

文章编号: 1674-5566(2020)02-0161-10

DOI: 10.12024/jsou.20190502645

基于转录组测序的半滑舌鳎长链非编码 RNA 的初步鉴定与分析

徐 浩^{1,2,3}, 徐锡文^{1,3}, 周 茜^{1,3}, 陈松林^{1,3}

(1. 中国水产科学研究院黄海水产研究所 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, 山东 青岛 266071; 2. 上海海洋大学 水产与生命学院, 上海 201306; 3. 农业农村部海洋渔业可持续发展重点实验室, 山东 青岛 266071)

摘 要: 以抗病家系与易感家系半滑舌鳎为材料, 进行哈维弧菌感染实验, 并对易感家系感染前(CsSU)、易感家系感染后(CsSC)、抗病家系感染前(CsRU)、抗病家系感染后(CsRC)4组进行转录组测序, 根据 RNA-seq 数据挖掘半滑舌鳎长链非编码 RNA 信息; 通过生物信息学分析, 筛选出与抗哈维弧菌病相关的差异长链非编码 RNA(long non-coding RNA, lncRNA)。结果显示, 共识别出 4 584 个 lncRNA 座位, 包含 5 714 个转录本; 其基本特征与编码基因的比较分析, lncRNA 的 GC 含量低于编码基因, 单外显子基因数多于编码基因, 转录本的平均长度长于编码基因, 基因表达量低于编码基因。对 4 组样品进行两两比较(CsRU vs CsSU、CsRC vs CsSC、CsRC vs CsRU、CsSC vs CsSU)分别筛选出 818、813、261、140 个差异表达 lncRNA, 其中 CsRU 与 CsSU 之间、CsRC 与 CsSC 之间 lncRNA 数目差异最多, 通过聚类分析确定了各实验组的表达模式之间的联系, CsRU 与 CsSU 之间的表达模式最为相近。通过共表达分析, 预测出 lncRNA 和 274 个编码基因可能存在 14 539 种相互关系, 并进行了功能注释, 进而筛选出 7 个关键 lncRNA。qRT-PCR 结果显示, 差异表达 lncRNAs 的表达模式和转录组数据得到的基本一致。研究结果为揭示 lncRNA 在半滑舌鳎抗哈维弧菌免疫调控反应中的作用提供重要的参考数据。

关键词: 半滑舌鳎; 哈维弧菌; 长链非编码 RNA; 免疫调控; 差异表达

中图分类号: S 917 **文献标志码:** A

非编码 RNA(non-coding RNA, ncRNA)是由 DNA 转录后但不翻译成蛋白质的 RNA, 主要包括短链非编码 RNA 和长链非编码 RNA。其中, 长链非编码 RNA(long non-coding RNA, lncRNA)长度在几百个碱基左右, 和 mRNA 结构相似, 转录过程中有着不同的剪接方式, 部分 lncRNA 作为短链非编码 RNA 的前体 RNA, lncRNA 广泛存在于真核生物中, 并具有组织特异性^[1]。

近年来, 在哺乳动物和鱼类中发现了越来越多的 lncRNA, 证实了其在表观遗传修饰、转录调控、生长发育、免疫、细胞分化、疾病发生等生物过程中发挥重要作用。在表观遗传学方面, lncRNA 作为支架分子结合表观遗传因子调控生物学过程^[2]; 在转录水平上, lncRNA 作为诱饵分

子抑制或促进启动子或者增强子的转录^[3]; 在转录后水平上, lncRNA 作为调控因子影响 mRNA 稳定性和翻译效率^[4]。因此, lncRNA 参与了细胞中多种生物学过程的调控, 在生物体生长发育和免疫等方面发挥了极其重要的作用。随着高通量测序技术的发展, 对 lncRNA 的研究越来越深入, 为全面分析转录组和研究基因功能调控提供了有力条件。

半滑舌鳎(*Cynoglossus semilaevis*)是我国重要的名贵海水养殖鱼类^[5]。然而近年来, 病害问题成为制约半滑舌鳎养殖业发展的重要因素, 其中弧菌病传播性强、致死率高, 对半滑舌鳎养殖业的影响尤为严重。哈维弧菌(*Vibrio harveyi*)是引起半滑舌鳎腐皮、烂尾等疾病的主要病原菌之

收稿日期: 2019-05-13 修回日期: 2019-10-18

基金项目: 国家自然科学基金(31530078); 国家海水鱼产业技术体系(CARS47G03); 青岛海洋科学与技术国家实验室资助的鳌山人才培养计划(2017ASTCP-OS15)

作者简介: 徐 浩(1989—), 男, 博士研究生, 研究方向为半滑舌鳎遗传育种。E-mail: zwzxyey4512688@126.com

通信作者: 陈松林, E-mail: chensl@ysfri.ac.cn

—^[6]。目前,通常使用广谱抗菌类药物处理病鱼,但是抗菌类药物的使用对养殖水环境和半滑舌鲷本身会造成若干不良后果。因此,有必要利用现代生物技术手段了解半滑舌鲷抗哈维弧菌病的免疫调控机理。本研究利用 Illumina 测序技术进行抗病家系和易感家系半滑舌鲷感染哈维弧菌前后转录组测序,获得大量的 lncRNA 数据,并通过比较分析,筛选出与哈维弧菌抗病相关的 lncRNA。

1 材料与方法

1.1 原始数据

所用数据来源于本实验室半滑舌鲷转录组测序数据^[7],目前已上传至 SRA 数据库(NCBI 登录号:SRP186770),见表 1。该数据以抗病家系(14L003 家系)与易感家系(14L013 家系)半滑舌

鲷为材料(图 1),在未感染哈维弧菌时,分别对抗病家系与易感家系半滑舌鲷各 3 尾取样测序;在感染哈维弧菌 24 h 时,再分别对抗病家系与易感家系半滑舌鲷各 3 尾取样测序。

