

罗氏沼虾(*Macrobrachium rosenbergii*)卵巢发育不同时期转录组分析*

陈雪峰^{1,2} 王春琳^{1①} 顾志敏^{2①} 徐宾朋² 张宇飞² 慎佩晶² 程海华²
彭 菲² 李喜莲² 黄振远²

(1. 宁波大学海洋学院 宁波 315211; 2. 国家罗氏沼虾遗传育种中心 浙江省淡水水产研究所 湖州 313001)

摘要 为发掘罗氏沼虾(*Macrobrachium rosenbergii*)卵巢发育过程中的重要功能基因, 采用 Illumina HiSeqTM 4000 高通量测序平台对罗氏沼虾卵巢发育四个时期卵巢组织进行转录组测序。所测序列经质控、组装后比对到 NR、String、Swiss-prot、Pfam、GO、KEGG 数据库中注释, 并进行差异基因聚类分析。结果显示, 四个样品共产生 221580604 个 Clean reads, 拼接获得 95379 个 Unigenes。分别对四个样品测序文库进行两两比较, 共检测到 6605 个差异表达基因, 其中显著上调基因 2410 个, 显著下调基因 4195 个。GO 功能分类显示, 共有 6422 条 unigenes 可分为分子功能、细胞组分、生物学过程 3 大类 59 分支, 差异基因主要涉及糖类转运、氨基酸合成、酶活性、细胞膜组成等。通过 KEGG 通路注释, 共有 8423 条 unigenes 被注释, 涉及 184 种代谢途径, 有 7 个代谢通路与卵巢发育相关, 差异基因的 KEGG 富集结果显示, 药物代谢细胞色素 P450 通路、氨基丁酸神经元突触、鞘糖脂生物合成、氮素代谢等通路富集显著。此外, 对 SSR 与 SNP 分子标记进行了鉴定。研究结果为进一步探索罗氏沼虾卵巢发育机制提供了基础信息。

关键词 罗氏沼虾; 转录组测序; 卵巢发育; 差异基因

中图分类号 Q523+.8; S917 **doi:** 10.11693/hyhz20181000239

罗氏沼虾(*Macrobrachium rosenbergii*)由于食性广、病害少、生长快、营养价值高, 因此具有重要的经济价值, 它是世界上许多地区内陆水产养殖中重要的经济甲壳动物之一(陈雪峰等, 2012)。我国于 1976 年从日本引进该虾, 经 40 多年的发展, 我国罗氏沼虾相关产业处于世界第一。据不完全统计, 至 2016 年, 我国苗种年生产量达到 200 亿尾, 养殖面积超过 3 万公顷, 总产量超过 15 万吨。近年来整个养殖业又呈现出向北方及内陆地区扩展的趋势, 养殖面积的迅速扩增对苗种的稳定供应提出了挑战。然而目前人工育苗过程中往往会出现雌虾产卵相对不集中的现象, 从而导致幼体捕获量参差不齐、育苗周期

延长、生产成本加大。掌握与运用性腺发育规律是进行罗氏沼虾的人工繁殖、规模化育苗的基础。

卵巢是甲壳动物的生殖器官, 具有产生成熟卵泡、分泌雌性激素、促进受精等重要功能(陆洋等, 2018)。其发育是复杂的组织发生过程, 包括原始生殖细胞形成、性别决定分化、卵母细胞发育成熟等, 该过程涉及大量重要功能基因的转录调控(谢波波, 2017)。目前为止, 针对罗氏沼虾卵巢的研究多集中于其形状、解剖结构以及相关基因的功能分析(赵维信等, 1995; 刘金磊等, 2018), 系统的卵巢发育分子机制和调控机理尚未被完全阐明。

转录组测序(RNA sequencing)是一种可以获得细

*浙江省农业(水产)新品种选育重大科技专项, 2016C02055-2 号。陈雪峰, 博士研究生, 副研究员, E-mail: chenfenghailan@126.com

① 通信作者: 王春琳, 博士生导师, 教授, E-mail: wangchunlin@nbu.edu.cn; 顾志敏, 硕士生导师, 研究员, E-mail: guzhimin2006@163.com

