文章编号：1000-0615(2020)05-0723-12　　水产学报, 2020, 44(5): 723−734

JOURNAL OF FISHERIES OF CHINA

DOI: 10.11964/jfc.20190211659

Science Press

长牡蛎性腺中调控型非编码RNA的生物信息学

王　雪[1,2,3]，王卫军[1,3,4*]，骆启豪[1,2,3]，孙国华[4]，冯艳微[4]，马敬俊[5]，杨建敏[1,3,4*]

(1. 上海海洋大学，水产科学国家级实验教学示范中心，上海　201306；
2. 上海海洋大学，上海水产养殖工程技术研究中心，上海　201306；
3. 山东省海洋资源与环境研究院，山东 烟台　264006；
4. 鲁东大学农学院，山东 烟台　264025；
5. 烟台市莱山区渔业海洋站，山东 烟台　264003)

摘要：本研究以日照海域同一家系的2龄长牡蛎性腺为研究对象，通过small RNA-seq和RNA-seq技术筛选和鉴定出大量的非编码微小RNA(miRNA)、长链非编码RNA(lncRNA)和环状RNA(circRNA)，并对其进行了生物学特征分析。结果显示，以斑马鱼为参考序列，获得25~30个已知miRNA成熟体和51~63个已知miRNA前体，预测到53~71个新miRNA成熟体和53~77个新miRNA前体；长牡蛎miRNA长度分布为18~26个核苷酸(nt)，其中分布在20~22 nt长度的miRNA数量最多，且miRNA首位碱基多为U。测序分析获得2 302~2 349个注释lncRNA转录本，预测到20 083~24 114个新lncRNA转录本，其中基因间型lncRNA(lincRNA)、内含子lncRNA(intronic lncRNA)、反义lncRNA(anti-sense lncRNA)分别占29.0%、62.1%和8.9%；长牡蛎lncRNA的基因组特征与其他真核生物的lncRNA基因组特征相似，与mRNA相比，外显子(exon)个数少，转录本长度较短，表达水平低。测序分析共获得383个circRNA，其中平均88.54%来源于exon，平均4.51%来源于intronic，平均6.95%来源于intergenic，且鉴定出内源性circRNA潜在大量的miRNA结合位点。研究结果为后续研究长牡蛎调控型ncRNA的表达规律和生物学功能奠定了基础。

关键词：长牡蛎；非编码微小RNA；长链非编码RNA；环状RNA；RNA-seq技术

中图分类号：Q 785; S 968.3　　**文献标志码：**A

根据RNA是否具有翻译蛋白质的能力，可将其分为两种类型：信使RNA(mRNA)和非编码RNA(non-coding RNA，ncRNA)。非编码RNA是指从基因组转录得到的不编码蛋白的功能性RNA分子[1]。目前，非编码RNA分为看家非编码RNA(housekeeping ncRNA)和调控型非编码RNA(regulatory ncRNA)两类，其中常见的调控型非编码RNA包括长度为19~25个核苷酸(nt)的微小RNA(microRNA，miRNA)、长度大于200 nt的长链非编码RNA(long non-coding RNA，lncRNA)和闭合单链的环状RNA(circular RNA，circRNA)[2]。miRNA是RNA诱导沉默复合体(RNA induced silencing complex，RISC)的重要组成部分，在动物中一般按碱基互补配对原则与靶基因3′非翻译区(3′untranslated region，3′UTR)特异结合，从而抑制靶基因的翻译水平或剪切降解靶基因，调控基因表达[3]。lncRNA通过顺式和反式两种方式调控靶基因，哺乳动物雌性其中一条X染色体的失活就是lncRNA调控作用的结果[4]。lncRNA和circRNA分子可以通过碱基互补原则与miRNA结

收稿日期：2019-02-13　　**修回日期：**2019-04-01

资助项目：国家自然科学基金(31402298)；山东省农业良种工程(2017LZGC009)；国家贝类产业技术体系专项(CARS-49)

通信作者：王卫军，E-mail：wwj2530616@163.com；杨建敏，E-mail：ladderup@126.com

合，类似海绵吸附作用，抑制miRNA与靶基因结合，调控靶基因的表达水平[5]。随着基因注释信息和高通量测序技术的不断完善，在真核生物中发现调控型ncRNA参与了诸如催化RNA前体中内含子的剪切、基因组重构、调控细胞发育和分化及表观遗传调控等[6-8]许多生命活动，在其中扮演着重要角色。

长牡蛎(*Crassostrea gigas*)，俗称太平洋牡蛎，隶属于软体动物门(Mollusca)，是一种分布广泛且具有重要经济价值的海水养殖贝类[9]。2017年中国牡蛎养殖产量达487.9万t，居海水养殖贝类产量首位[10]。海洋生物的ncRNA研究起步较晚，且主要集中在海洋脊椎动物。2003年Lim等[11]首次在斑马鱼(*Danio rerio*)中鉴定出38个miRNA。之后对虹鳟(*Oncorhynchus mykiss*)、尼罗罗非鱼(*Oreochromis niloticus*)、马氏珠母贝(*Pinctada martensii*)、明钩虾(*Parhyale hawaiensis*)、长牡蛎等多个物种进行miRNA测序，研究表明，miRNA在海洋生物的免疫、发育和生殖等方面起重要作用[12-16]。Pauli等[17]研究发现，斑马鱼lncRNA发育阶段的特异性强于mRNA。之后Yu等[18]发现长牡蛎基因间型lncRNA (lincRNA)与幼虫变态发育相关，Feng等[19]发现长牡蛎外套膜lncRNA与壳色素相关。目前对长牡蛎circRNA的研究尚未见报道。本实验利用illumina HiSeq™ 2500等测序平台对长牡蛎的性腺组织进行miRNA、lncRNA、circRNA测序和生物信息学分析，为今后深入研究长牡蛎ncRNA的调控机理奠定基础。

1 材料与方法

1.1 实验材料

本实验所用样品为养殖于日照海域同一家系的2龄长牡蛎，于2018年3月选取3个性状优良的个体，取其性腺组织，分别装入无RNase的离心管中，液氮保存。长牡蛎总RNA的提取、转录组的建库测序和拼接组装委托北京诺禾致源科技股份有限公司完成。