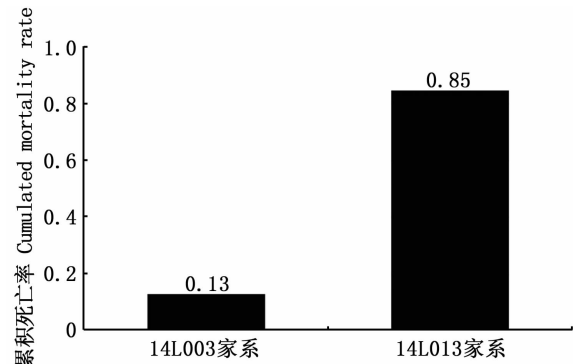


图 1 所用半滑舌鲷家系死亡率的比较

Fig. 1 Comparison of cumulated mortality rate in the family of Chinese tongue sole

表 1 12 个半滑舌鲷样本信息概览

Tab. 1 Summary statistics for twelve samples of *Cynoglossus semilaevis*

样品名 Sample name	物种名 Species	家系类别 Family	处理方法 Treatment	原始数据 Raw reads	组别 Group
SU1	半滑舌鲷	易感家系	未感染哈维弧菌	93 383 974	CsSU
SU2	半滑舌鲷	易感家系	未感染哈维弧菌	104 672 142	CsSU
SU3	半滑舌鲷	易感家系	未感染哈维弧菌	80 095 718	CsSU
RU1	半滑舌鲷	抗病家系	未感染哈维弧菌	85 660 884	CsRU
RU2	半滑舌鲷	抗病家系	未感染哈维弧菌	91 134 342	CsRU
RU3	半滑舌鲷	抗病家系	未感染哈维弧菌	91 226 452	CsRU
SC1	半滑舌鲷	易感家系	感染哈维弧菌 24h	101 900 430	CsSC
SC2	半滑舌鲷	易感家系	感染哈维弧菌 24h	104 216 082	CsSC
SC3	半滑舌鲷	易感家系	感染哈维弧菌 24h	100 320 038	CsSC
RC1	半滑舌鲷	抗病家系	感染哈维弧菌 24h	106 581 728	CsRC
RC2	半滑舌鲷	抗病家系	感染哈维弧菌 24h	105 878 506	CsRC
RC3	半滑舌鲷	抗病家系	感染哈维弧菌 24h	100 828 908	CsRC

1.2 实验方法

1.2.1 lncRNA 的识别

从转录组数据中识别 lncRNA,由图 2 可知具体步骤:使用 RNA-QC-Chain 软件^[8]对原始数据进行质量控制,去除含有 Poly-A、接头的序列以及低质量序列。使用 HISAT2 软件^[9]将测序数据分别比对到半滑舌鲷参考基因组(NCBI 登录号:GCF_000523025.1)上。使用 Cufflinks 软件^[10]基于半滑舌鲷基因组对转录组数据进行组装。并使用 TACO 软件^[11]对各组样品的转录本进行整

合。去除疑似 rRNA、tRNA、miRNA 前体的转录本。使用 Cuffcompare 软件和转录组注释文件相比较后,去除编码基因的转录本,将剩下长度大于 200 bp 的非编码基因转录本作为候选 lncRNA。使用 hmmscan 软件^[12]与 Pfam 数据库^[13]比对去除含有结构功能域的转录本。使用 FEELnc 软件^[14]预测候选 lncRNA 的编码能力,去除具有编码能力的转录本。使用 RSEM 软件^[15]对转录本进行定量,确定所有转录本的 FPKM 均小于 0.5。

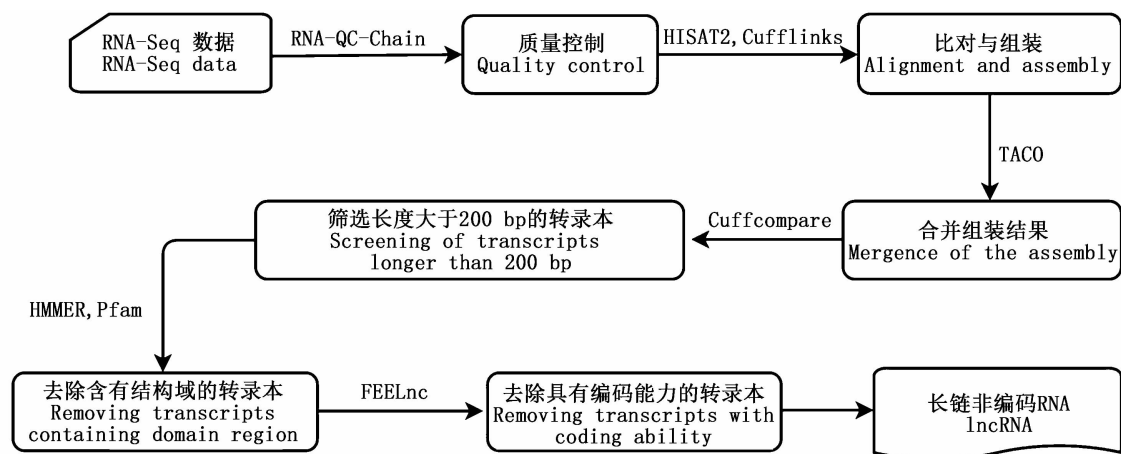


图 2 长链非编码 RNA 的识别流程

Fig. 2 The flow chart to identify lncRNA

1.2.2 不同实验组 lncRNA 表达谱 PCA 分析

为研究 4 组实验组中 3 个生物学重复样本间 lncRNA 的表达情况,使用 R 语言中的主成分分析方法(PCA)对 12 个样本的 lncRNA 表达量进行比较分析。

1.2.3 lncRNA 的差异表达分析

使用 HTSeq-count 软件^[16]统计读段数目,随后通过 DESeq2 软件^[17]检测两组之间的差异表达 lncRNA,以 adjusted P -value < 0.01 作为评价 lncRNA 表达差异统计学意义的阈值,使用 R 语言中的 Heatmaps 软件对 4 个实验组的所有差异表达基因进行系统分析,得到聚类结果。

