

DOI: 10.3724/SP.J.1118.2019.19087

大黄鱼雌雄性腺长链非编码 RNA 的挖掘与差异分析

王伟佳^{1,2}, 韩兆方¹, 李完波², 王志勇¹

1. 集美大学农业农村部东海海水健康养殖重点实验室, 福建 厦门 361021;

2. 青岛海洋科学与技术试点国家实验室, 海洋渔业科学与食物产出过程功能实验室, 山东 青岛 266071

摘要: 为探究长链非编码 RNA 在大黄鱼性腺发育与分化中的作用, 从雌雄各 3 尾大黄鱼(*Larimichthys crocea*)的性腺中提取总 RNA, 进行去除 rRNA 的链特异性转录组建库和二代测序。将测序数据比对到大黄鱼参考基因组, 经比对、组装共得到来自 31675 个基因的 66088 个转录本, 严格筛选得到来自 3984 个基因位点的 5162 条 lncRNA。进一步分析获得了在大黄鱼雌雄性腺中差异表达的 mRNA 9341 个, lncRNA 2782 个, 高度相关的 lncRNA-mRNA 对 1227 个; 有多个 lncRNA 靶向已知的性别分化和发育相关基因, 其中 lncRNA MSTRG.24346 与大黄鱼的性别决定候选基因 *dmrt1* 距离相近, 且相关性极显著。该研究表明 lncRNA 可能在大黄鱼性别分化中起到重要作用, 值得深入研究阐明机制。

关键词: 大黄鱼; 性腺; lncRNA; 性别分化

中图分类号: S917

文献标志码: A

文章编号: 1005-8737-(2019)05-0852-09

长链非编码 RNA (long non-coding RNA, lncRNA) 是一种长度超过 200 个碱基且不翻译产生蛋白质的 RNA^[1]。*Xist* 是最早被发现的 lncRNA 之一, 它的功能是在雌性哺乳动物中控制两条 X 染色体中的一条染色体沉默, 从而起到剂量补偿效应的作用^[2]。多项研究表明, lncRNA 在剂量补偿效应^[3]、调控 mRNA 的表达^[4]和降解^[5]、构成细胞核内亚结构的结构骨架^[6]、作为染色体重塑的调节因子^[7]等方面发挥着重要的作用。lncRNA 与 mRNA 皆是由 RNA 聚合酶 II 转录而后拼接形成的, 部分 lncRNA 还具有多聚核苷酸的尾巴, 但 lncRNA 与 mRNA 的不同之处是其不含编码区并且保守性较低^[8]。lncRNA 主要分布在生物的细胞核中, 但也有少部分 lncRNA 分布在细胞质或细胞器中^[9]。根据 lncRNA 与其相邻的编码基因的位置关系, 有研究者将 lncRNA 分为 3 类: (1)位于编

码基因与编码基因之间的 lncRNA, 又称 lincRNA (long intergenic noncoding RNA); (2)编码基因内含子区 lncRNA; (3)编码基因反义链 lncRNA^[10]。lncRNA 在生物体内的作用机制主要可以分为 5 种: (1)信号作用, 如在 DNA 受损后, P53 结合到 *cdkn1a* 的启动子区域, 激活 lncRNA *panda*, *panda* 进一步作用于 NF- κ B 蛋白质, 从而限制凋亡前体基因的表达, 最终使细胞周期停滞^[11]; (2)诱饵作用, 如 lncRNA *gas5* 可以竞争性地与糖皮质激素受体 DNA 结合区域结合, 进而抑制受体和染色体的相互作用^[12]; (3)海绵作用, 如 lncRNA *ptenp1* 可能吸附某些特定 miRNA, 从而调控某些受该 miRNA 靶基因的表达^[13]; 具备该作用的 lncRNA 也被称为 ceRNA (competing endogenous RNA), 即竞争性内源 RNA^[14]; (4)引导作用, 如植物中的 lncRNA *coldair* 可以引导 PRC2 复合体到达 *flc* 基

收稿日期: 2019-02-22; 修订日期: 2019-04-04.

基金项目: 国家自然科学基金重点支持项目(U1705231); 国家海水鱼产业技术体系项目(CARS-47-G04); 福建省自然科学基金项目(2018J01450); 福建省教育厅中青年教师教育科研项目(JA15290).

