主要被划分为两大类。一类在幼鱼变态前后基因有较高的表达,如图 4 中聚类的前 15 个 *Sox* 基因,其中,*CseSox4a*、*CseSox4b*、*CseSox9a* 及 *CseSox9b* 等基因变态前后表达差异显著,*CseSox4* 基因变态前

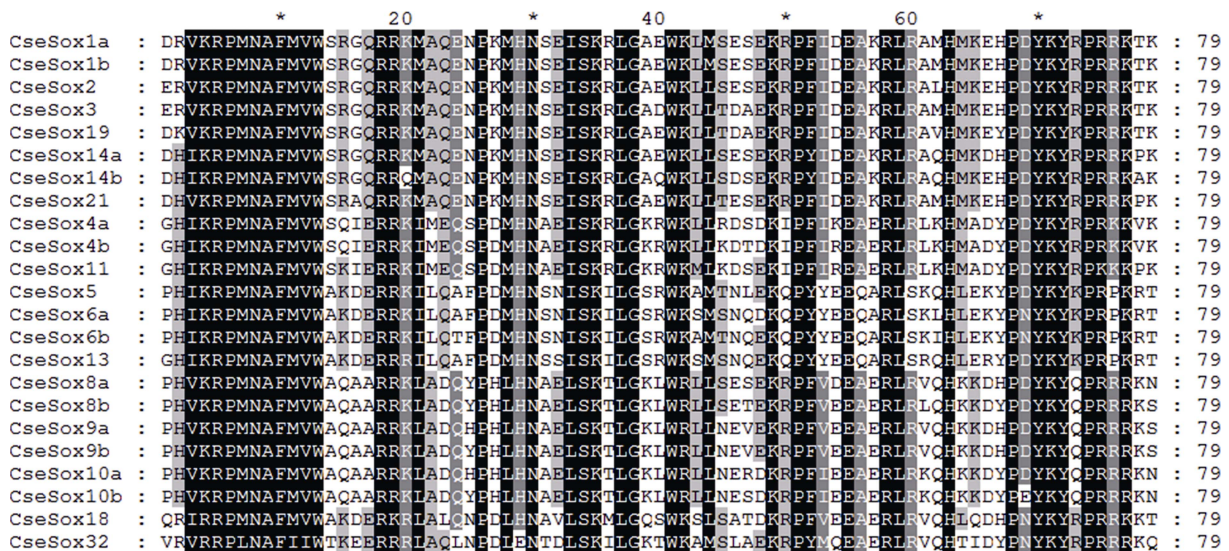


图 1 半滑舌鲷 *Sox* 基因家族保守结构域分析
Fig.1 Conserved domains analysis of *Sox* gene family in *C. semilaevis*