1.2 small RNA文库的制备与测序

测序数据质量控制　为了保证信息分析的质量，去除测序得到的原始序列(raw reads)中含有带接头的、低质量的、被污染的及含有polyA/T/G/C的序列，得到干净的序列(clean reads)。最后筛选18~35 nt clean reads用于后续分析。

small RNA分类注释　用bowtie软件[20]将长度筛选后的small RNA定位到参考序列上，将能比对到参考序列的small RNA与斑马鱼已经注释的RNA进行比对。考虑到某个small RNA能同时比对上多个注释信息，为了使每个small RNA具有唯一的注释，按照已知miRNA>rRNA>tRNA>snRNA>snoRNA>repeat>gene的优先级顺序进行注释。对于没有注释的small RNA，利用miREvo[21]和mirdeep2[22]平台来分析其是否含有miRNA的特点，如可能的miRNA前体的二级结构、Dicer酶切位点信息等，预测分析新miRNA。并且对各样品中匹配到miRBase中已知miRNA前体序列的small RNA和预测的新miRNA的首位点碱基分布情况进行统计。

1.3 lncRNA文库的制备与测序

测序原始数据质量评价及序列组装分析　为了保证信息分析的质量，去除测序得到的raw reads中含有带接头的、低质量的、含有poly-N的序列，得到clean reads。通过Hisat2[23]将clean reads比对到基因组上，利用Scripturet[24]和Cufflink[25]两款组装软件将能对比到基因组上的clean reads进行拼接。

lncRNA筛选鉴定与基因组特征分析　将拼接好的转录本按照以下步骤进行lncRNA的筛选，最后得到的即为预测的lnRNA：①选择外显子(exon)个数≥2的转录本；②选择长度>200 bp的转录本；③通过Cuffcompare软件[25]，筛除与数据库注释exon区域有重叠的转录本，并将数据库中与本次拼接转录本exon区域有重叠的lncRNA作为数据库注释lncRNA纳入到后续分析；④通过Cuffquant软件[25]计算每条转录本的表达量，选择每百万片段中来自某一基因每千碱基长度的片段数目(expected number of fragments per kilobase of transcript sequence per millions base pairs sequenced，FPKM)≥0.5的转录本；⑤通过CPC[26]和Pfam scan[27]两款软件进行编码潜能筛选，对预测的lncRNA进行基因组特征分析，并将其与mRNA进行比较，对比参数包括exon个数、开放阅读框(ORF)长度、转录本核酸长度以及物种间序列保守性，以了解长牡蛎lncRNA的基因组特点。

circRNA筛选鉴定与基因组特征分析　在构建的lncRNA文库中筛选鉴定circRNA，由于circRNA鉴定存在假阳性高的现象[28]，使用find_circ[29]和CIRI2[30]两款软件对circRNA进行筛选，

对两款软件鉴定出的circRNA的结果进行合并，取二者的交集。

circRNA-miRNA结合位点分析　用miRanda软件[31]预测剪切后的circRNA的miRNA的结合位点。

2　结果

2.1　small RNA建库测序原始数据质量评价及序列组装分析

通过Illumina HiSeq™2500平台完成RNA-Seq测序，共得到了44 569 271条raw reads，去除其中的接头序列和低质量序列，获得约96.92% clean reads，碱基位置的测序错误率(error rate)为0.01%<0.5%，GC含量均值为49.10%，在40%~60%的区间内，Q30均值为94.46%>85%(表1)。上述结果说明本实验中文库的构建和RNA-Seq测序的结果良好，可进行后续分析。对样品的clean reads进行18~35 nt范围内的small RNA筛选(图1)，共获得40 541 719条序列可进行后续分析。将长度筛选后的small RNA定位到参考序列上，平均约83.45%的序列可以比对到参考序列上，比对到参考序列方向相同链的序列占51.08%，比对到参考序列方向相反链的序列占32.39%。

2.2　miRNA鉴定分析

将上述比对上的序列，与miRBase中的斑马鱼数据库进行比对。3个样品分别检测到25、26和30个已知miRNA成熟体和52、51和63个已知miRNA前体。这些miRNA长度分布于18~26 nt，其中，分布在20、21和22 nt长度的miRNA数量最多。长度18~30 nt的已知miRNA首位碱基偏好性分析结果显示，不同长度的miRNA的首位碱基偏好性差异明显，长度为21 nt的miRNA首位碱基为A和U，其他长度miRNA首位碱基多为U(图2)。在鉴定的已知miRNA中，dre-miR-100-5p、dre-miR-10a-5p、dre-miR-184、dre-miR-7a、dre-miR-1、dre-let-7a、dre-miR-10c-5p、dre-miR-133a-3p、

表 1　Small RNA文库测序数据过滤和基因组定位信息

Tab. 1　Sequencing data filtering and reads mapping to the reference in three small RNA libraries

项目名 items	样品1 sample 1	样品2 sample 2	样品3 sample 3
原始序列/条　raw reads	12 777 844	15 403 032	16 388 395
错误率/%　error rate	0.01	0.01	0.01
Q20/%	96.99	97.56	97.62
Q30/%	93.65	94.81	94.90
GC含量/%　GC content	48.15	47.92	51.24
N>10%	0	0	0
低质量序列/条　low quality	28 095 (0.22%)	41 084 (0.27%)	53 021 (0.32%)
5′接头污染序列/条 5′_adapter_contamine	9 476 (0.07%)	6 322 (0.04%)	9 472 (0.06%)
无3′ 接头或没有插入片段的序列/条 3′_adapter_null or insert_null	204 006 (1.60%)	251 553 (1.63%)	661 159 (4.03%)
含有ployA/T/G/C的序列/条 with ployA/T/G/C	59 091 (0.46%)	68 226 (0.44%)	16 772 (0.10%)
干净序列/ 条　clean reads	12 477 176 (97.65%)	15 035 847 (97.62%)	15 647 971 (95.48%)
总small RNA/条 total small RNA	12 007 144	14 273 654	14 260 921
small RNA比对到参考基因组/条 mapped small RNA	9 504 317 (79.16%)	11 537 915 (80.83%)	12 893 409 (90.41%)
small RNA比对到参考基因组相同链/条 "+" mapped small RNA	5 717 391 (47.62%)	6 472 619 (45.35%)	8 593 879 (60.26%)
small RNA比对到参考基因组相反链/条 "-" mapped small RNA	3 786 926 (31.54%)	5 065 296 (35.49%)	4 299 530 (30.15%)

注：Q20、Q30分别表示 Phred 数值大于20、30的碱基占总体碱基的百分比；N. 无法确定碱基信息；括号中数值为该项占对应总序列数的百分比；下同

Notes: Q20, Q30 indicated the percentage of bases with phred values greater than 20 and 30 of the total bases, respectively; N. the base information cannot be determined; the values between parentheses are the percentage of the corresponding total number; the same below