收稿日期: 2018-10-17, 收修改稿日期: 2018-11-16

胞在某一状态下所能转录出来的 mRNA、small RNA 和 non-coding RNA (ncRNA)的高通量测序分析技术(王跃等, 2017)。与传统芯片技术相比, RNA-seq 能在全基因组范围内以单碱基分辨率检测和量化转录本, 具有信噪比高、灵敏度高、分析范围广、成本低、重复性好等优势, 从而全面剖析基因表达谱, 发现未知基因, 并精确地识别可变剪切位点及单核苷酸多态性(SNP)位点和非翻译区(UTR)。该技术已经应用于多种经济虾蟹类的性腺发育相研究(Liao *et al*, 2013; 刘金磊等, 2018)。本研究采用 RNA-Seq 测序对罗氏沼虾卵巢发育的四个代表性时期的卵巢组织进行转录组学研究, 期望通过对卵巢发育过程中细胞进程相关基因的分析, 从宏观上获取与卵巢发育和卵子发生相关的基因, 以全面了解在这四个关键时期卵巢组织中基因的表达变化情况, 从而为分子水平探索罗氏沼虾卵巢发育调控机制提供理论基础。

1 材料与方法

1.1 实验虾

在浙江国家级罗氏沼虾遗传育种中心育种基地的无特定病原(SPF)罗氏沼虾核心育种群体中采集大小规格基本一致的雌虾共 30 尾, 平均体重约 20g。实验虾暂养于水温 28℃、盐度 0.5‰ 的人工配置海水中。取下每尾虾的卵巢组织分成 2 份, 1 份存放液氮中, 1 份存放在组织固定液中备用。

1.2 方法

1.2.1 卵巢组织 RNA 提取 通过组织切片观察确定不同发育时期卵巢组织(颜素芬等, 2004)。不同发育期卵巢组织 RNA 的提取使用 Invitrogen 公司的 TRIzol[®] Reagent 试剂盒进行。抽提 RNA 满足 OD₂₆₀/OD₂₈₀ 为 1.8—2.2, RNA 总量 ≥ 5μg, 浓度 ≥ 200ng/μL, RIN ≥ 8, 进行下一步的文库构建。

1.2.2 文库构建与测序 采用 TruSeq RNA Sample Preparation Kit (Illumina)进行卵巢 cDNA 文库构建。取 5μg 总 RNA, 以带有 Oligo dT 的磁珠富集 mRNA, 加入 fragmentation Buffer, 将得到的 mRNA 随机断裂成 200bp 左右的小片段。以 mRNA 小片段为模板, 利用随机引物, 进行反转录, 合成双链 cDNA, 加入 End Repair Mix 补平末端, 在 3'末端加 A, 连接接头。连接产物纯化后, 进行 PCR 扩增(15 个循环), 扩增产物用 2%琼脂糖胶回收 200bp 左右的目的条带, 得到测序用双末端文库。通过 TBS380 Picogreen 对文库进行定量, 用 Illumina 公司的簇生成

试剂盒(TruSeq PE Cluster Kit v3-cBot-HS)进行桥式 PCR 扩增, 簇生成反应后, 用高通量测序平台 Illumina HiSeq[™] 4000 进行 2×151bp 测序, 获得的测序数据进行质量控制。

1.2.3 转录组 *de novo* 组装 Illumina 测序得到的原始图像数据经过 Base Calling 转化为序列数据, 对原始数据(Raw Data)进行过滤, 去除含测序接头的 reads、低质量 reads(质量值 Q ≤ 10 的碱基数占整个 read 的 20%以上)、N 率 ≥ 10% 的 reads 及长度过短序列, 得到高质量的测序数据(Clean reads)。使用 Trinity 软件在默认设置下对测序数据(Clean reads)进行组装, 首先将 reads 拼接成长片段 contigs, 然后去除冗余 contigs, 将 contigs 进行分组, 连接成两端不能再延长的转录片段(Unigenes)。

1.2.4 转录组注释 注释前, 利用 Trinity 软件提供的 ORF 预测流程对组装得到的所有转录本序列进行基因预测。通过 HMMER3 (Eddy, 1998)程序对组装的转录本在 Pfam 数据库中进行蛋白家族注释。设定期望值 E Value < 1e-5, 使用 BlastX (版本号: 2.2.25)软件对拼接所得的所有核苷酸序列分别与 NR(邓决决等, 2006)、String(柳军涛, 2017)、Swiss-prot (Apweiler *et al*, 2004)、GO (Ashburner *et al*, 2000)、KEGG (Kanehisa *et al*, 2004)数据库进行比对获得相应的注释信息。

1.2.5 差异基因的表达分析 使用软件 bowtie (Langmead *et al*, 2009)将 clean reads 比对到转录组序列上, 使用软件 RSEM 对 bowtie 比对结果进行表达量统计分析, 通过 RPKM (FPKM) 值 (Reads/ (Fragments) Per Kilobase of exon model per Million mapped reads), 来衡量基因的表达量水平。使用 edgeR 软件进行基因表达差异分析。不同样本比较后 FDR (False discovery rate)值 ≤ 0.05 且基因表达量差异倍数以 2 为底数的对数值 |log₂FC(Sample2/Sample1)| ≥ 1, 则该基因为显著差异表达基因。