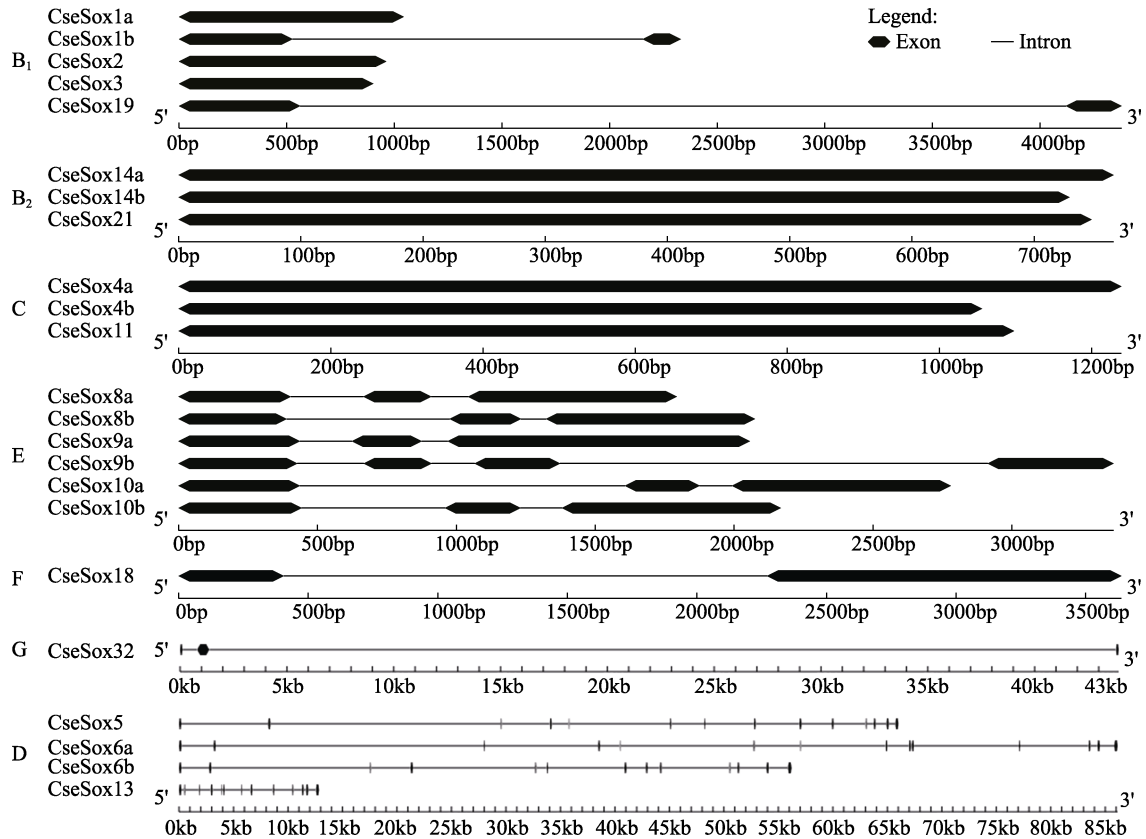


图 2 半滑舌鳎 *Sox* 基因家族基因结构示意图
Fig.2 Gene structure of *Sox* genes in *C. semilaevis*

基因表达高于变态后, 而 *CseSox9* 和 *CseSox9b* 则相反。另一类则在幼鱼变态前后表达较低, 但在性腺中却有较高的表达, 且在不同种类的性腺中具有表达差异存在, 这类的基因主要有 *CseSox1a*、*CseSox14b*、*CseSox32*、*CseSox18* 及 *CseSox13* 等。

3 讨论

基于完成测序的基因组数据, 从全基因组层面对基因家族进行分离鉴定、分类、序列特征、结构及进化等方面的分析已成为生物学研究领域的热点和要点。*Sox* 基因家族在动物特别是脊椎动物体内广泛存在, 且参与多个发育过程, 而 *Sox* 基因资源的挖掘也为其进一步的功能研究提供了参考借鉴。迄今为止, 通过分子克隆或基因组序列筛选等方法, 在线虫、果蝇等无脊椎动物中共鉴定出约 10 个 *Sox* 基因, 而在脊椎动物中已经鉴定出大约 40 个左右 *Sox* 基因 (Koopman *et al.*, 2004)。本研究从目前公开的半滑舌鳎基因组中, 使用生物信息学的方法一共鉴定出 23 个半滑舌鳎 *Sox* 基因。由于全基因组序列组装过程中存

在可能的 gaps (Cui *et al.*, 2011), 使得这一数目只是初步研究的结果, 随着 *Sox* 基因的克隆和基因组的组装、注释工作的不断完善或有增减。

与其他硬骨鱼类相似, 半滑舌鳎的基因家族成员被划分至 B–K 共 7 个亚族。蛋白结构域及基因结构分析发现, 其 HMG 结构域在种间及种内绝大部分成员中都是高度保守的, 而半滑舌鳎 *CseSox32* 基因的保守结构域虽然与其他家族成员有着一定的差异性, 但与其他物种的 K 亚族基因却保持着高度的一致性 (杨超, 2013)¹⁾。D 亚族基因相较于其他亚族含有更多内含子, 使其基因结构具有显著特点, 这可能是由于 D 亚族有着特殊的进化历程所导致的。基因结构示意图显示 (图 2), *CseSox13* 基因跨越长度接近 12 kb, 而 *CseSox5*、*CseSox6a* 及 *CseSox6b* 基因的跨越长度更是超过 50 kb。与 B、C 亚族内的单外显子基因相比, D 亚族基因在进化过程中趋于保守使大量内含子得以保留, 而那些单外显子基因则在进化过程中发生内含子丢失现象 (Roose *et al.*, 1999)。此外, 半滑舌鳎 *Sox* 基因染色体定位发现, *Sox* 基因在染色体上