作者简介: 王伟佳(1991-), 男, 硕士研究生, 研究方向为鱼类遗传育种. E-mail: viru@jmu.edu.cn

通信作者: 王志勇(1963-), 男, 教授, 博士生导师, 从事水生生物遗传育种与生物技术研究. E-mail: zywang@jmu.edu.cn

因的染色体位置,通过 H3K27 的三甲基化调控基因表达^[15]; (5) 支架作用, lncRNA *hotair* 可以与 PRC2 蛋白复合体结合,同时还可以与另一个包含 LSD1、CoREST 及 REST 等蛋白因子的去甲基化蛋白复合体结合,从而起到连接支架的作用^[16]。

多项研究表明, lncRNA 参与了生物的有性生殖^[17]、性腺发育^[18]、减数分裂^[19]、配子形成^[20-21]等过程。在黑龙江鲟(*Acipenser schrenckii*)^[22]、中华鳖(*Pelodiscus sinensis*)^[23]和小鼠^[24]中,通过生物信息学方法发现了多条 lncRNA 靶向与生殖相关的基因。Dmrt (double-sex and mab-3 related transcription factor) 基因家族是指编码具有高度保守的 DNA 结合结构域的转录因子的一类基因,该基因家族编码的蛋白可以以锌指方式与特定 DNA 序列结合并调控下游基因的表达,从而参与性别决定和分化等有关生理发育的过程^[25]。*dmrt1* 是 DMRT 家族的一个成员,其在青鳉(*Oryzias latipes*)^[26]和半滑舌鳎(*Cynoglossus semilaevis*)^[27-28]中被鉴定为性别决定基因。在小鼠中, lncRNA *Dmr* 参与 *Dmrt1* 转录后的剪接过程^[29]。Wen 等^[30]使用基因敲除技术敲除了果蝇中 105 个在睾丸中特异表达的 lncRNA,其中的 33 个 lncRNA 的缺失会导致雄性果蝇精子发生可见的发育缺陷,并失去生殖能力。以上研究都表明 lncRNA 在动物的性别分化与发育、成熟中发挥着重要的作用。

大黄鱼(*Larimichthys crocea*)是中国重要的海洋经济鱼类之一。大黄鱼在个体大小和生长速率等方面存在显著的雌雄二态性,雌鱼的生长速度明显快于雄鱼,性成熟时雌鱼的均重明显比雄鱼高,因此繁育大黄鱼的全雌群体被认为是提高大黄鱼养殖效率的一种有效方式。虽然本实验室之前就已经建立了通过将人工雌核发育产生的大黄鱼苗诱导性反转培育伪雄鱼(XX♂),再通过伪雄鱼与雌鱼配组生产遗传全雌大黄鱼的技术^[31],但是仍存在一些问題,例如伪雄鱼生长发育迟缓,个体严重偏小,产精量比较低。为解决上述问题,需要开发新的大黄鱼性别控制育种技术,为此对大黄鱼性别决定和性腺发育与分化调控机制开展深入研究十分必要。

本实验室通过构建大黄鱼高密度遗传图谱并

结合 QTL 定位技术,已将雄性性别决定区间定位到了 9 号染色体的 28~37.1 cM 区域^[32],在此区域内发现了 68 个基因, *dmrt1* (doublesex and mab-3 related transcription factor 1) 是该区域中在雌雄性腺中差异表达最显著的基因^[33]。lncRNA 在黑龙江鲟^[22]、中华鳖^[23]、小鼠^[29]、果蝇^[30]等生物性腺中的研究表明 lncRNA 在动物性腺中发挥着重要作用,由此推测 lncRNA 在大黄鱼的性腺发育中也可能起着重要作用。因此,本研究利用生物信息学的方法挖掘大黄鱼性腺中的 lncRNA,以期筛选到与性腺发育和分化相关的 lncRNA,为阐明大黄鱼性别决定与分化的机制提供基础。

1 材料与方法

1.1 材料

在福建宁德的大黄鱼养殖基地,从生长状态良好、体长 20~24 cm 的 2 年龄性成熟大黄鱼养殖群体中随机选取雄鱼和雌鱼各 3 尾。分别剖取部分精巢和卵巢,用 RNA 保护液固定后保存于 -80℃ 超低温冰箱。使用 RNA 试剂盒(TransZol Up Plus RNA Kit, 全式金生物有限公司)对所取性腺提取总 RNA, DNase I 消化去除可能混有的 DNA。检测合格的样品,用 Epicentre Ribo-zero™ rRNA Removal Kit (Epicentre, USA) 去除 rRNA 后,构建链特异性转录组文库,使用 Illumina HiSeq4000 平台测序(PE150)。

1.2 转录组的比对组装、注释

使用 FastQC v0.11.7^[34]查看测序后的数据质量,随后用 Trimmomatic v0.36^[35]对测序数据进行过滤:去除接头,截去连续 4 个碱基平均质量值低于 20 的部分,舍去长度短于 50 bp 的 reads (参数: ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10SLIDINGWINDOW:4:20 MINLEN: 50)。

使用 HISAT2 v2.1.0^[36]将过滤后的测序数据比对到本实验室利用三代测序技术组装的大黄鱼基因组(未发表)。对得到的 SAM 文件,使用 samtools v1.4^[37]排序并转换为 BAM 文件。随后使用 StringTie v1.3.4d^[38]进行拼接、组装。用 gffread^[39]得到组装后的转录本序列。