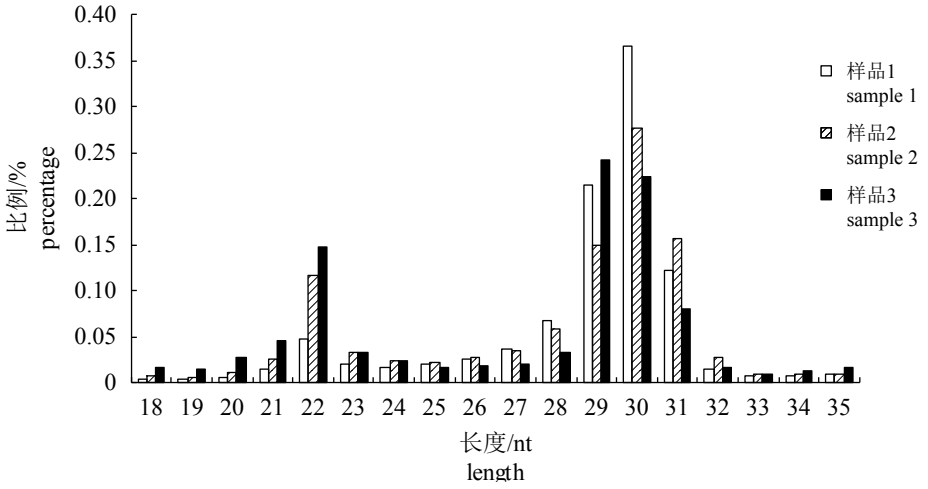

图 1 所得total small RNA片段的长度分布统计

Fig. 1 Length distribution of total small RNA

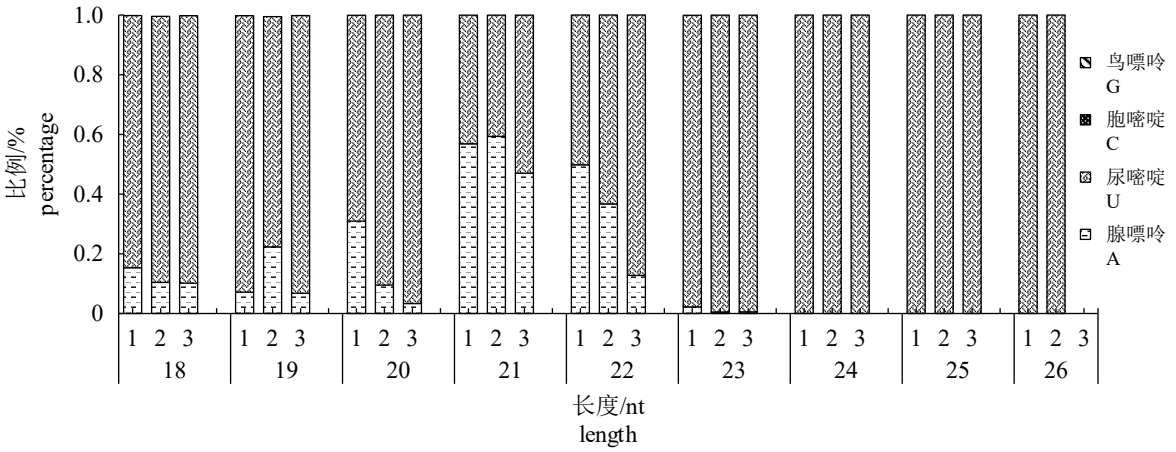

图 2 长度18~30 nt的已知miRNA首位碱基偏好性分析

1. 样品1，2. 样品2，3. 样品3，下同

Fig. 2 Base bias analysis of known miRNA with a length of 18-30 nt on the first nucleotide position

1. sample 1, 2. sample 2, 3. sample 3, the same below

dre-miR-99和dre-miR-9-5p表达量较高，部分匹配结果见表2。3个样品分别预测到60、71和53个新miRNA成熟体与66、77和53个新miRNA前体，对预测到的新miRNA首位碱基偏好性分析结果显示，其首位碱基多为U，与已知miRNA首位碱基分析结果相一致(图3)。

2.3 lncRNA建库测序原始数据质量评价及序列组装分析

通过Illumina HiSeq™2500平台完成RNA-Seq测序，共得到了300 331 640条raw reads，去除其中的接头序列和低质量序列，获得用于进行后续各项生物信息学分析的clean reads 283 548 118条，碱基位置的测序错误率低于0.5%，GC含量均值为42.87%，在40%~60%的区间内，Q30均值为92.83%>85%(表3)，上述结果说明本实验中文库的构建和RNA-Seq测序的结果良好，可进行后续分析。经过TopHat软件比对发现，平均75.20%的clean reads可以比对到长牡蛎参考基因组上，并且平均67.06%的序列具有唯一的基因组位置，平均33.50%序列比对到基因组正链，平均约33.56%序列比对到基因组负链。

使用HTSeq软件，对3个样品进行已知类型基因的定量分析，根据表达量统计，得到样品中各类型基因的表达情况。数据统计结果显示，exon平均约占0.12%，ncRNA平均占1.60%，protein coding平均占68.98%(表4)。

表 2　匹配上的miRNA成熟体的数量

Tab. 2　Number of matched miRNA matures

miRNA成熟体id miRNA mature id	匹配到该成熟体的数量/个 number of matched matures		
	样品1 sample 1	样品2 sample 2	样品3 sample 3
dre-let-7a	14 859	10 766	7 723
dre-let-7b	8	1	0
dre-let-7c-5p	711	294	592
dre-let-7d-5p	1	0	2
dre-let-7f	28	164	963
dre-let-7g	4	5	6
dre-miR-1	21 991	82 378	96 769
dre-miR-100-5p	217 933	467 823	197 162
dre-miR-101a	2	3	56
dre-miR-10a-5p	143 028	524 751	928 026
dre-miR-10b-5p	46	172	708
dre-miR-10c-5p	366	1 640	3 504
dre-miR-124-3p	0	0	24
dre-miR-125b-5p	9	25	183
dre-miR-133a-3p	263	1 529	752
dre-miR-133b-3p	3	8	13
dre-miR-141-3p	0	0	33
dre-miR-142a-5p	0	0	15
dre-miR-143	0	2	0
dre-miR-153a-3p	12	25	0
dre-miR-153c-3p	0	1	0
dre-miR-183-5p	0	4	15
dre-miR-184	69 366	245 937	428 977
dre-miR-200a-3p	0	0	33
dre-miR-206-3p	0	0	18
dre-miR-29a	3	5	1
dre-miR-29b	25	23	14
dre-miR-30e-5p	0	0	10
dre-miR-7a	24 022	52 801	54 570
dre-miR-7b	16	50	107
dre-miR-92a-3p	1	0	4
dre-miR-92b-3p	16	8	1
dre-miR-9-5p	108	588	1 586
dre-miR-99	148	396	1 941