1) 杨超. 尼罗罗非鱼 *Sox* 家族生物信息学分析及 *Sox9a/b* 在性腺发育中的可能作用. 西南大学博士研究生学位论文, 2013, 2–23

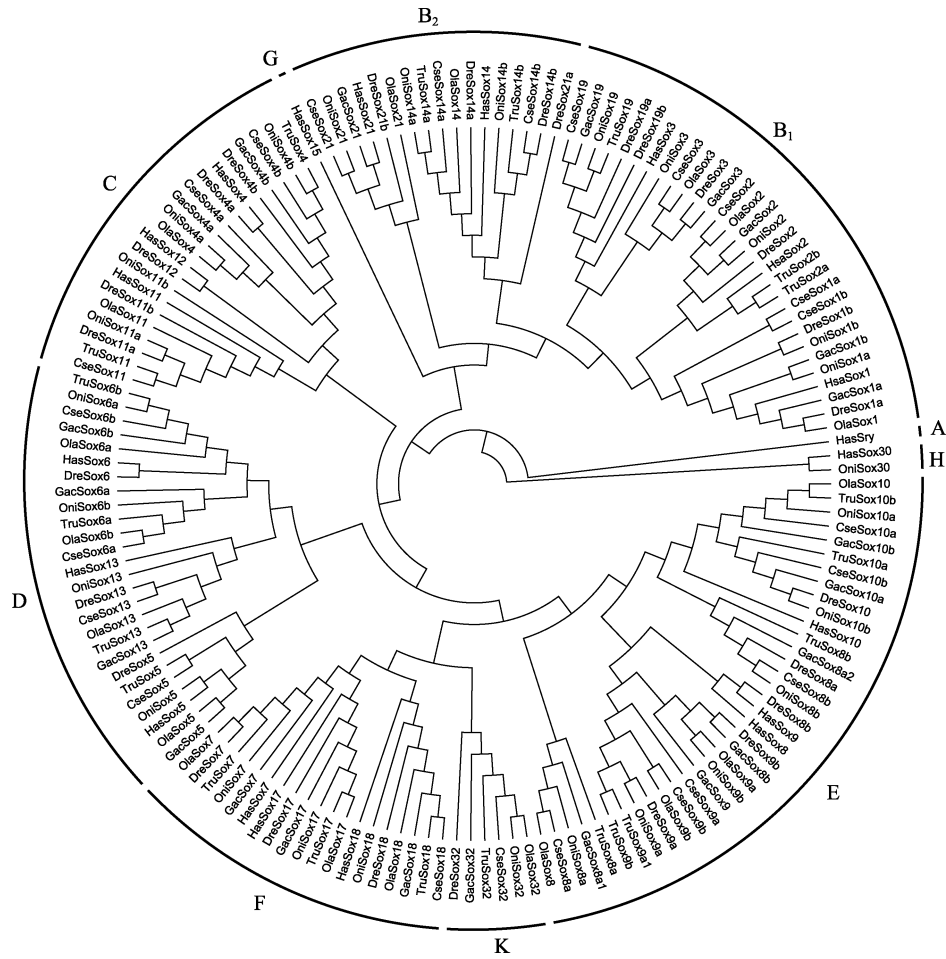


图 3 半滑舌鲷、人、斑马鱼、青鳉、尼罗罗非鱼、河豚、三棘刺鱼 Sox 家族进化树

Fig.3 Phylogenetic tree of *Cynoglossus semilaevis*, *Homo sapiens*, *Danio rerio*, *Oryzias latipes*, *Oreochromis niloticus*, *Takifugu rubripes* and *Gasterosteus aculeatus* Sox family