1.2.4 qRT-PCR 验证差异表达 lncRNA

在 4 组比较中分别随机选择 6 个差异表达 lncRNA,使用 Geneious 软件设计引物(表 2),使用 qRT-PCR 验证差异表达 lncRNA。

采用 TaKaRa 荧光定量试剂盒,使用 ABI 7 500 fast real-time PCR 仪器,转录组测序的 cDNA 为模板,验证分析引物的扩增效率,并确定 cDNA 的稀释倍数。以半滑舌鳎 β -actin 为内参基因,qRT-PCR 的反应体系如下:10 μ L SYBR Primix Ex Taq,1 μ L 正向引物,1 μ L 反向引物,2 μ L cDNA,6 μ L 灭菌水,共 20 μ L。qPCR 采用两步法,反应条件:95 $^{\circ}$ C, 30 s;95 $^{\circ}$ C, 5 s,60 $^{\circ}$ C, 34 s,40 个循环;反应完成后直接进行熔解曲线分析。

根据 qPCR 得出的 Ct 值,采用相对定量的 $2^{-\Delta\Delta Ct}$ 法,分析出基因的相对表达量,由 Mean \pm SE 来表示。使用 Prism version 7.0a 统计软件对

数据进行样本显著性差异分析及作图。

1.2.5 半滑舌鳎 lncRNA 共表达分析

使用 4 个实验组识别出的 lncRNA 和已知的编码基因的表达水平分析 lncRNA 和编码基因的共表达。使用 perl 脚本计算 lncRNA 和编码基因的表达水平之间的 Pearson 相关系数。使用 DAVID 和 KOBAS 软件进行功能注释分析。

2 结果

2.1 半滑舌鳎 lncRNA 的识别

对半滑舌鳎的转录组数据全部分析获得 4 584 个 lncRNA 座位,包含 5 714 个转录本。对 lncRNA 和编码基因的基本特征进行比较分析,结果显示,lncRNA 的 GC 含量低于编码基因(图 3 a);lncRNA 的单外显子数目多于编码基因(图 3 b);lncRNA 转录本平均长度大于编码基因(图 3 c);lncRNA 的平均表达量低于编码基因(图 3 d)。

2.2 半滑舌鳎 lncRNA 表达水平 PCA 分析

使用 R 语言中的主成分分析方法(PCA)对半滑舌鳎的 4 个实验组(CsSU、CsSC、CsRU、CsRC)的 12 个样本(SU1、SU2、SU3、SC1、SC2、SC3、RU1、RU2、RU3、RC1、RC2、RC3)进行比较分析。如图 4 所示,两个主成分(PC1 和 PC2)方差之和为 98.1%,表明分析基本包含了所有样本的 lncRNA 数据。此外,12 个样本中感染哈维弧菌半滑舌鳎聚为一类,未感染哈维弧菌的半滑舌鳎聚为一类。