2.4　lncRNA筛选与鉴定

经过筛选，3个样品中分别筛选出2 302、2 349和2 316个注释lncRNA和24 114、21 921、20 083个新lncRNA。对lncRNA转录本进一步分析，得到不同种类的lncRNA占比情况(图4)，其中基因间型lncRNA占29.0%，内含子lncRNA(intronic lncRNA)占62.1%，反义lncRNA(anti-sense lncRNA)占8.9%。为了深入解析长牡蛎lncRNA的基因组特征，对鉴定得到的lncRNA与已知mRNA进行生物信息学对比，涉及exon数、转录本长度和ORF长度(图5)，其中lncRNA的exon个数少，转录本长度和ORF的长度较短。

2.5　circRNA鉴定分析

在3个长牡蛎样品的测序文库中鉴定出383个circRNA，对circRNA进行转录本长度分析发现，在3个样品中鉴定出的circRNA转录本长度范围跨度很大，最短长度为153 nt，最长的长度为60 025 nt，有79.13%的circRNA在10 000 nt以下(图6)。对circRNA进行来源统计分析，平均88.54% circRNA来源于exon，平均4.51% circRNA来源于intronic，平均6.95% circRNA来源于基因间(图7)。

2.6　circRNA-miRNA结合位点分析

为了探索长牡蛎性腺中circRNA是否具有与其他模式生物circRNA相似的miRNA吸附海绵的功能，对鉴定出的383个circRNA进行miRNA的靶向预测。结果显示，长牡蛎的circRNA序列上潜在有大量miRNA的靶位点、部分检测到的circRNA与其靶miRNA(表5)。

3　讨论

动物small RNA的长度区间一般为18~35 nt，本研究检测得到的新miRNA成熟体和新miRNA前体与其他物种如日本吸血虫(*Schistosoma japonicum*)成虫、马氏珠母贝[32-33]的miRNA长度相似。长牡蛎miRNA的碱基偏好性分析发现，其首位碱基多为U，该现象可能与miRNA成熟体形成过程中Dicer酶切产物特性相关，偏向U更有助于与AGO蛋白结合形成蛋白复合体，与靶基因结合，调控靶基因的表达[34]。性腺发育是水产动物繁育生殖的基础，众多的生化过程参与其中，且miRNA在其发育过程中起重要作用。在

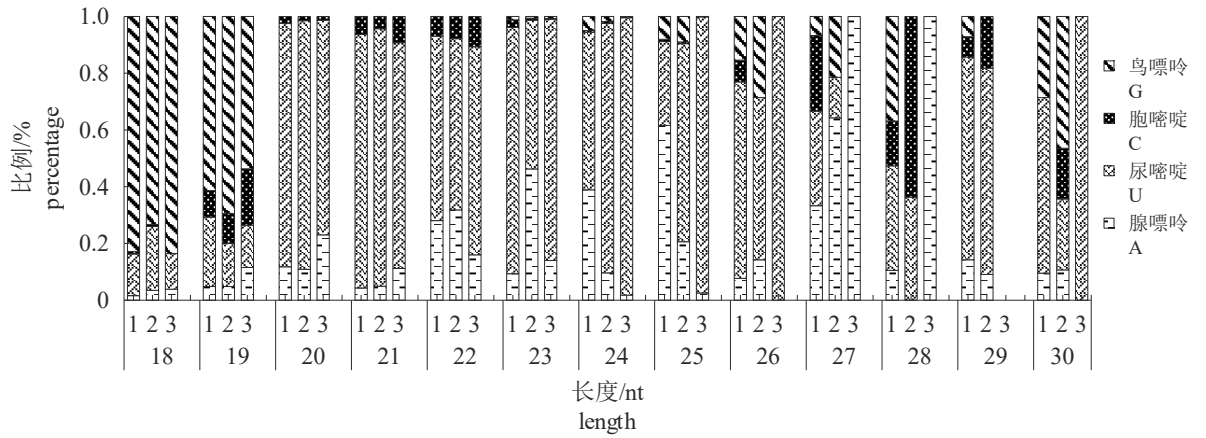

图 3 长度18~30 nt的新miRNA首位碱基偏好性

Fig. 3 Base bias analysis of novel miRNA with a length of 18-30 nt on the first nucleotide position

表 3 lncRNA文库测序数据过滤和基因组定位

Tab. 3 Sequencing data filtering and reads mapping to the reference in three lncRNA libraries

项目名 item	样品1 sample 1	样品2 sample 2	样品3 sample3
原始序列/条 raw reads	102 088 824	100 882 698	97 360 118
干净序列/条 clean reads	99 074 286	93 482 494	90 991 338
干净的碱基数/G clean bases	14.86	14.02	13.65
错误率/% error rate	0.01	0.02	0.02
Q20/%	97.81	97.05	96.37
Q30/%	94.35	92.86	91.27
GC含量/% GC content	44.63	41.47	42.5
比对上总数/条 total mapped	73 175 827 (73.86%)	70 630 915 (75.56%)	69 326 512 (76.19%)
多个基因组位置的序列数/条 multiple mapped	6510 490 (6.57%)	7 298 644 (7.81%)	9 148 034 (10.05%)
唯一基因组位置的序列数/条 uniquely mapped	66 665 337 (67.29%)	63 332 271 (67.75%)	60 178 478 (66.14%)
第1次比对序列数/条 read-1	33 558 489 (33.87%)	32 119 880 (34.36%)	30 776 583 (33.82%)
第2次比对序列数/条 read-2	33 106 848 (33.42%)	31 212 391 (33.39%)	29 401 895 (32.31%)
比对到相同链的序列数/条 reads map to '+'	33 312 466 (33.62%)	31 607 568 (33.81%)	30 088 023 (33.07%)
比对到相反链的序列数/条 reads map to '–'	33 352 871 (33.66%)	31 724 703 (33.94%)	30 090 455 (33.07%)

鉴定表达量较高的已知miRNA中，dre-miR-184、dre-let-7a和dre-miR-133有类似的调控性腺发育的功能[35-37]，dre-miR-100-5p、dre-let-7a、dre-miR-1、dre-miR-9-5p和dre-miR-133a-3p有类似的调控生长发育的功能[38-40]。Bizuayehu等[35]对调控大西洋庸鲽(*Atlantic halibut*)雄性性成熟早于雌性这一现象研究时发现，同龄精巢成熟而卵巢未成熟的雌雄个体中，let-7a、miR-143和miR-202-3p参与了性腺发育的调控过程。He等[36]和Song等[37]对中华绒螯蟹(*Eriocheir sinensis*)性腺的miRNA测序分析发现，miR-184和miR-133分别在卵巢和精巢中差异表达，且miR-133可以调控*cyclin B*基因的表达。