Cse: 半滑舌鲷; Has: 人; Dre: 斑马鱼; Ola: 青鳉; Oni: 罗非鱼; Tru: 河豚; Gac: 三棘刺鱼

Cse: *Cynoglossus semilaevis*; Has: *Homo sapiens*; Dre: *Danio rerio*; Ola: *Oryzias latipes*; Oni: *Oreochromis niloticus*; Tru: *Takifugu rubripes*; Gac: *Gasterosteus aculeatus*

的分布是散乱无规律的, 没有在特定染色体上发生优先主导现象, 同时表明其没有出现基因簇, 这一结果与其他物种上 Sox 基因家族的染色体分布情况相似 (Pevny *et al*, 1997)。

半滑舌鲷 Sox 基因系统进化分析表明, Sox 基因不同亚族之间在进化方向上表现出了差异性, 这也预示着其基因功能的多样性。多物种联合构建的系统发育树显示, 不同物种间及种内的 Sox 基因在总体上呈现出进化的趋同性, 但在不同物种中的 Sox 基因亚族类型及数目还是存在差异的。在 6 种硬骨鱼类中都存在 Sox32 基因, 而人类中却不存在 K 亚族的 Sox 基因, 但 A 亚族的 Sry 和 G 亚族的 Sox15 基因却是这 7 种物种中人所独有的。无脊椎动物 Sox 基因家族虽然成员数目较少, 但其亚族分类与脊椎动物 Sox 基因的亚族分类却是一致的 (Crémazy *et al*, 2000)。由此推测,

在整个动物进化过程中的某一时刻, Sox 基因家族亦随着动物的进化出现爆发性的拓群现象。在随后的一段稳定期内, HMG 基序的序列可能发生重复, 或者出现某一个多余的拷贝并进化为一个新的基因, 又或者从基因组中消失 (Force *et al*, 1999)。半滑舌鲷 Sox 基因表达信息分析结果显示, Sox 基因家族在半滑舌鲷性腺发育及早期幼鱼变态过程中具有一定的特异性。Dong 等 (2011) 对半滑舌鲷 Sox9a 基因及表达的克隆分析发现, CseSox9a 功能与性反转、性别分化、胚胎发育中细胞分化及生精细胞形成相关联。Kobayashi 等 (2008) 研究表明, 尼罗罗非鱼 Sox9a 基因在性腺分化及性反转过程中亦发挥重要作用。使用促性腺激素释放激素类似物 (GnRHa) 处理黑鲷导致 Sox3 及 DRMT1 的表达上调, 表明 Sox3 和 DRMT1 参与了黑鲷精巢发育过程 (Shin *et al*, 2009)。此外, 在小

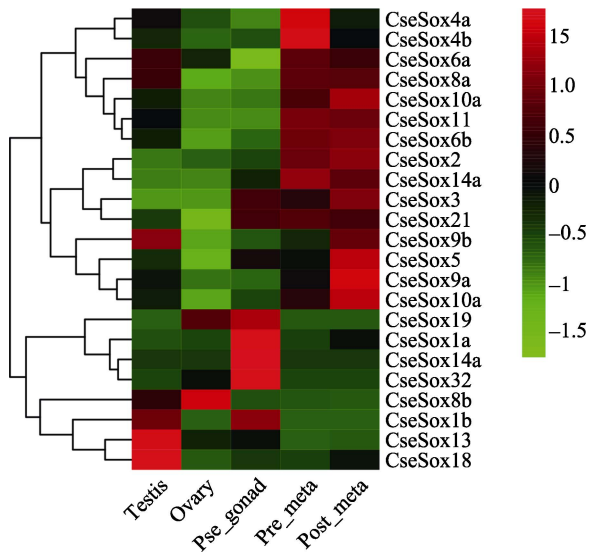


图 4 半滑舌鳎 *Sox* 基因表达模式

Fig.4 The expression profiles of *C. semilaevis Sox* genes

Testis: 精巢; Ovary: 卵巢; Pse_gonad: 伪雄鱼性腺;
Pre_meta: 幼鱼变态前; Post_meta: 幼鱼变态后
Pse_gonad: gonad of Pseudo-male; Pre_meta:
pre-metamorphosis; Post_meta: post-metamorphosis