选取每个基因最长的转录本,使用 egg-

mapper v1.0.3^[40]的 diamond 模式将序列比对到 eggnog 蛋白数据库 v4.5.1^[41]进行基因功能注释。

1.3 差异表达基因分析

上述组装出的转录本,使用 Python 脚本 prepDE.py 得到基因的 reads 计数矩阵。随后,使用 DESeq2^[42]进行基因层面的差异分析。由于大黄鱼卵巢与精巢中的基因表达差异巨大,因此执行较严格的筛选条件 $\text{padj} < 0.01$, $\log_2(\text{fold change})$ 绝对值大于 2。选取前 300 个差异表达基因对两组雌雄样品进行主成分分析(PCA),鉴别组内样本的重复性以及组间样品的差异水平。

1.4 候选 lncRNA 的筛选

将拼接得到的转录本使用 eggog-mapper 的 diamond 模式比对到 eggnog 的蛋白数据库,无法比对到蛋白数据库的转录本用 CPC2^[43]和 CNCI^[44]进行编码潜能的分析,同时用 pfam_scan.pl 比对 Pfam32.0^[45]数据库。最后将 CPC2、CNCI 预测得到的没有编码能力的转录本,与未比对到 Pfam32.0 数据库的转录本取交集,再将其比对到 rfam^[46]数据库,去除 tRNA、rRNA,得到最后的候选 lncRNA 集。

1.5 lncRNA 的靶基因预测

在基因组序列上,取所有差异 lncRNA 上下游 50 kb 的所有编码基因,分别计算该 lncRNA 与此区域中各编码基因的皮尔森相关系数(Pearson correlation coefficient),若某编码基因与 lncRNA 的相关系数大于 0.8, $P < 0.05$,则认为该编码基因是受该 lncRNA 调控的靶基因。

2 结果与分析

2.1 测序数据

6 尾大黄鱼卵巢与精巢组织测序共得到 282457758 条 reads,每个个体的测序量为 13.38~15.32 Gb,平均 14.24 Gb。去除接头序列,截去连续 4 个碱基平均质量值低于 20 的部分,舍去长度少于 50 bp 的 reads。共得到 259935783 条 reads。过滤后每个个体的测序量为 12.36~14.07 Gb,平均 12.98 Gb。每个个体的测序量为每 4 个碱基的平均质量值均大于 20。每个个体的测序深度为 17.23 X~19.62 X,平均测序深度为 18.10 X。

2.2 基因注释

使用 HISAT2 将过滤后的测序数据比对到本实验室组装的大黄鱼三代参考基因组(未发表),平均比对率为 91.42%(89.78%~93.60%),平均的唯一比对率为 74.14%(70.60%~78.03%),通过 StringTie 拼接得到来自 31675 个基因的 66088 个转录本。26277 个基因可用 eggong-mapper 比对到蛋白家族数据库,其中有 Gene symbol 的共有 13928 (43.97%)个。

2.3 lncRNA 筛选

将未能比对到蛋白数据库的转录本,首先过滤掉长度小于 200 bp 的转录本。然后,分别使用 CPC2、CNCI 和 PfamScan 3 个软件进行编码能力的预测分析,得到无编码能力的转录本数目分别是 5625、5841 和 6957 条,将 3 个软件得到的数据取交集得到 5251 条 lncRNA (图 1)。然后通过比对到 rfam 数据库去除 tRNA 和 rRNA,最终共得到来自 3984 个基因的 5162 条 lncRNA。

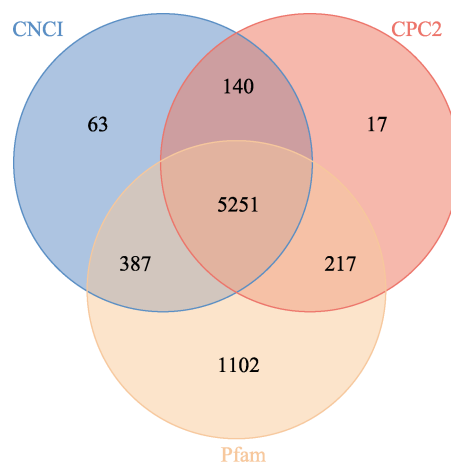


图 1 CPC2、CNCI、Pfam 预测 lncRNA 的韦恩图

Fig. 1 Venn diagram of lncRNAs predicted by CPC2, CNCI and Pfam

2.4 差异表达的 mRNA 和 lncRNA

通过 DESeq2 对数据进行基因层次的处理和统计分析。根据 $\text{FDR} < 0.01$ 以及 $\text{Fold Change} > 2$ 的标准,共发现卵巢与精巢中差异 mRNA 有 9341 个,其中卵巢相对于精巢表现上调的 mRNA 有 3184 个,下调的 mRNA 有 6157 个。在卵巢中特异性表达的 mRNA 有 22 个,在精巢中特异性表达的 mRNA 有 1109 个。

根据 $FDR < 0.01$ 以及 $Fold\ Change > 2$ 的标准, 共发现卵巢与精巢中差异 lncRNA 为 2782 个, 其中卵巢相对于精巢上调的 lncRNA 492 个, 下调的 lncRNA 2290 个。在卵巢中特异性表达的 lncRNA 有 61 个, 在精巢中特异性表达的 lncRNA 有 803 个。