表 2 实验所需引物

Tab. 2 Primers used in this study

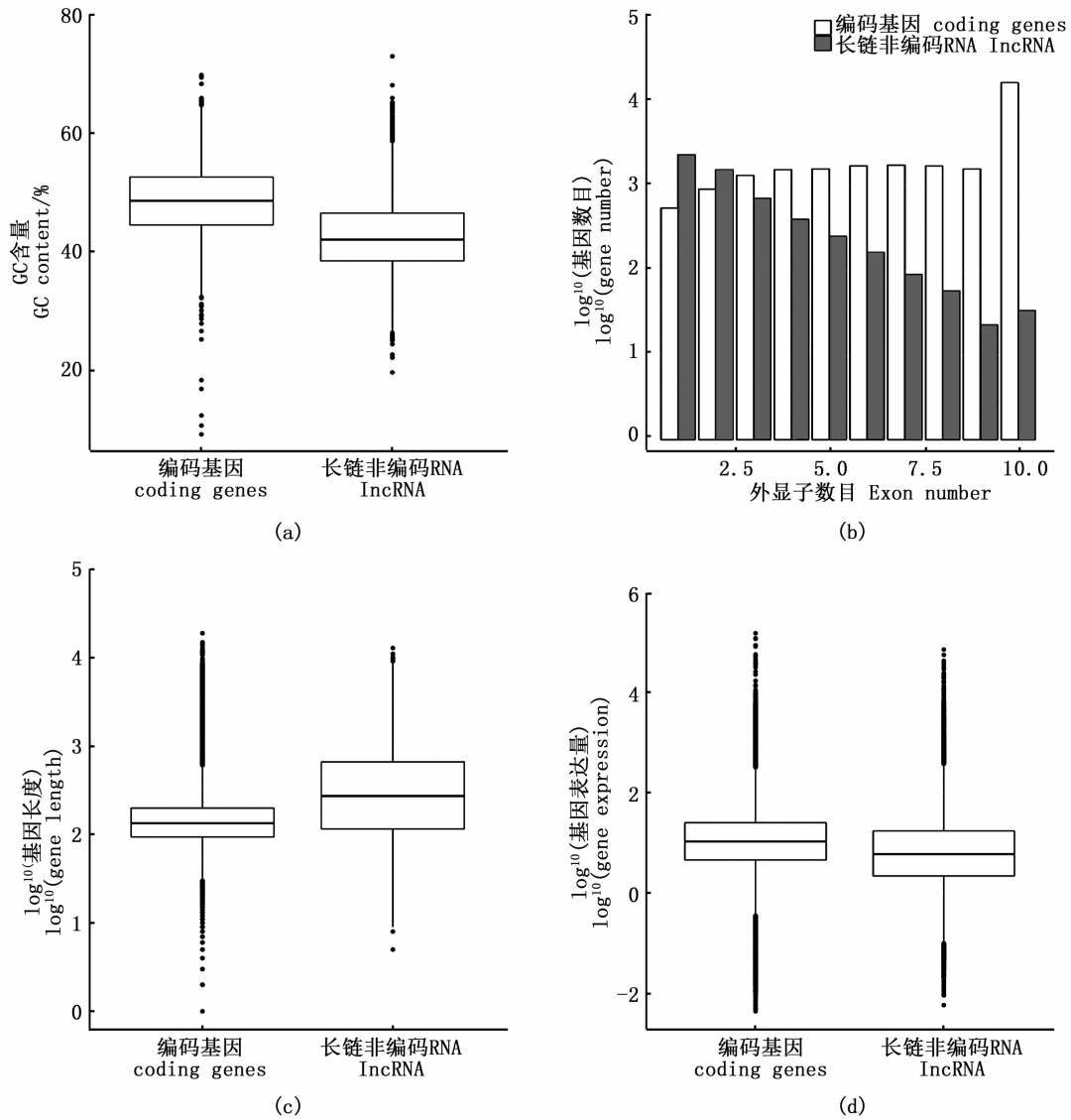
长链非编码 RNA lncRNA	正向引物 Forward Primer	反向引物 Reverse Primer
lnc_MSTRG. 3538. 10	TCAGAGTTTGAACAACCGCAACC	ATTTCCAGAGCAGACCTGAGACC
lnc_MSTRG. 398. 5	AAACAGTCGCACCAACTGAAGCTG	CTCCTCATGCCAGTTTCCTTCAGA
MSTRG. 1648. 1	TGGAGAGCAGTCTGATGGACATG	TTCATCTTGGCGGCAGAATGTTC
lnc_MSTRG. 5697. 1	AAAATCACCCGACGTTTCACAGT	GACGCTGCTGCACATGAATAGAC
lnc_MSTRG. 15174. 1	GTAGGGTGAAGAGCAGAGGACTG	CTCTGTAAGTGGGTGCTAGGCCTT
lnc_MSTRG. 2920. 9	GAGAGTGAAGTCCCTGCATCACT	ACACGACATCTGTGGGGAAGAAA
lnc_MSTRG. 17735. 2	CTGCTCATCTCACCTATGCTGT	AGACCTGCAGCATCCTAGGAAAG
lnc_MSTRG. 15378. 1	TCTGGGACTGGTGTACATGAAGC	CCCAGTCCAACATACATGTCCAT
lnc_MSTRG. 15001. 1	CCAATCAGGTCAGGCAGAGAGAA	CCCATACTCTGCATCACTCACA
lnc_MSTRG. 16516. 1	TGTCTCTGTGCCTCATGTCAAGA	CCACAACCTCTCATGTCAATGCC
lnc_MSTRG. 16915. 2	GAACCTTTTACCAGTGGAGCTG	GGCAAACCTAGTCCCTTTTCGATG
lnc_MSTRG. 17031. 1	TGGGCCTGGACTCCTTTTAAACA	ACATACTGCATCCTCGACGTGAA
lnc_MSTRG. 15376. 1	GGAATTTGCCGACGCATTTCAC	TAGAACTGCCGTACACACACACA
lnc_MSTRG. 15229. 1	GCATTGAGCATTGAGGCAGGATT	TCGATGTACTGAGGGAGCGAAAG
lnc_MSTRG. 15228. 1	TCCATGTACTGAGGGAGCGAAAG	GCATTGAGCATTGAGGCAGGATT
lnc_MSTRG. 9032. 1	GATCGGACTTCTCGAAGAAGCT	CGGCTAACGTAAGACTGTCTCGA
lnc_MSTRG. 9860. 1	AGGCACAAGGATATCAACAGCT	ACGAGAGAACAAGATCTGGGAGC
lnc_MSTRG. 3988. 1	CATTCTTCTCGGAGCTCCAGACA	TGAGATTTGGCTGACGGAGAGAG
lnc_MSTRG. 11713. 1	TTCCAAAACCCCTGGAGGAGTTT	CTGAGGTTCCCTCTGAGGTCCATG
lnc_MSTRG. 8260. 22	TCTGTGCCAAGTGTITAGGTCCA	CTGGACATGGTTCTTGTGTGGAAC
lnc_MSTRG. 8260. 21	CGACCACTGTGATGACAAACCTG	TGTCTGTCCCAACATGTTTGC
lnc_MSTRG. 191. 2	TCTGGTTGGAGAGCTGTGTGTT	TGATCTGGCTACTCCTCTGGCTA
lnc_MSTRG. 8275. 1	TGATGAGACACACCAACCTGT	GGACATCAATGGTCTGTGCCAAG
lnc_MSTRG. 8260. 7	TCTGTGCCAAGTGTITAGGTCCA	AAGACACCTGGACTTCCTTGGAC

2.3 差异表达 lncRNA 的筛选

使用 DEseq2 对 4 组比较 (CsRU vs CsSU, CsRC vs CsSC, CsRC vs CsRU, CsSC vs CsSU) 进行差异表达 lncRNA 进行筛选。如图 5 所示,在上述 4 组比较中,共筛选出 2 032 个差异表达 lncRNA,其中,CsRU 和 CsSU 的比较分析中,共筛选到 818 个差异表达 lncRNA,529 个 lncRNA 在 CsRU 的表达量显著高于 CsSU;在 CsRC 和 CsSC 的比较分析中,共筛选到 813 个差异表达 lncRNA,213 个 lncRNA 在 CsRC 的表达量显著高于 CsSC;在 CsRC 和 CsRU 的比较分析中,共筛选到 261 个差异表达 lncRNA,135 个 lncRNA 在 CsRC 的表达量显著高于 CsRU;在 CsSC 和 CsSU

的比较分析中,共筛选到 140 个差异表达 lncRNA,59 个 lncRNA 在 CsSC 的表达量显著高于 CsSU。

为评价差异表达 lncRNA 在不同实验条件下的表达模式,对 4 个实验组半滑舌鲷的 lncRNA 进行聚类分析(图 6 a),发现未感染哈维弧菌的抗病家系与易感家系半滑舌鲷的表达模式最为相似。同时,为进一步研究不同组比较的差异表达 lncRNA 之间的联系,我们利用 CsRU vs CsSU, CsRC vs CsSC, CsRC vs CsRU, CsSC vs CsSU 比较组中产生的差异表达 lncRNA 构建维恩图(图 6 b),结果显示没有筛选到在所有 4 组比较中均有的差异表达 lncRNAs。