1.2.6 GO 分类与 Pathway 富集分析 利用 Blast2go 软件在 GO 数据库中, 进行 Unigenes 参与的生物学过程、细胞组分、分子功能分类注释。利用 KEGG 数据库获得 Unigenes 的代谢通路注释信息。

1.2.7 SSR与SNP分子标记鉴定 采用MISA软件对Unigenes进行SSR检测, 设置待检测SSR核心序列的碱基数分别为1、2、3、4、5、6, 对应的核心序列的最少重复次数分别为10、6、5、5、5、5。此外, 以组装好的转录本为模板序列, 利用Samtools和

VarScan v.2.2.7软件寻找候选SNP。

2 结果

2.1 转录组测序结果

2.1.1 测序数据统计与组装 根据罗氏沼虾卵巢发育的4个不同时期的卵巢样品，建立了4个单独的测序文库，通过Illumina HiSeq™ 4000高通量测序，获得卵巢4个期Raw reads序列条数分别为47927784、70553888、57962050、54377814，经原始数据过滤后，得到Clean reads共221580604条，4个时期分别为45860872、67775028、55624580、52320124条。原始数据过滤后，Q20% (Phred数值大于20的碱基占总体碱基的百分比)有明显提高，达到98%左右。使用Trinity软件对转录组进行*de novo*组装，结果表明，组装得到的Unigene长度范围为201—14658bp，4个时期样品总共组装得到平均长度为766.29bp的Unigene序

列95379条，平均GC含量为38.96%(表1)。

表 1 组装结果统计
Tab.1 Assembly quality of *M. rosenbergii* ovary transcriptome

项目	数量
序列总数	95379
总碱基数	73087563
GC 含量(%GC)	38.96
最长转录本长度	14658bp
最短转录本长度	201bp
平均长度	766.29bp

所有Unigenes长度分布统计结果表明，Unigenes序列长度主要分布在1—4000bp，其中长度范围在1—1000bp有77247个，占80.98%，长度范围在1001—2000bp的有9409个，占9.87%，长度范围在2001—4000bp的有6284个，占6.59%。大于4000bp的Unigenes仅有2439个，占2.56%(图1)。

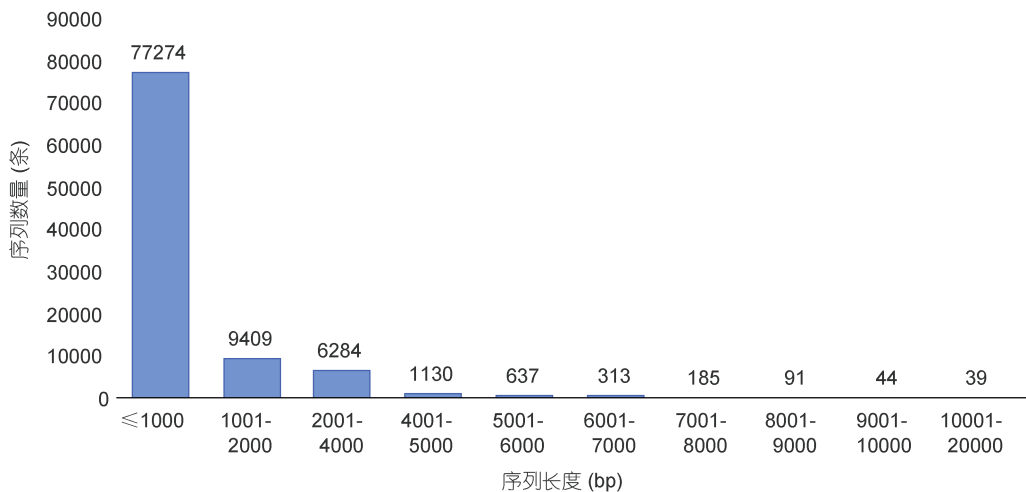


图 1 Unigenes 序列长度分布

Fig.1 The distribution of sequence length of the unigenes

2.1.2 转录组注释 通过与Pfam、NR、String、Swissprot、KEGG数据库比对，有52029个Unigenes与已知基因同源，占整个组装转录本的54.55%，Pfam、NR、String、Swissprot、KEGG数据库中同源序列条数分别为11308(11.86%)、16417(17.21%)、4281(4.49%)、11600(12.16%)、8423(8.83%)(表2)。物种同源性分析表明，本研究罗氏沼虾unigenes与NR库中无脊椎动物的基因具有同源性，与NR数据库比对的16417个unigenes中，与内华达古白蚁(*Zootermopsis nevadensis*)同源序列比例最高，达到10.99%(1804个)，其次是与淡水枝角水蚤(*Daphnia*

pulex)的同源序列比例，达到6.62%(1087个)，而与*Danaus plexippus*的同源序列比例最低，为0.98%(161个)(图2)。

表 2 基因注释量统计
Tab.2 Statistics of annotation gene number

数据库类别	unigenes 总数	注释 unigenes 量	比例
Pfam	95379	11308	11.86%
NR	95379	16417	17.21%
String	95379	4281	4.49%
Swissprot	95379	11600	12.16%
KEGG	95379	8423	8.83%