主成分分析显示, 雌雄两组间的差异显著, 而在组内差异性较小。说明组内样本的重复性良好(图 2)。

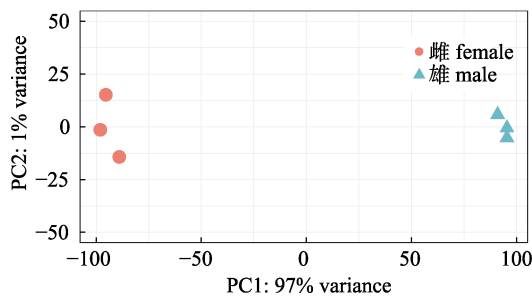


图 2 性腺 lncRNA 表达量主成分分析
Fig. 2 Principle component analysis of gonadal lncRNA expression

2.5 差异表达 mRNA 的富集分析

对上述得到的差异表达 mRNA 进行 GO 富集分析, 其结果如图 2 所示。在 biological process (生物学过程) 方面, 富集结果最多的有 reproduction

(繁殖), 富集到了 137 个差异的 mRNA。在 cellular component (细胞组成) 方面, 富集结果最多的是 cytoskeleton (细胞骨架), 富集到了 144 个差异的 mRNA。在 molecular function (分子功能) 方面, 富集结果最多的是 cytoskeletal protein binding (细胞骨架绑定), 富集到了 58 个差异的 mRNA。

在 reproduction (繁殖) 的 GO 条目中的基因大多与生物繁殖或细胞分裂相关, 精巢相对于卵巢表达上调的基因有 *piwil1*、*cetn1*、*hormad2* 等共 93 个, 卵巢相对于精巢表达上调的基因有 *bmp15*、*zp3*、*nanos3* 等 44 个。

2.6 lncRNA 的靶基因分析

以距离及表达量的相关性作为筛选条件, 经过计算共得到 892 个 lncRNA 与 865 个 mRNA 组成的 1227 对 lncRNA-mRNA。其中有 44 条 lncRNA 靶向调控了 reproduction 条目的 25 个基因, 共组成 45 对 lncRNA-mRNA, 其中包括与精子的发育密切相关的基因 *piwil1*、*cetn1*、*hormad2*, 以及与多种动物性别决定密切相关的基因 *dmrt1* (表 1)。

lncRNA MSTRG.24346 位于大黄鱼 *dmrt1* 基因下游 16.2 kb 的负链上(图 3), 只在精巢中表达, 而在卵巢中不表达。

表 1 部分性别相关 lncRNA 及其靶基因的位置信息、表达量及相关系数
Tab. 1 Location, expression and correlation coefficient of partial gender-based lncRNA and its target genes

$n = 3; \bar{x} \pm SD$

LncRNA 染色体位置 location of lncRNA	编号 number	卵巢 表达量 expression of ovary	精巢 表达量 expression of testis	靶基因 染色体位置 location of target gene	靶基因名 gene symbol of target gene	卵巢 表达量 expression of ovary	精巢 表达量 expression of testis	相关系数 correlation coefficient	P
chr14: 6644569-6658563	MSTRG.6100	0.037± 0.014	3.772± 1.128	chr14: 6599490-6601718	<i>cetn1</i>	0.196± 0.042	18.078± 1.093	0.925	0.008
chr14: 6644448-6645188	MSTRG.6099	0.002± 0.004	0.201± 0.068	chr14: 6599490-6601718	<i>cetn1</i>	0.196± 0.042	18.078± 1.093	0.928	0.008
chr12: 27688799-27690572	MSTRG.4543	0.071± 0.047	0.759± 0.264	chr12: 27645741-27661879	<i>hormad2</i>	0.950± 0.085	27.952± 2.134	0.908	0.012
chr12: 27661937-27672151	MSTRG.4561	0.067± 0.069	19.583± 7.167	chr12: 27645741-27661879	<i>hormad2</i>	0.950± 0.085	27.952± 2.134	0.941	0.005
chr9: 25856698-25938693	MSTRG.25008	9.853± 2.036	0.268± 0.097	chr9: 25942376-25950206	<i>piwil1</i>	8.471± 7.302	115.021± 27.862	-0.941	0.005
chr9: 7055233-7116761	MSTRG.24346	0.000± 0.000	1.723± 1.156	chr9: 7032048-7038995	<i>dmrt1</i>	0.000± 0.000	18.610± 9.610	0.992	8.48×10 ⁻⁵

注: 卵巢表达量和精巢表达量的值为 FPKM 值。
Note: Expression level of ovary and testis are FPKM.