(a). GC 含量比较; (b). 外显子数目比较; (c). 转录本长度比较; (d). 表达量比较
 (a). The comparison of GC content; (b). The comparison of exon number; (c). The comparison of transcript length; (d). The comparison of gene expression

图 3 lncRNA 和编码基因的比较分析

Fig. 3 The comparison of lncRNA genes and coding genes

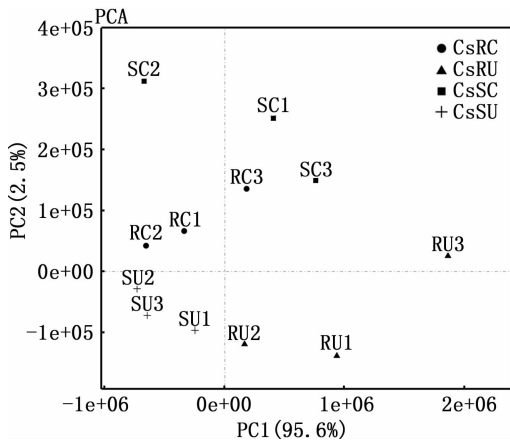


图 4 不同样品的 PCA 分析图
 Fig. 4 PCA analysis of different samples

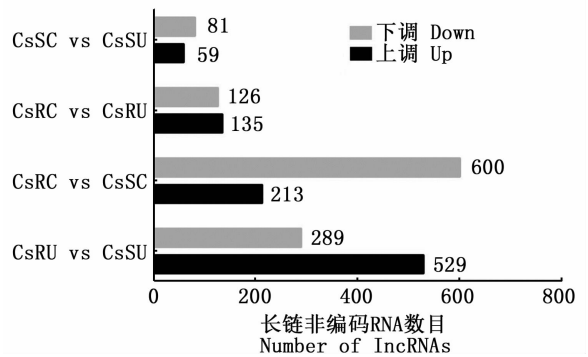
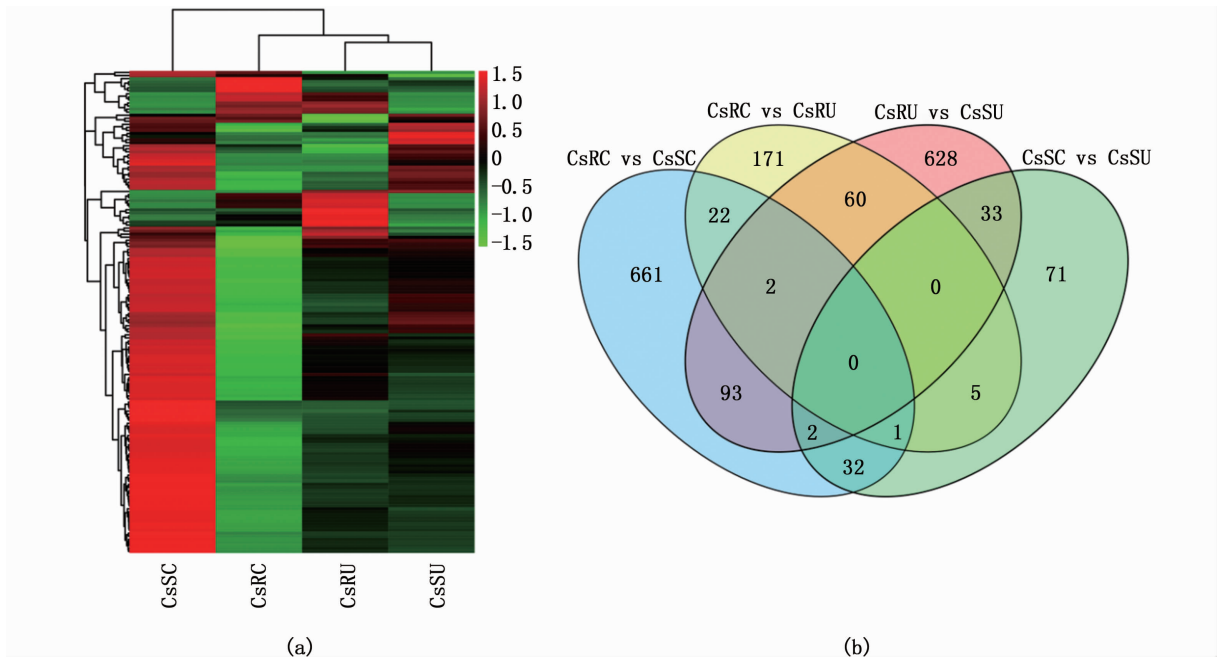


图 5 lncRNA 的统计情况
 Fig. 5 Numbers of upregulated and downregulated lncRNA



(a). 差异表达 lncRNA 整体 FPKM 层次聚类图; (b). 差异表达 lncRNA 维恩图
 (a). Hierarchical cluster analysis of differentially expressed lncRNA of the four groups; (b). Venn diagram of the number of differentially expressed lncRNA between contrasts

图 6 差异表达 lncRNA 的分析

Fig. 6 Analysis of differentially expressed lncRNA in the RNA-seq library comparisons

2.4 半滑舌鳎 lncRNA 共表达基因的功能注释

使用共表达分析和 ceRNA 理论预测了 lncRNA 在编码基因中的潜在靶基因。在 lncRNA 和编码基因之间共检测到 14 539 个相互作用关系(14 157 个正相关和 382 个负相关),其中对应半滑舌鳎参考基因组中 274 个编码基因。根据 lncRNA 对应的靶基因进行功能注释,结果(图 7)显示共表达基因共富集到 127 个 GO 条目,其中生物学过程(biological process)下 58 个,细胞组分(cellular component)下 33 个,分子功能(molecular function)下 36 个。值得注意的是,其中包含多个免疫相关的 GO 条目:免疫反应