长牡蛎lncRNA的基因组特征与mRNA相比，具有exon个数少，转录本长度较短，表达水平低的特点。由于软体动物lncRNA数据库中缺乏牡蛎的lncRNA的注释信息，研究过程中未对牡蛎中lncRNA的保守性进行评估，但是之前的研究报道中指出在长牡蛎中lincRNA保守性低[18]。本实验基于基因组的位置，将lncRNA分为3类，

表 4　已知类型的基因序列分布情况

Tab. 4　Gene distribution of reads in known types

类型 types	序列的类型分布数量/条 number of read types		
	样品1 sample 1	样品2 sample 2	样品3 sample 3
外显子　exon	40 688 (0.13%)	36 368 (0.12%)	35 639 (0.12%)
庞杂类RNA分子　miscellaneous RNA	20 476 (0.06%)	19 905 (0.06%)	16 867 (0.06%)
非编码RNA　ncRNA	464 828 (1.44%)	516 138 (1.66%)	497 693 (1.70%)
蛋白编码类　protein coding	21 974 117 (67.92%)	21 510 556 (69.39%)	20 360 545 (69.63%)
核糖体RNA　rRNA	77 952 (0.24%)	73 342 (0.24%)	68 730 (0.24%)
转运RNA　tRNA	128 561 (0.40%)	109 717 (0.35%)	144 725 (0.49%)
其他　others	9 644 876 (29.81%)	8 734 947 (28.18%)	8 118 910 (27.76%)

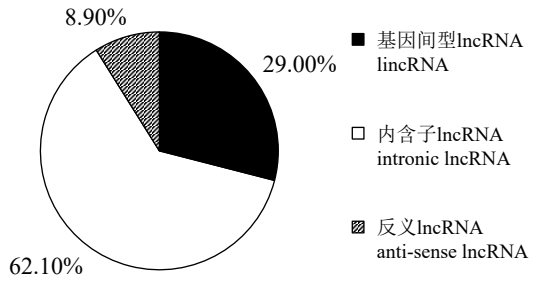

图 4　lncRNA 类型分布图

Fig. 4　Classification of lncRNA

其中lincRNA占29.0%，intronic lncRNA占62.1%，anti-sense lncRNA占8.9%。intronic lncRNA指的是完全从编码基因的intronic区域转录出来的lncRNA，其保守性远低于lincRNA，相比于lincRNA和anti-sense lncRNA，intronic lncRNA的研究相对较少。lincRNA位于编码基因间的区域，既可以与编码基因在同一条链上，也可以在相对链上，相对于其他的lncRNA，lincRNA不和编码基因重叠，验证一条新的lincRNA不用对其非编码性进行额