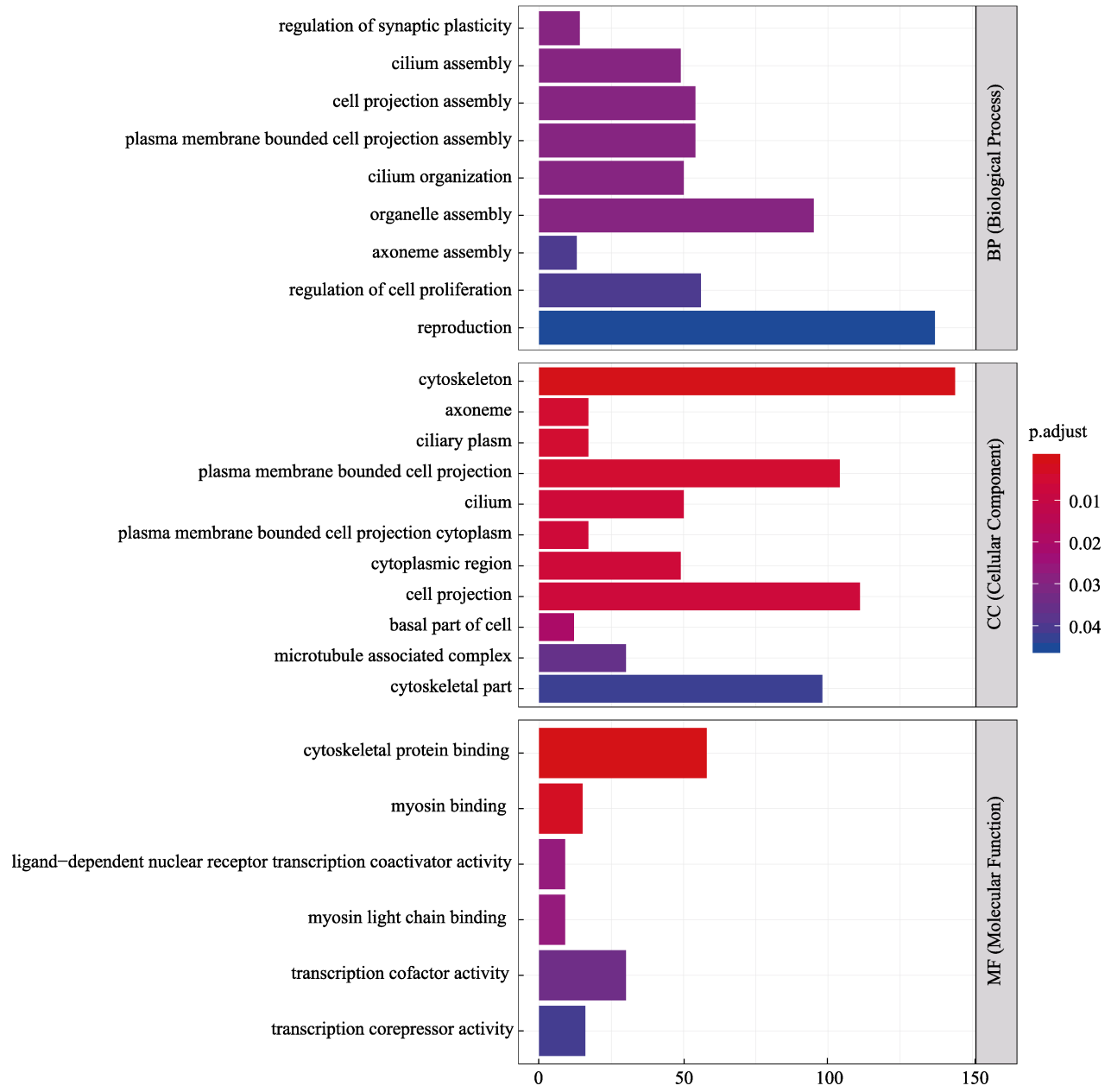


图 3 差异表达基因的 GO 富集

Fig. 3 Gene ontology (GO) functional annotation of differentially expressed genes

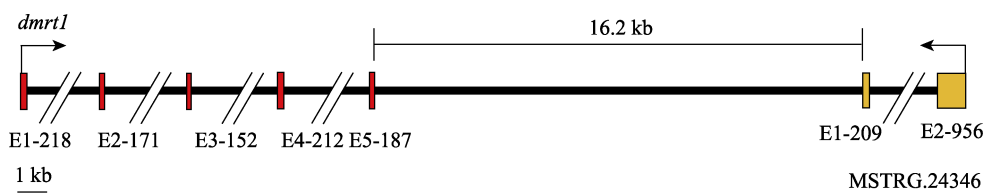


图 4 *dmrt1* 与 lncRNA MSTRG.24346 位置示意图

方框表示外显子，外显子编号及长度标注在其下侧，双斜线表示内含子。

Fig. 4 The location diagram of *dmrt1* and lncRNA MSTRG.24346

Exons are represented by boxes, while the numbers given below it represent the relative base pair sizes of exons.

Introns are represented by double slant.