鼠性别决定过程中, 利用双敲除的方法验证了 *Sox8* 和 *Sox9* 基因在精巢分化中起着关键性作用, 且 *Sox8* 对 *Sox9* 表达具有增强作用(Chaboissier *et al.*, 2004)。上述研究表明, *Sox* 基因家族在性别决定及性腺分化的过程中可能发挥重要作用, 而其在半滑舌鳎中的具体功能还有待进一步的实验验证。

4 结论

综上所述, 本研究首次运用生物信息学方法对半滑舌鳎 *Sox* 基因家族进行了全基因组层面的鉴定、保守结构域分析、染色体定位、基因结构分析、进化及表达分析。结果不仅表明了半滑舌鳎 *Sox* 基因家族在性别决定及早期发育过程中可能发挥着重要的作用, 也为其进一步的基因功能验证及日益丰富的水产基因组资源的挖掘利用提供参考。

参 考 文 献

丁安明, 李凌, 屈旭, 等. 番茄 PPR 基因家族的鉴定与生物信息学分析. 遗传, 2014, 36(1): 77-84
于雪, 于红, 孔令锋, 等. 太平洋牡蛎酪氨酸酶基因家族的系统发生分析. 遗传, 2014, 36(2): 135-144
汪锐, 程汉华, 郭一清, 等. 脊椎动物 *Sox* 基因家族的系统发生分析. 遗传学报, 2002, 29(11): 990-994
常重杰, 杜启艳, 邵红伟. *Sox* 基因家族研究的新进展. 遗传, 2002, 24(4): 470-476
梁小娟, 王志萌, 孟琼, 等. 猪 14-3-3 蛋白家族基因生物信息

学分析. 畜牧与兽医, 2015, 47(5): 31-37
Azim E, Jabaudon D, Fame RM, *et al.* SOX6 controls dorsal progenitor identity and interneuron diversity during neocortical development. Nat Neurosci, 2009, 12(10): 1238-1247
Barrionuevo F, Scherer G. SOXE genes: SOX9 and SOX8 in mammalian testis development. Int J Biochem Cell B, 2010, 42(3): 433-436
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
Chaboissier MC, Kobayashi A, Vidal VI, *et al.* Functional analysis of Sox8 and Sox9 during sex determination in the mouse. Development, 2004, 131(9): 1891-1901
Chardard D, Chesnel A, Gozé C, *et al.* Pw Sox-1: the first member of the Sox gene family in Urodeles. Nucleic Acids Res, 1993, 21(15): 35-76
Chen SL, Zhang GJ, Shao CW, *et al.* Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. Nature Genet, 2014, 46(3): 253-260
Crémazy F, Berta P, Girard F. Sox neuro, a new Drosophila Sox gene expressed in the developing central nervous system. Mech Develop, 2000, 93(1-2): 215-219
Cui J, Shen X, Zhao H, *et al.* Genome-wide analysis of Sox genes in Medaka (*Oryzias latipes*) and their expression pattern in embryonic development. Cytogenet Genome Res, 2011, 134(4): 283-294
Deng SP, Chen SL. cDNA cloning, tissues, embryos and larvae expression analysis of Sox10 in half-smooth tongue-sole, *Cynoglossus semilaevis*. Mar Genom, 2008, 1(3-4): 109-114
Dong XL, Chen SL, Ji XS, *et al.* Molecular cloning, characterization and expression analysis of Sox9a and Foxl2 genes in half-smooth tongue sole (*Cynoglossus semilaevis*). Acta Oceanol Sin, 2011, 30(1): 68-77
Finn RD, Coggill P, Eberhardt RY, *et al.* The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res, 2016, 44(7): 279-285
Force A, Lynch M, Pickett FB, *et al.* Preservation of duplicate genes by complementary, degenerative mutations. Genetics, 1999, 151(4): 1531-1545
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
Hosking B, François M, Wilhelm D, *et al.* Sox7 and Sox17 are strain-specific modifiers of the lymphangiogenic defects caused by Sox18 dysfunction in mice. Development, 2009, 136(14): 2385-2391
Hu B, Jin J, Guo AY, *et al.* GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics, 2015, 31(8): 1296-1297
Kobayashi T, Kajiura-Kobayashi H, Guan GJ, *et al.* Sexual dimorphic expression of DMRT1 and Sox9a during gonadal differentiation and hormone-induced sex reversal in the teleost fish Nile tilapia (*Oreochromis niloticus*). Dev Dynam, 2008, 237(1): 297-306
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(3): 177-186
Laudet V, Stehelin D, Clevers H. Ancestry and diversity of the HMG box superfamily. Nucleic Acids Res, 1993, 21(10): 2493-2501

- Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res*, 2015, 43(7): 257–260
- Pevny LH, Lovell-Badge R. Sox genes find their feet. *Curr Opin Genet Dev*, 1997, 7(3): 338–344
- Roose J, Korver W, Boer R, *et al.* The Sox-13 gene: structure, promoter characterization, and chromosomal localization. *Genomics*, 1999, 57(2): 301–305
- Schepers GE, Teasdale RD, 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
- Shin HS, An KW, Park MS, *et al.* Quantitative mRNA expression of sox3 and DMRT1 during sex reversal, and expression profiles after GnRHα administration in black porgy, *Acanthopagrus schlegelii*. *Comp Biochem Phys B Biochem Mol Biol*, 2009, 154(1): 150–156
- Tamura K, Stecher G, Peterson D, *et al.* MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol*, 2013, 30(12): 2725–2729

(编辑 冯小花)

Bioinformatics Analysis of Sox Gene Family in *Cynoglossus semilaevis*

GAO Jin^{1,2,3}, MA Jialu^{2,3,4}, LIU Yang^{1,2,3}, SHAO Changwei^{2,3}, JIA Xiaodong², CHEN Songlin^{1,2,3}①

(1. Wuxi Fisheries College, Nanjing Agricultural University, Wuxi 214081; 2. Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071; 3. Function Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266200; 4. College of Aquaculture, Ocean University of China, Qingdao 266003)

Abstract Sox genes are developmental regulators characterized by the presence of an HMG (high mobility group) DNA-binding domain, and they exist extensively in animal bodies in nature. Members of the Sox gene family have been shown to be conserved during evolution and play crucial roles in a wide variety of developmental processes, including growth and development, hemopoiesis, the development and formation of the nervous system, especially, sex determination and differentiation. Genome resources are growing faster and faster, and it provides a platform for the study gene families by scanning whole genomic sequences. We study Sox genes in *Cynoglossus semilaevis* by using bioinformatics analysis on the standard of whole genome. A total of 23 Sox genes were identified systematically from *C. semilaevis* and classified into 7 subfamilies which had not been reported yet. They are subfamily B₁, B₂, C, D, E, F and K, each of these groups may have distinct and specific functions. Sequence analysis of conserved domain of Sox genes in *C. semilaevis* shows that, all of members in the family contain 9 amino acid residues (RPMNAFMVW) as a highly conserved motif except *CseSox32*. On the other hand, the motif in *Sox32* is identical in all specie as RPLNAF. Research on the highly conserved HMG box in *C. semilaevis* causes the discovery of these Sox genes, and phylogenetic analysis supports the classification of subfamily. To date, about 10 and 40 Sox genes have been identified in invertebrates and vertebrates separately. The phylogenetic analysis also suggests that the groups of Sox genes might be extended in the process of evolution. Subfamilies have got convergence between different species, and got otherness inside one of these species. Gene location results show that these 23 Sox genes were distributed randomly on 15 chromosomes, and we do not find gene cluster. The analysis of gene structure divides all of the Sox genes into two categories, evaluation on the basis of the exon numbers of the gene are less or more than two. The expression profiles of *C. semilaevis* Sox genes show a characteristic of different kinds of gonad and metamorphosis specific, and it indicated Sox genes may play key roles during sex determination, gonad differentiation and early development in *C. semilaevis*. To sum up the above, we enhance understanding of Sox genes in *C. semilaevis* with bioinformatics analysis, such as system evolution, gene structure, chromosomal location and gene expression patterns. Our results suggest that bioinformatics analysis may contribute to further functional verification of gene family and be good for genome resources mining and utilization.

Key words Sox genes; HMG motif; *Cynoglossus semilaevis*; Bioinformatics

① Corresponding author: CHEN Songlin, E-mail: chensl@ysfri.ac.cn