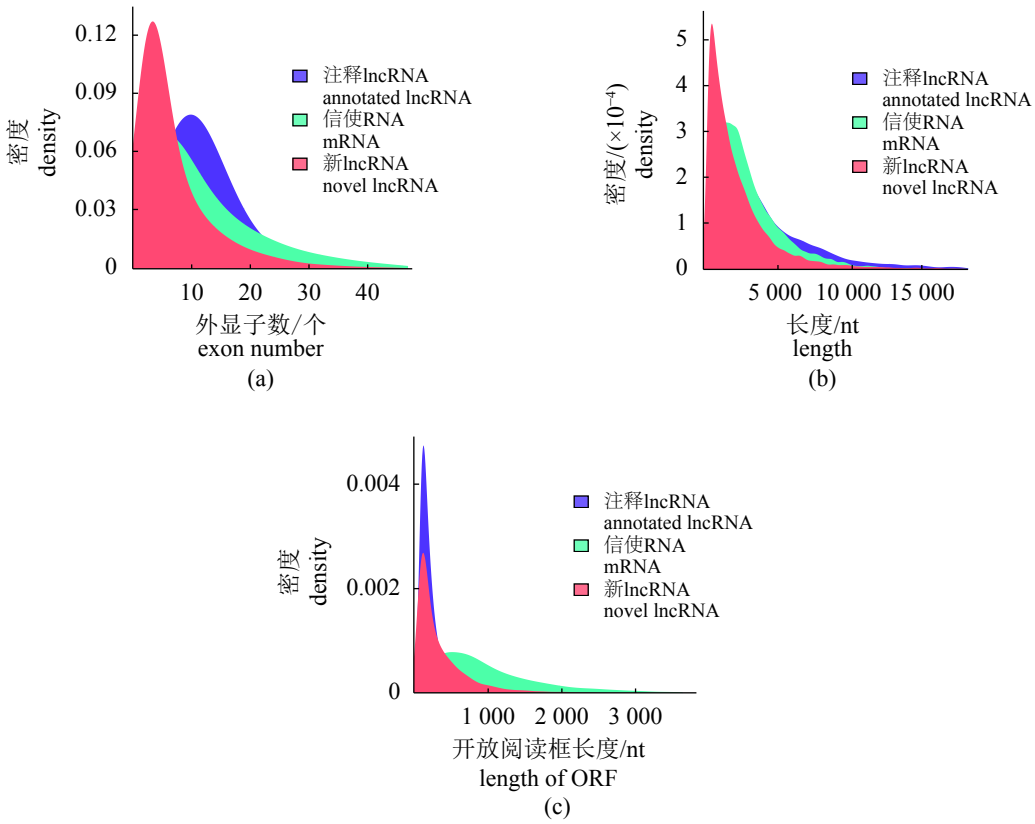

图 5　lncRNA的基因组特征

(a) exon数量；(b) 转录本长度；(c) ORF长度

Fig. 5　Genomic features of candidate lncRNA

(a) exon number; (b) transcript length; (c) ORF length

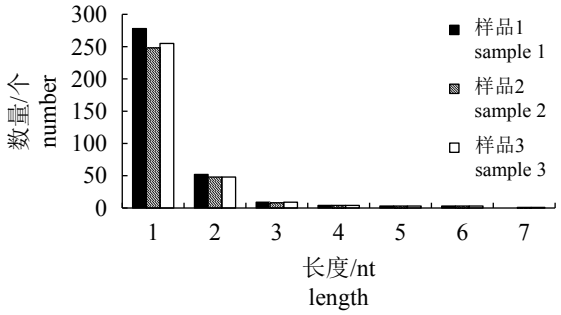

图 6 circRNA的长度分布

Fig. 6 Length distribution of circRNA

1. 10 000, 2. 20 000, 3. 30 000,4. 40 000, 5. 50 000, 6. 60 000, 7. 70 000

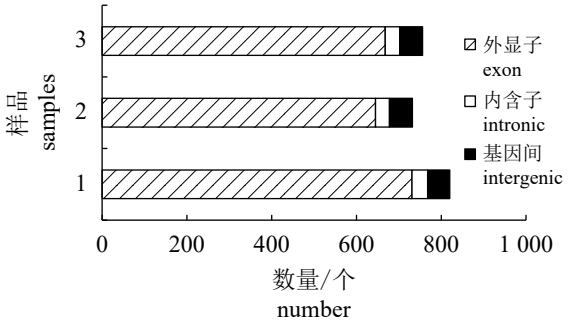

图 7 circRNA的来源

Fig. 7 Origins of circRNA

表 5 预测的部分circRNA的靶miRNA

Tab. 5 Predicted target miRNA of some circRNA

新circRNA id novel circRNA id	潜在靶miRNA potential target miRNA
novel_circ_0 000 012	dre-miR-181b-5p、dre-miR-181a-5p、dre-miR-29b、dre-miR-29a、dre-miR-92b-5p、dre-miR-181c-5p
novel_circ_0 002 303	dre-miR-7b、dre-miR-7a、dre-miR-203a-3p、dre-miR-204-5p、dre-miR-103、dre-miR-107a-3p、dre-miR-130c-5p、dre-miR-146a、dre-miR-203b-3pdre-miR-1388-5p、dre-miR-107b
novel_circ_0 004 099	dre-miR-10a-5p、dre-miR-10b-5p、dre-miR-10c-5p、dre-miR-10d-5p、dre-miR-19a-3p、dre-miR-19b-3p、dre-miR-19c-3p、dre-miR-19d-3p、dre-miR-22a-5p、dre-miR-22b-5p、dre-miR-107a-5p、dre-miR-129-3p、dre-miR-193a-5p、dre-miR-218a、dre-miR-218b、dre-miR-301b-5p、dre-miR-430a-5p、dre-miR-430a-4-5p、dre-miR-430a-11-5p、dre-miR-430a-12-5p、dre-miR-430a-13-5p、dre-miR-430a-14-5p、dre-miR-430a-15-5p、dre-miR-430a-16-5p、dre-miR-430a-17-5p、dre-miR-430i-5p、dre-miR-461
novel_circ_0 000 415	dre-miR-181b-5p、dre-miR-205-5p、dre-miR-181a-5p、dre-miR-217、dre-miR-26a-5p、dre-miR-26b、dre-miR-133b-5p、dre-miR-181c-5p、dre-miR-202-3p、dre-miR-499-3p
novel_circ_0 002 273	dre-miR-10a-5p、dre-miR-10b-5p、dre-miR-10c-5p、dre-miR-10d-5p、dre-miR-22a-3p、dre-miR-22b-3p、dre-miR-96-3p、dre-miR-730
novel_circ_0 000 835	dre-miR-10a-5p、dre-miR-10a-3p、dre-miR-10b-5p、dre-miR-205-5p、dre-miR-10c-5p、dre-miR-10d-5p、dre-miR-15b-3p、dre-miR-16c-3p、dre-miR-23a-3p、dre-miR-23b、dre-miR-25-3p、dre-miR-92a-3p、dre-miR-92b-3p、dre-miR-130c-5p、dre-miR-140-5p、dre-miR-363-3p、dre-miR-723-5p、dre-miR-1 788-5p、dre-miR-2 189
novel_circ_0 000 642	dre-miR-10a-5p、dre-miR-10b-5p、dre-miR-10c-5p、dre-miR-10d-5p、dre-miR-18a、dre-miR-18b-5p、dre-miR-18b-3p、dre-miR-18c、dre-miR-26a-5p、dre-miR-26b、dre-miR-126a-3p、dre-miR-148、dre-miR-152、dre-miR-193a-5p、dre-miR-730、dre-miR-2 196、dre-miR-126b-3p、dre-miR-2 195
novel_circ_0 000 843	dre-miR-451、dre-miR-30e-3p、dre-miR-125a、dre-miR-125a、dre-miR-125b-5pdre-miR-125c-5p、dre-miR-125c-5p、dre-miR-143、dre-miR-731、dre-miR-2 196

外的排除实验。研究表明，lincRNA在细胞凋亡、胚胎发育、细胞分化[17,41-42]等生物过程中发挥重要的调节作用，这些特点使lincRNA成为目前研究最多、最为深入的lncRNA。

长牡蛎中circRNA的来源与秀丽隐杆线虫(*Caenorhabditis Elegans*)、人类(*Homo sapiens*)和小鼠(*Mus musculus*)等相似，除大量来源于exon外，还存在少量来源于intronic的circRNA。大部分的intronic在剪切作用后会被降解，但有些存在特殊核苷酸序列上的intronic不会在剪切之后被脱分支酶降解，从而形成circRNA[43-44]。从目前的报道中看出，circRNA在生物体内的含量丰富，且具有一定程度的序列保守性[45]。Hansen等[46]和Memczak等[29]研究发现，circRNA作为miRNA的海绵可以特异性诱捕内源性miRNA，进而改变miRNA-mRNA互作模式，提高靶基因的表达水平，从而参与基因的表达。Hansen等[46]在研究过程中还发现，circRNA中ciRS-7对应73个miR-7潜在靶位点，Sry RNA有16个miR-138潜在靶位点。在对植物、鱼类、猴[47-49]等研究中均发现circRNA具有miRNA的潜在靶位点。在本研究中，通过生物信息学预测，同样鉴定出内源性circRNA潜在大量的miRNA结合位点。研究过程中发现novel_circ_0 000 012潜在dre-miR-29a结合位点，Tian等[50]研究发现，miR-29a在珍珠层生物矿化过程和免疫过程中发挥重要调控作用。研究中还发现novel_circ_0 000 415潜在dre-miR-202-3p结合位点、novel_circ_0 000 843潜在dre-miR-143结合

位点，之前研究表明miR-143和miR-202-3p在性腺发育的过程中起调控作用[35]。本实验通过高通量测序技术和生物信息学技术对长牡蛎性腺进行系统的鉴定分析，这为后续对长牡蛎非编码RNA功能机制的研究奠定了坚实的基础。

参考文献：

[1] Chan J J, Tay Y. Noncoding RNA: RNA regulatory networks in cancer[J]. International Journal of Molecular Sciences, 2018, 19(5): 1310.

[2] 杨福兰, 饶周舟, 陈汉春. 非编码RNA与基因表达调控[J]. 生命的化学, 2014, 34(1): 119-125.
Yang F L, Rao Z Z, Chen H C. RNA regulation for gene expression[J]. Chemistry of Life, 2014, 34(1): 119-125(in Chinese).

[3] Qu K, Wang Z, Lin X L, *et al*. MicroRNAs: key regulators of endothelial progenitor cell functions[J]. Clinica Chimica Acta, 2015, 448: 65-73.