(immune response, GO: 0006955)、刺激反应(response to stimulus, GO:0050896)、细胞因子活性(cytokine activity, GO:0005125)和趋化因子活性(chemokine activity, GO:0008009)等。同时根据 GO 注释结果筛选出 7 个关键 lncRNA(表 3),且其连接度均大于 14。

2.5 差异表达 lncRNA 的验证

使用 qPCR 的方法验证筛选到的差异表达 lncRNA 的表达模式。由图 8 可知,使用 qPCR 检测 24 个 lncRNA 差异表达水平的分析结果与 RNA-seq 的分析结果基本一致,说明筛选差异表达 lncRNA 的数据是真实可靠的。

表 3 哈维弧菌胁迫下半滑舌鳎的关键 lncRNA

Tab. 3 The key lncRNA of *Cynoglossus semilaevis* under *Vibrio harveyi* infection

长链非编码 RNA lncRNAs	连接度 Degree	GO 编号 GO ID	GO 条目名称 GO name
linc_MSTRG. 16107	18	GO:0006955	immune response
linc_MSTRG. 15347	14	GO:0006955	immune response
linc_MSTRG. 15118	121	GO:0006954	inflammatory response
linc_MSTRG. 15732	21	GO:0006954	inflammatory response
linc_MSTRG. 17133	99	GO:0006954	inflammatory response
linc_MSTRG. 17225	94	GO:0006954	inflammatory response
linc_MSTRG. 9953	130	GO:0006954	inflammatory response