3 讨论

在其他物种的研究中, 已发现 lncRNA 与性别分化、性腺发育、配子发生等过程密切相关。在黑龙江鲟精巢与卵巢的 lncRNA 研究中发现, 多条 lncRNA 靶向多个与性腺发育和配子发生相关的基因^[22]。在中华鳖的精巢与卵巢的 lncRNA 研究中, 发现多条 lncRNA 的靶基因中包括 *dmrt1*、*sox9*、*cyp19a*、*sox3* 等与性别分化相关的基因^[23]。通过基因敲除技术, 证实了某些 lncRNA 对果蝇的精子发育以及维持生育能力是不可或缺的^[30]。

Avasthi 等^[47]报道, 在雌性小鼠中敲除 *cetn1* 后雌性小鼠仍然具有正常生育能力, 但是在雄性小鼠中敲除 *cetn1* 基因后, 虽然该小鼠的睾丸大小正常, 精原细胞和精母细胞也发育正常, 但是精子数量却十分稀少, 为数不多的精子大多畸形, 失去了正常生育能力。这说明 *cetn1* 基因在雄性小鼠的精子发育晚期发挥着重要作用。Kogo 等^[48]的研究表明, *hormad2* 是雄性小鼠的精子发生过程中必需的基因, 雄性小鼠敲除 *hormad2* 基因后, 其精母细胞凋亡, 睾丸和附睾中检测不到精子的存在, 而敲除雌性小鼠的 *hormad2* 基因则不影响其生育能力。Guo 等^[49]研究发现在无精子症患者中 *piwi(piwil1)* 突变, 并通过构建 *piwi(piwil1)* 突变的小鼠, 证明了这种突变可以导致雄性小鼠不育, 从而说明 *piwi(piwil1)* 的突变是人类不育的一个因素。本研究通过对大黄鱼精巢和卵巢 RNA 测序数据进行分析, 共挖掘到了 3984 个 lncRNA, 并预测得到了 1227 对 lncRNA-mRNA; 在 lncRNA 调控的基因中包括 *cetn1*、*hormad2*、*piwil1* 3 个精子发生和发育所必需的基因, 其中, lncRNA MSTRG.6099、MSTRG.6100 靶向 *cetn1*, lncRNA MSTRG.4543、MSTRG.4561 靶向 *hormad2*, lncRNA MSTRG.25008 靶向 *piwil1*。这些 lncRNA 与其靶基因的皮尔森相关系数的绝对值都大于 0.9, *P* 值均小于 0.05。MSTRG.25008 与其靶基因 *piwil1* 之间呈现负相关, 其余的 lncRNA 与其靶基因都呈现高度的正相关。说明大黄鱼性腺发育(至少是精子发生和发育)过程可能也受到 lncRNA 的调控。

除此之外, 本研究还在大黄鱼精巢的转录组

中发现一条 lncRNA (lncRNA MSTRG.24346), 位于大黄鱼的 *dmrt1* 下游 16.2 kb 的负链上, 表达模式与 *dmrt1* 呈极显著的正相关, 皮尔森相关系数达 0.992 ($P < 0.0001$)。本实验室林爱强^[33]采用荧光定量 PCR 技术对大黄鱼 *dmrt1* 的时空表达模式进行了研究, 发现 *dmrt1* 基因只在雄鱼及伪雄鱼精巢中特异性表达, 在其他组织及雌鱼各组织(包括卵巢)中都检测不到其表达。鉴于 *dmrt1* 在多个物种的性别决定起重要作用^[26-28], 并且本实验室通过 QTL 定位找到的性别决定区域中包含有 *dmrt1*, 可以推测 *dmrt1* 是大黄鱼的性别决定候选基因。本研究在精巢转录组中发现 lncRNA MSTRG.24346 的靶基因为 *dmrt1*, 两者之间表达模式完全一致, 说明 lncRNA 可能在大黄鱼的性别决定中也发挥着一定的作用。

综上所述, 本研究在大黄鱼性腺中挖掘到了 3984 个基因位点的 5162 条 lncRNA, 2782 个 lncRNA 在精巢与卵巢中差异表达, 高度相关的 lncRNA-mRNA 对 1227, 并在精巢中发现了多条靶向 *dmrt1*、*cetn1*、*hormad2*、*piwil1* 等与雄性性别决定和性腺发育相关基因的 lncRNA, 说明大黄鱼的性别决定及性腺发育和分化也受到 lncRNA 的调控。本研究为大黄鱼雌雄性别决定与分化机制研究提供了一个新的研究方向, 为最终研究阐明其分子调控机理提供了重要的基础资料。

参考文献:

- [1] Spizzo R, Almeida M I, Colombatti A, et al. Long non-coding RNAs and cancer: A new frontier of translational research?[J]. *Oncogene*, 2012, 31(43): 4577-4587.
- [2] Lee J T. The X as model for RNA's niche in epigenomic regulation[J]. *Cold Spring Harbor Perspectives in Biology*, 2010, 2(9): a003749.
- [3] Wutz A, Rasmussen T P, Jaenisch R. Chromosomal silencing and localization are mediated by different domains of Xist RNA[J]. *Nature Genetics*, 2002, 30(2): 167-174.
- [4] Zhou Z Y, Li A M, Adeola A C, et al. Genome-wide identification of long intergenic noncoding RNA genes and their potential association with domestication in pigs[J]. *Genome Biology and Evolution*, 2014, 6(6): 1387-1392.
- [5] Gong C G, Maquat L E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu

- elements[J]. *Nature*, 2011, 470(7333): 284-288.
- [6] Clemson C M, Hutchinson J N, Sara S A, et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles[J]. *Molecular Cell*, 2009, 33(6): 717-726.
- [7] Gupta R A, Shah N, Wang K C, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis[J]. *Nature*, 2010, 464(7291): 1071-1076.
- [8] Rinn J L, Chang H Y. Genome regulation by long noncoding RNAs[J]. *Annual Review of Biochemistry*, 2012, 81: 145-166.
- [9] Derrien T, Johnson R, Bussotti G, et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression[J]. *Genome Research*, 2012, 22(9): 1775-1789.
- [10] Moran V A, Perera R J, Khalil A M. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs[J]. *Nucleic Acids Research*, 2012, 40(14): 6391-6400.
- [11] Hung T, Wang Y L, Lin M F, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters[J]. *Nature Genetics*, 2011, 43(7): 621-629.
- [12] Kino T, Hurt D E, Ichijo T, et al. Noncoding RNA Gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor[J]. *Science Signaling*, 2010, 3(107): ra8.
- [13] Poliseno L, Salmena L, Zhang J W, et al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology[J]. *Nature*, 2010, 465(7301): 1033-1038.
- [14] Salmena L, Poliseno L, Tay Y, et al. A ceRNA hypothesis: The Rosetta Stone of a hidden RNA language?[J]. *Cell*, 2011, 146(3): 353-358.
- [15] Heo J B, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA[J]. *Science*, 2011, 331(6013): 76-79.
- [16] Tsai M C, Manor O, Wan Y, et al. Long noncoding RNA as modular scaffold of histone modification complexes[J]. *Science*, 2010, 329(5992): 689-693.
- [17] Zhang Y C, Liao J Y, Li Z Y, et al. Genome-wide screening and functional analysis identify a large number of long non-coding RNAs involved in the sexual reproduction of rice[J]. *Genome Biology*, 2014, 15(12): 512.
- [18] Taylor D H, Chu E T, Spektor R, et al. Long non-coding RNA regulation of reproduction and development[J]. *Molecular Reproduction and Development*, 2015, 82(12): 932-956.
- [19] van Werven F J, Neuert G, Hendrick N, et al. Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast[J]. *Cell*, 2012, 150(6): 1170-1181.
- [20] Herrera L, Ottolenghi C, Garcia-Ortiz J E, et al. Mouse ovary developmental RNA and protein markers from gene expression profiling[J]. *Developmental Biology*, 2005, 279(2): 271-290.
- [21] Laiho A, Kotaja N, Gyenesei A, et al. Transcriptome profiling of the murine testis during the first wave of spermatogenesis[J]. *PLoS ONE*, 2013, 8(4): e61558.
- [22] Zhang X, Shi J L, Sun Y L, et al. Transcriptome analysis provides insights into differentially expressed genes and long noncoding RNAs involved in sex-related differences in Amur sturgeon (*Acipenser schrenckii*)[J]. *Molecular Reproduction and Development*, 2019, 86(2): 132-144.
- [23] Zhang J, Yu P, Zhou Q Y, et al. Screening and characterisation of sex differentiation-related long non-coding RNAs in Chinese soft-shell turtle (*Pelodiscus sinensis*)[J]. *Scientific Reports*, 2018, 8: 8630.
- [24] Sun J, Lin Y, Wu J. Long non-coding RNA expression profiling of mouse testis during postnatal development[J]. *PLoS ONE*, 2013, 8(10): e75750.
- [25] Raymond C S, Shamu C E, Shen M M, et al. Evidence for evolutionary conservation of sex-determining genes[J]. *Nature*, 1998, 391(6668): 691-695.
- [26] Matsuda M, Nagahama Y, Shinomiya A, et al. DMY is a Y-specific DM-domain gene required for male development in the medaka fish[J]. *Nature*, 2002, 417(6888): 559-563.
- [27] Chen S L, Zhang G J, Shao C W, et al. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle[J]. *Nature Genetics*, 2014, 46(3): 253-260.
- [28] Cui Z K, Liu Y, Wang W W, et al. Genome editing reveals dmrt1 as an essential male sex-determining gene in Chinese tongue sole (*Cynoglossus semilaevis*)[J]. *Scientific Reports*, 2017, 7: 42213.
- [29] Zhang L, Lu H, Xin D Z, et al. A novel ncRNA gene from mouse chromosome 5 trans-splices with Dmrt1 on chromosome 19[J]. *Biochemical and Biophysical Research Communications*, 2010, 400(4): 696-700.
- [30] Wen K J, Yang L J, Xiong T L, et al. Critical roles of long noncoding RNAs in *Drosophila* spermatogenesis[J]. *Genome Research*, 2016, 26(9): 1233-1244.
- [31] Wang Z Y, Cai M Y, Xie Y J. Female nuclear development and sex control techniques in large yellow croaker[M]//Chen S L. *Fish Sex Control and Cell Engineering Breeding*. Beijing: Science Press, 2013: 274-296. [王志勇, 蔡明夷, 谢仰杰. 大黄鱼雌核发育和性别控制技术[M]//陈松林. *鱼类性别控制与细胞工程育种*. 北京: 科学出版社, 2013:

- 274-296]
- [32] Wang Z Y, Cai M Y. Artificial gynogenesis and sex control in large yellow croaker[M]//Wang H P, Piferrer F, Chen S L, et al. Sex Control in Aquaculture. Hoboken: John Wiley & Sons, 2018: 751-773.
- [33] Lin A Q. Study of sex-specific molecular markers and some sexrelated genes in large yellow croaker (*Larimichthys crocea*)[D]. Xiamen: Jimei University, 2017. [林爱强. 大黄鱼性别特异分子标记及部分性别相关基因的初步研究[D]. 厦门: 集美大学, 2017.]
- [34] Andrews S. Fast QC: A quality control tool for high throughput sequence data[CP/OL]. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- [35] Bolger A M, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data[J]. Bioinformatics, 2014, 30(15): 2114-2120.
- [36] Kim D, Langmead B, Salzberg S L. HISAT: A fast spliced aligner with low memory requirements[J]. Nature Methods, 2015, 12(4): 357-360.
- [37] Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools[J]. Bioinformatics, 2009, 25(16): 2078-2079.
- [38] Pertea M, Pertea G M, Antonescu C M, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads[J]. Nature Biotechnology, 2015, 33(3): 290-295.
- [39] Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks[J]. Nature Protocols, 2012, 7(3): 562-578.
- [40] Huerta-Cepas J, Forslund K, Coelho L P, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper[J]. Molecular Biology and Evolution, 2017, 34(8): 2115-2122.
- [41] Huerta-Cepas J, Szklarczyk D, Forslund K, et al. eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences[J]. Nucleic Acids Research, 2016, 44(D1): D286-D293.
- [42] Love M I, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2[J]. Genome Biology, 2014, 15(12): 550.
- [43] Kang Y J, Yang D C, Kong L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features[J]. Nucleic Acids Research, 2017, 45(W1): W12-W16.
- [44] Sun L, Luo H T, Bu D C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts[J]. Nucleic Acids Research, 2013, 41(17): e166.
- [45] El-Gebali S, Mistry J, Bateman A, et al. The Pfam protein families database in 2019[J]. Nucleic Acids Research, 2019, 47(D1): D427-D432.
- [46] Kalvari I, Argasinska J, Quinones-Olvera N, et al. Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families[J]. Nucleic Acids Research, 2018, 46(D1): D335-D342.
- [47] Avasthi P, Scheel J F, Ying G, et al. Germline deletion of Ctn1 causes infertility in male mice[J]. Journal of Cell Science, 2013, 126(14): 3204-3213.
- [48] Kogo H, Tsutsumi M, Inagaki H, et al. HORMAD2 is essential for synapsis surveillance during meiotic prophase via the recruitment of ATR activity[J]. Genes to Cells, 2012, 17(11): 897-912.
- [49] Gou L T, Kang J Y, Dai P, et al. Ubiquitination-deficient mutations in human *Piwi* cause male infertility by impairing histone-to-protamine exchange during spermiogenesis[J]. Cell, 2017, 169(6): 1090-1104.

The identification and analysis of long noncoding RNA in testes and ovaries of the large yellow croaker (*Larimichthys crocea*)

WANG Weijia^{1,2}, HAN Zhaofang¹, LI Wanbo², WANG Zhiyong¹

1. Key Laboratory of Healthy Mariculture for the East China Sea of Ministry of Agriculture and Rural Affairs, Jimei University, Xiamen 361021, China;
2. Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao 266071, China

Abstract: To explore the functions of long non-coding RNA in the differentiation of gonads, we performed next-generation sequencing on strand-specific RNA libraries where rRNA was depleted from total RNA extracted from three testes and three ovaries of the large yellow croaker (*Larimichthys crocea*). In total, 66088 transcripts derived from 31675 genes were obtained from the sequencing data of the large yellow croaker; 5162 lncRNAs from 3984 gene loci were retained after stringent filtering. There were 9341 differentially expressed mRNAs and 2782 lncRNAs between the male and female gonads at the gene level. Furthermore, Pearson correlation coefficients were calculated between the differentially expressed lncRNAs and their adjacent genes using the expression data. We detected 1227 highly correlated lncRNA-mRNA pairs, among which, multiple lncRNAs targeted genes related to sex differentiation. lncRNA MSTRG.24346 was close to (<17 kb) and highly significantly correlated with the sex-differentiation candidate gene *dmrt1* in the large yellow croaker. Thus, this study showed that lncRNA may play important roles in the sex differentiation of the large yellow croaker.

Key words: *Larimichthys crocea*; gonad; lncRNA; sex differentiation

Corresponding author: WANG Zhiyong. E-mail: zywang@jmu.edu.cn