[4] 陆绮. X染色体失活现象与机制[J]. 自然杂志, 2017, 39(1): 25-30.
Lu Q. Mechanisms of X chromosome inactivation[J]. Chinese Journal of Nature, 2017, 39(1): 25-30(in Chinese).

[5] Chen J, Li Y, Zheng Q P, *et al*. Circular RNA profile identifies circPVT1 as a proliferative factor and prognostic marker in gastric cancer[J]. Cancer Letters, 2017, 388: 208-219.

[6] Mattick J S, Makunin I V. Non-coding RNA[J]. Human Molecular Genetics, 2006, 15 Suppl 1: R17-R29.

[7] Ma L, Bajic V B, Zhang Z. On the classification of long non-coding RNAs[J]. RNA Biology, 2013, 10(6): 925-933.

[8] 于红. 表观遗传学: 生物细胞非编码RNA调控的研究进展[J]. 遗传, 2009, 31(11): 1077-1086.
Yu H. Epigenetics: advances of non-coding RNAs regulation in mammalian cells[J]. Hereditas (Beijing), 2009, 31(11): 1077-1086(in Chinese).

[9] 王如才, 王昭萍. 海水贝类养殖[M]. 青岛: 中国海洋大学出版社, 2008: 116-174.
Wang R C, Wang Z P. Science of marine shellfish culture[M]. Qingdao: China Ocean University Press, 2008: 116-174(in Chinese).

[10] 农业农村部渔业渔政管理局. 中国渔业统计年鉴-2018[M]. 北京: 中国农业出版社, 2018.
Fisheries and Fishery Administration, Ministry of Agriculture and Rural Affairs. China fishery statistical yearbook 2018[M]. Beijing: China Agriculture Press, 2018(in Chinese).

[11] Lim L P, Glasner M E, Yekta S, *et al*. Vertebrate microRNA genes[J]. Science, 2003, 299(5612): 1540.

[12] Salem M, Xiao C D, Womack J, *et al*. A microRNA repertoire for functional genome research in rainbowtrout (*Oncorhynchus mykiss*)[J]. Marine Biotechnology, 2010, 12(4): 410-429.

[13] Yan B, Wang Z H, Zhu C D, *et al*. MicroRNA repertoire for functional genome research in tilapia identified by deep sequencing[J]. Molecular Biology Reports, 2014, 41(8): 4953-4963.

[14] Jiao Y, Zheng Z, Du X D, *et al*. Identification and characterization of microRNAs in pearl oyster *Pinctada martensii* by Solexa deep sequencing[J]. Marine Biotechnology, 2014, 16(1): 54-62.

[15] Blythe M J, Malla S, Everall R, *et al*. High through-put sequencing of the *Parhyale hawaiensis* mRNAs and microRNAs to aid comparative developmental studies[J]. PLoS One, 2012, 7(3): e33784.

[16] Martín-Gómez L, Villalba A, Kerkhoven R H, *et al*. Role of microRNAs in the immunity process of the flat oyster *Ostrea edulis* against bonamiosis[J]. Infection, Genetics and Evolution, 2014, 27: 40-50.

[17] Pauli A, Valen E, Lin M F, *et al*. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis[J]. Genome Research, 2012, 22(3): 577-591.

[18] Yu H, Zhao X L, Li Q. Genome-wide identification and characterization of long intergenic noncoding RNAs and their potential association with larval development in the Pacific oyster[J]. Scientific Reports, 2016, 6: 20796.

[19] Feng D D, Li Q, Yu H, *et al*. Transcriptional profiling of long non-coding RNAs in mantle of *Crassostrea gigas* and their association with shell pigmentation[J]. Scientific Reports, 2018, 8(1): 1436.

[20] Langmead B, Trapnell C, Pop M, *et al*. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome[J]. Genome Biology, 2009, 10(3): R25.

[21] Wen M, Shen Y, Shi S H, *et al*. miREvo: an integrative microRNA evolutionary analysis platform for next-

generation sequencing experiments[J]. BMC Bioinformatics, 2012, 13: 140.

[22] Friedlander M R, Mackowiak S D, Li N, *et al*. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades[J]. Nucleic Acids Research, 2012, 40(1): 37-52.

[23] Pertea M, Kim D, Pertea G M, *et al*. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown[J]. Nature Protocols, 2016, 11(9): 1650-1667.

[24] Guttman M, Garber M, Levin J Z, *et al*. *Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs[J]. Nature Biotechnology, 2010, 28(5): 503-510.

[25] 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.

[26] Kong L, Zhang Y, Ye Z Q, *et al*. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine[J]. Nucleic Acids Research, 2007, 35(Web Server issue): W345-W349.

[27] Mistry J, Bateman A, Finn R D. Predicting active site residue annotations in the Pfam database[J]. BMC Bioinformatics, 2007, 8: 298.

[28] Hansen T B, Venø M T, Damgaard C K, *et al*. Comparison of circular RNA prediction tools[J]. Nucleic Acids Research, 2016, 44(6): e58.

[29] Memczak S, Jens M, Elefsinioti A, *et al*. Circular RNAs are a large class of animal RNAs with regulatory potency[J]. Nature, 2013, 495(7441): 333-338.

[30] Gao Y, Zhang J Y, Zhao F Q. Circular RNA identification based on multiple seed matching[J]. Briefings in Bioinformatics, 2018, 19(5): 803-810.

[31] Witkos T M, Koscianska E, Krzyzosiak W J. Practical aspects of microRNA target prediction[J]. Current Molecular Medicine, 2011, 11(2): 93-109.

[32] 喻祎哲, 杨杰, 曾凡胜, 等. 日本血吸虫成虫的非编码RNA高通量测序分析[J]. 热带病与寄生虫学, 2017, 15(1): 31-35.
Yu Y Z, Yang J, Zeng F S, *et al*. Analysis of the non-coding RNA in adult *Schistosoma japonicum* by high-throughput sequencing[J]. Journal of Tropical Diseases and Parasitology, 2017, 15(1): 31-35(in Chinese).

[33] Zheng Z, Jiao Y, Du X D, *et al*. Computational prediction of candidate miRNAs and their potential functions in biomineralization in pearl oyster *Pinctada martensi*[J]. Saudi Journal of Biological Sciences, 2016, 23(3): 372-378.

[34] 张方, 胡子乔, 景炅婕, 等. 绵羊不同部位脂肪组织microRNA高通量测序及生物信息学分析[J]. 畜牧兽医学报, 2016, 47(6): 1093-1101.
Zhang F, Hu Z Q, Jing J J, *et al*. High-throughput sequencing and bioinformatics analysis on microRNAs expressed in adipose tissues of sheep[J]. Acta Veterinaria et Zootechnica Sinica, 2016, 47(6): 1093-1101(in Chinese).