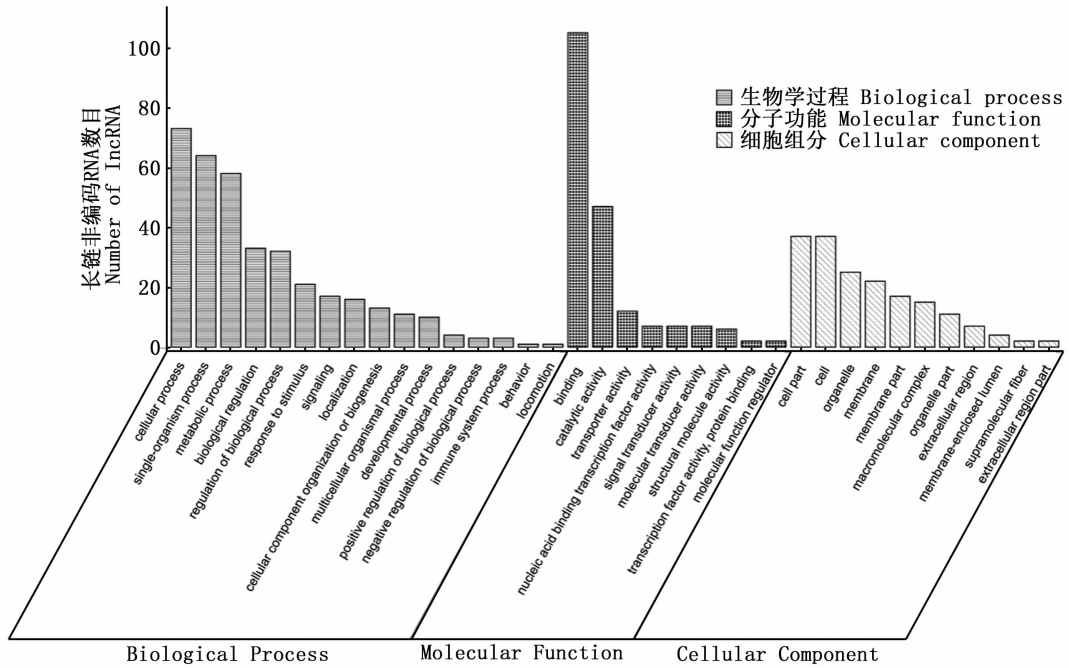


图 7 长链非编码 RNA 对应靶基因的 GO 注释

Fig.7 Functional annotation of the potential targets of lncRNA

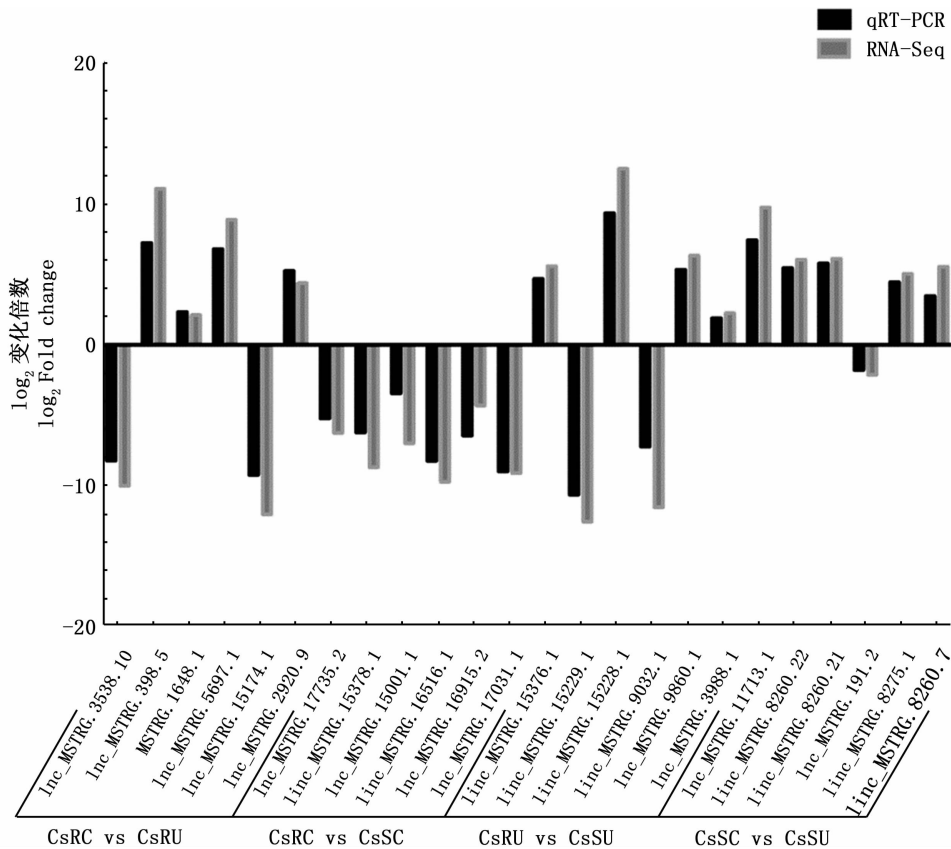


图 8 差异表达 lncRNA 验证

Fig.8 Validation of differentially expressed lncRNA by qPCR

3 讨论

lncRNA 是一类内源性 RNA,它可以调控基因表达,从而影响生物性状。本研究通过 lncRNA 识别从转录组数据中获得大量的 lncRNA 座位。通过基因组等容线结构分析发现 GC 含量高的基因组区域有更多的编码基因^[18],因此编码基因的 GC 含量高于 lncRNA; lncRNA 多为单外显子, lncRNA 中单外显子所占的比例远高于编码基因,然而就外显子和转录本的平均长度而言, lncRNA 均大于具有相同外显子数目的编码基因; RAN 等^[19]发现编码基因的平均表达量高于 lncRNA,与本实验统计结果相似。通过 lncRNA 和编码基因基本特征比较结果说明, lncRNA 基本特性符合普遍规律,满足后续分析。

lncRNA 已经在模式生物的免疫调控研究上广泛应用。在斑马鱼中,利用转录组数据在两个实验组中分别挖掘出 2 842 和 3 433 个差异 lncRNA,并通过 qPCR 获得差异表达较大的 lncRNA,随后利用原位杂交技术筛选出 3 条 lncRNA,证明它们调控的 *plk3* 和 *syt10* 基因与免疫应答反应相关^[20]。关于 lncRNA 在半滑舌鲷免疫功能上的研究未见报道。本研究通过 lncRNA 差异表达分析,共筛选出 2 032 个差异表达 lncRNA。值得注意的是,抗病家系和易感家系之间差异 lncRNA 数目占了大多数,这些差异 lncRNA 可能与抗病家系和易感家系的免疫力差异有着密切联系。在共表达分析中,基于表达相关系数,检测出 285 个 lncRNA 与 274 个编码基因可能存在 14 539 种共表达关系。其中,160 个编码基因被成功注释,主要涉及信号转导、转录调控和生物合成过程,有 3 个基因注释为“免疫反应”和“炎症反应”功能,例如 *il-1 β* 、*ccl20*、*ccl28*。YU 等^[21]发现 *il-1 β* 基因在半滑舌鲷注射鳃弧菌 8 h 时头肾和肝脏的表达量显著升高,表明 *il-1 β* 基因在半滑舌鲷免疫应答中扮演着重要作用^[21]。SCHUTYSER 等^[22]研究发现 LPS、肿瘤坏死因子、干扰素 γ 都能诱导 *ccl20* 基因表达,随后 *ccl20* 基因在炎症反应中引起 CD4 + T 细胞的增加。此外,*ccl28* 基因通过趋化因子受体 CCR3 和 CCR10 调节 B 细胞和 T 细胞的迁移^[23]。研究结果表明 7 个关键 lncRNA 可能在半滑舌鲷免疫调控方面起到了重要作用。

lncRNA 在免疫应答调控、表观遗传学等方面扮演着重要的角色^[24-26]。然而半滑舌鲷 lncRNA 的研究仍处于初步阶段,本研究从半滑舌鲷转录组数据中识别出大量的 lncRNA,并通过比较分析,筛选出可能与抗哈维弧菌相关的差异 lncRNA,并通过共表达分析获得 274 个候选的靶基因,通过功能注释筛选出 7 个与半滑舌鲷免疫应答中的关键 lncRNA,为后续研究 ceRNA 网络构建奠定数据基础。

参考文献:

- [1] PONTING C P, OLIVER P L, REIK W. Evolution and functions of long noncoding RNAs [J]. *Cell*, 2009, 136 (4): 629-641.
- [2] RINN J L, KERTESZ M, WANG J K, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs [J]. *Cell*, 2007, 129 (7): 1311-1323.
- [3] MARTIANOV I, RAMADASS A, BARROS A S, et al. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript [J]. *Nature*, 2007, 445 (7128): 666-670.
- [4] TRIPATHI V, ELLIS J D, SHEN Z, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation [J]. *Molecular Cell*, 2010, 39 (6): 925-938.
- [5] 陈松林, 杜民, 杨景峰, 等. 半滑舌鲷家系建立及其生长和抗病性能测定 [J]. *水产学报*, 2010, 34 (12): 1789-1794.
CHEN S L, DU M, YANG J F, et al. Development and characterization for growth rate and disease resistance of families in half-smooth tongue sole (*Cynoglossus semilaevis*) [J]. *Journal of Fisheries of China*, 2010, 34 (12): 1789-1794.
- [6] 戴欢, 刘洋, 王文文, 等. 半滑舌鲷抗哈维氏弧菌病相关微卫星标记筛选及 QTL 定位 [J]. *中国水产科学*, 2017, 24 (1): 22-30.
DAI H, LIU Y, WANG W W, et al. Detection of SSR and a QTL analysis of *Vibrio harveyi* resistance in *Cynoglossus semilaevis* [J]. *Journal of Fishery Sciences of China*, 2017, 24 (1): 22-30.
- [7] XU H, XU X W, LI X H, et al. Comparative transcriptome profiling of immune response against *Vibrio harveyi* infection in Chinese tongue sole [J]. *Scientific Data*, 2019, 6 (1): 224.
- [8] ZHOU Q, SU X Q, JING G C, et al. RNA-QC-chain: comprehensive and fast quality control for RNA-Seq data [J]. *BMC Genomics*, 2018, 19 (1): 144.
- [9] KIM D, LANGMEAD B, SALZBERG S L. HISAT: a fast

- spliced aligner with low memory requirements [J]. *Nature Methods*, 2015, 12(4): 357-360.
- [10] TRAPNELL C, WILLIAMS B A, PERTEA G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation [J]. *Nature Biotechnology*, 2010, 28(5): 511-515.
- [11] NIKNAFS Y S, PANDIANB, IYER H K, et al. TACO produces robust multisample transcriptome assemblies from RNA-seq [J]. *Nature Methods*, 2017, 14(1): 68-70.
- [12] FINN R D, CLEMENTS J, EDDY S R. HMMER web server: interactive sequence similarity searching [J]. *Nucleic Acids Research*, 2011, 39(s2): W29-W37.
- [13] FINN R D, COGGILL P, EBERHARDT R Y, et al. The Pfam protein families database: towards a more sustainable future [J]. *Nucleic Acids Research*, 2016, 44(D1): D279-D285.
- [14] WUCHER V, LEGEAI F, HÉDAN B, et al. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome [J]. *Nucleic Acids Research*, 2017, 45(8): e57.
- [15] LI B, DEWEYC N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome [J]. *BMC Bioinformatics*, 2011, 12(1): 323.
- [16] ANDERS S, PYL P T, HUBER W. HTSeq—a Python framework to work with high-throughput sequencing data [J]. *Bioinformatics*, 2015, 31(2): 166-169.
- [17] 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.
- [18] GLÉMIN S, CLÉMENT Y, DAVID J, et al. GC content evolution in coding regions of angiosperm genomes: a unifying hypothesis [J]. *Trends in Genetics*, 2014, 30(7): 263-270.
- [19] RAN M L, CHEN B, LI Z, et al. Systematic identification of long noncoding RNAs in immature and mature porcine testes [J]. *Biology of Reproduction*, 2016, 94(4): 77.
- [20] WANG X D, LIN J B, LI F H, et al. Screening and functional identification of lncRNAs under β -diketone antibiotic exposure to zebrafish (*Danio rerio*) using high-throughput sequencing [J]. *Aquatic Toxicology*, 2017, 182: 214-225.
- [21] YU Y, ZHONG Q W, LI C M, et al. Molecular cloning and characterization of interleukin-1 β in half-smooth tongue sole *Cynoglossus semilaevis* [J]. *Veterinary Immunology and Immunopathology*, 2012, 146(3/4): 270-276.
- [22] SCHUTYSER E, STRUYF S, MENTEN P, et al. Regulated production and molecular diversity of human liver and activation-regulated chemokine/macrophage inflammatory protein-3 α from normal and transformed cells [J]. *The Journal of Immunology*, 2000, 165(8): 4470-4477.
- [23] SUN C, ZHANG Y Y, TANG C L, et al. Chemokine CCL28 induces apoptosis of decidual stromal cells via binding CCR3/CCR10 in human spontaneous abortion [J]. *Molecular Human Reproduction*, 2013, 19(10): 676-686.
- [24] 王智诚, 周晓旭, 王昊泽, 等. 长链非编码 RNA 及其在斑马鱼中的研究进展 [J]. *大连海洋大学学报*, 2017, 32(2): 248-254.
- WANG Z C, ZHOU X X, WANG H Z, et al. Long non-coding RNA and its research advances in zebrafish *Danio rerio*: a review [J]. *Journal of Dalian Ocean University*, 2017, 32(2): 248-254.
- [25] MORLANDO M, FATICA A. Alteration of epigenetic regulation by long noncoding RNAs in cancer [J]. *International Journal of Molecular Sciences*, 2018, 19(2): 570.
- [26] WANG W, ZHAO Z, YANG F, et al. An immune-related lncRNA signature for patients with anaplastic gliomas [J]. *Journal of Neuro-Oncology*, 2018, 136(2): 263-271.

Identification and characterization of long non-coding RNAs from *Cynoglossus semilaevis* based on transcriptomics

XU Hao^{1,2,3}, XU Xiwen^{1,3}, ZHOU Qian^{1,3}, CHEN Songlin^{1,3}

(1. National Laboratory for Qingdao Marine Science and Technology, Laboratory for Marine Fisheries Science and Food Production Processes, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, Shandong, China; 2. College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China; 3. Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture and Rural Affairs, Qingdao 266071, Shandong, China)

Abstract: Long non-coding RNA (lncRNA) plays an important role in the regulation of epigenetics, transcription, and post-transcription. To understand the regulation mechanism of immune response to *Vibrio harveyi*, we performed lncRNA identification and differential expression analysis from RNA-seq data. We obtained a total of 4 584 lncRNA loci, containing 5 714 transcripts in this study. The comparison of basic characteristics between lncRNA and the coding gene showed that the GC content of lncRNAs was lower than that of coding genes, and single exon genes number was greater than that of coding genes, and the average transcripts length of lncRNAs was longer than that of coding genes, and the average expression of lncRNAs was slightly lower than coding genes. We screened 818, 813, 261 and 140 differentially expressed lncRNAs by differential expression analysis in four comparisons of *Cynoglossus semilaevis* groups (CsRU vs CsSU, CsRC vs CsSC, CsRC vs CsRU, CsSC vs CsSU), respectively. Additionally, we preliminarily analyzed the expression patterns and the relationship of differential expression of lncRNAs from those four groups. Through co-expression analysis, a total of 14 539 types of interaction relationships were detected between 285 lncRNA and 274 coding genes. We performed GO annotation and screened 7 key lncRNA. The qRT-PCR confirmed that the expression patterns of lncRNAs were consistent with their expression levels bases on transcriptome data. This study provides some of reference data for studying the role of lncRNA in the immune response against *V. harveyi*.

Key words: *Cynoglossus semilaevis*; *Vibrio harveyi*; lncRNA; immune response; differentially expressed lncRNA