[35] Bizuayehu T T, Lanes C F C, Furmanek T, *et al*. Differential expression patterns of conserved miRNAs and isomiRs during Atlantic halibut development[J]. BMC Genomics, 2012, 13: 11.

[36] He L, Wang Y L, Li Q, *et al*. Profiling microRNAs in the testis during sexual maturation stages in *Eriocheir sinensis*[J]. Animal Reproduction Science, 2015, 162: 52-61.

[37] Song Y N, Shi L L, Liu Z Q, *et al*. Global analysis of the ovarian microRNA transcriptome: implication for miR-2 and miR-133 regulation of oocyte meiosis in the Chinese mitten crab, *Eriocheir sinensis* (Crustacea: Decapoda)[J]. BMC Genomics, 2014, 15: 547.

[38] Chen S, McKinney G J, Nichols K M, *et al*. *In silico* prediction and *in vivo* validation of *Daphnia pulex* micrornas[J]. PLoS One, 2014, 9(1): e83708.

[39] Yan B, Guo J T, Zhu C D, *et al*. miR-203b: a novel regulator of MyoD expression in tilapia skeletal muscle[J]. Journal of Experimental Biology, 2013, 216: 447-451.

[40] Zhu X, Chen D X, Hu Y, *et al*. The microRNA signature in response to nutrient restriction and refeeding in skeletal muscle of Chinese perch (*Siniperca chuatsi*)[J]. Marine Biotechnology, 2015, 17(2): 180-189.

[41] Johnsson P, Lipovich L, Grandér D, *et al*. Evolutionary conservation of long non-coding RNAs; sequence, structure, function[J]. Biochimica et Biophysica Acta (BBA)-General Subjects, 2014, 1840(3): 1063-1071.

[42] Zhao W M, Mu Y L, Ma L, *et al*. Systematic

identification and characterization of long intergenic non-coding RNAs in fetal porcine skeletal muscle development[J]. Scientific Reports, 2015, 5: 8957.

[43] Suzuki H, Zuo Y H, Wang J H, *et al*. Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing[J]. Nucleic Acids Research, 2006, 34(8): e63.

[44] Zhang Y, Zhang X O, Chen T, *et al*. Circular intronic long noncoding RNAs[J]. Molecular Cell, 2013, 51(6): 792-806.

[45] Werfel S, Nothjunge S, Schwarzmayr T, *et al*. Characterization of circular RNAs in human, mouse and rat hearts[J]. Journal of Molecular and Cellular Cardiology, 2016, 98: 103-107.

[46] Hansen T B, Jensen T I, Clausen B H, *et al*. Natural RNA circles function as efficient microRNA sponges[J]. Nature, 2013, 495(7441): 384-388.

[47] Zuo J H, Wang Q, Zhu B Z, *et al*. Deciphering the roles of circRNAs on chilling injury in tomato[J]. Biochemical and Biophysical Research Communications, 2016, 479(2): 132-138.

[48] Xu S B, Xiao S J, Qiu C L, *et al*. Transcriptome-wide identification and functional investigation of circular RNA in the teleost large yellow croaker (*Larimichthys crocea*)[J]. Marine Genomics, 2017, 32: 71-78.

[49] Abdelmohsen K, Panda A C, De S, *et al*. Circular RNAs in monkey muscle: age-dependent changes[J]. Aging, 2015, 7(11): 903-910.

[50] Tian R R, Zheng Z, Huang R L, *et al*. miR-29a participated in nacre formation and immune response by targeting Y2R in *Pinctada martensii*[J]. International Journal of Molecular Sciences, 2015, 16(12): 29436-29445.

Bioinformatics analysis of regulatory non-coding RNA in gonad of *Crassostrea gigas*

WANG Xue [1,2,3], WANG Weijun [1,3,4*], LUO Qihao [1,2,3], SUN Guohua [4], FENG Yanwei [4], MA Jingjun [5], YANG Jianmin [1,3,4*]

(1. *National Demonstration Center for Experimental Fisheries Science Education*, *Shanghai Ocean University*, *Shanghai* 201306, *China*;
2. *Shanghai Engineering Research Center of Aquaculture*, *Shanghai Ocean University*, *Shanghai* 201306, *China*;
3. *School of Agriculture*, *Ludong University*, *Yantai* 264025, *China*;
4. *Shandong Marine Resource and Environment Research Institute*, *Yantai* 264006, *China*;
5. *Laishan Marine Fishery Station*, *Yantai* 264003, *China*)

Abstract: A plenty of non-coding RNAs (ncRNAs) have been identified through the application of high-throughput analysis of the transcriptome, and this has led to an intensive search for possible biological functions attributable to these transcripts. In this study, the gonad tissue of the two-year-old *Crassostrea gigas* of same family cultured in Rizhao Huanghai area was used to identify a large number of miRNA, lncRNA and circRNA by small RNA-seq and RNA-seq, and their biological characteristics were analyzed. The results showed that, with *Danio rerio* as a reference, 25-30 known miRNA matures and 51-63 known miRNA precursors was obtained, 53-71 new miRNA matures and 53-77 new miRNA precursors were predicted. The length of miRNA in *C. gigas* ranged from 18-26 nt, where the largest number was in the 20-22 nt and the first nucleotide position tended to be U. 2 302-2 349 known lncRNA transcripts were obtained, and 20 083-24 114 new lncRNA were predicted. Among them, the percentage of the intergenic lncRNA, intronic lncRNA and antisense lncRNA was 29.0%, 62.1%, and 8.9%, respectively. The data showed that genomic characteristics of lncRNA in *C. gigas* were similar to those of other eukaryotes. Compared with mRNA, the transcript and open reading frame of lncRNA were much shorter at length and much lower at expression level. 383 circRNA transcripts were obtained, of which the average percentage of 88.54% came from exon, 4.51% came from intronic and 6.95% came from intergenic. The data showed that the endogenous circRNA have a lot of miRNA target sites. This study revealed the basic biological characteristics of miRNA, lncRNA and circRNA in *C. gigas*. The results laid the foundation for the subsequent research on the expression rules and biological function of regulatory non-coding RNA in *C. gigas*.

Key words: *Crassostrea gigas*; miRNA; lncRNA; circRNA; RNA-seq

Corresponding authors: WANG Weijun. E-mail: wwj2530616@163.com;
YANG Jianmin. E-mail: ladderup@126.com

Funding projects: National Natural Science Foundation of China (31402298); Shandong Provincial Agricultural Variety Project (2017LZGC009); China Agriculture Research System (CARS-49)