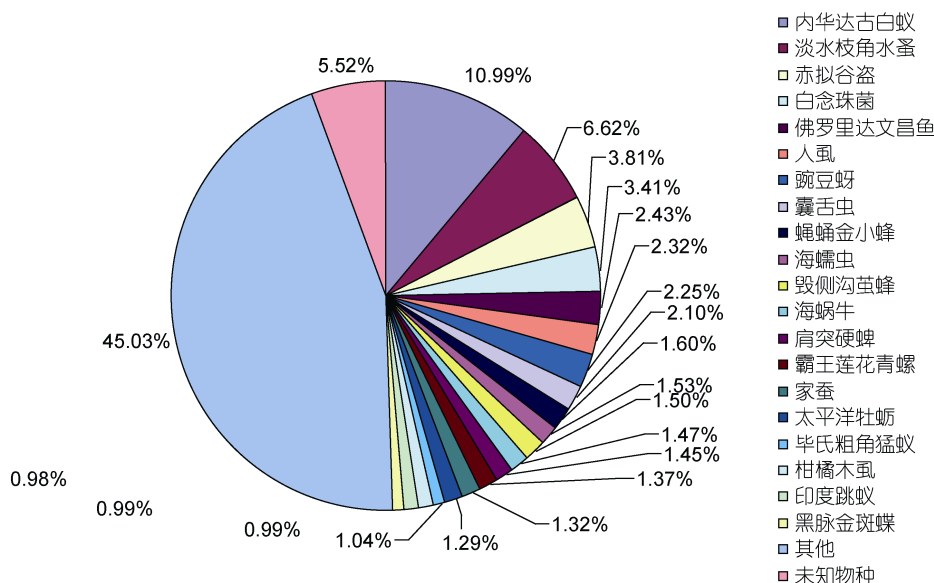


图2 转录组序列同源物种分布统计

Fig.2 Species distribution of the BLASTX results against nr database

2.2 差异表达基因

对罗氏沼虾卵巢发育 4 个时期转录组测序文库进行两两比较分析, 根据不同样本比较后 FDR(False discovery rate)值 ≤ 0.05 且基因表达量差异倍数以 2 为底数的对数值 $|\log_2FC(\text{Sample2}/\text{Sample1})| \geq 1$ 获得各个时期的基因差异表达情况, 共检测到 6605 个差异表达基因, 其中显著上调基因 2410 个, 显著下调基

因 4195 个。II 期与 III、IV、V 期比较, 表达量差异显著的基因数目分别为 1217(上调 329, 下调 888)、1435(上调 1080, 下调 355)、207(上调 87, 下调 120)个; III 期与 IV、V 期比较, 表达量差异显著的基因数目分别为 350(上调 179, 下调 171)、1508(上调 328, 下调 1180)个; IV、V 期比较, 表达量差异显著的基因数目为 1888(上调 407, 下调 1481)个。

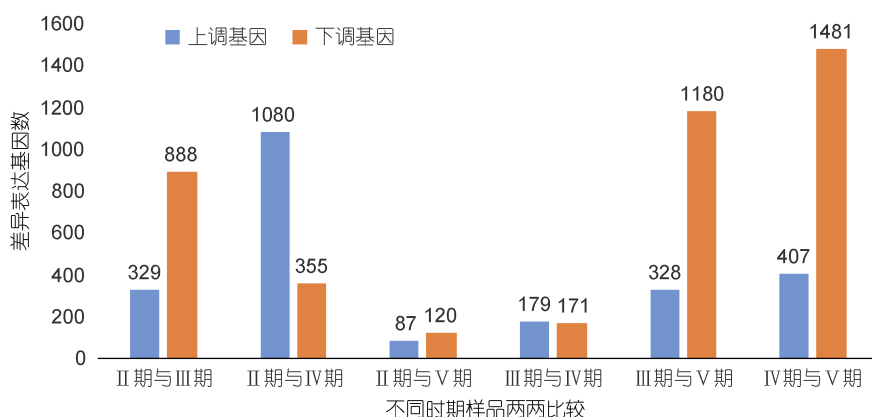


图3 差异表达基因数

Fig.3 The number of differently expressed unigenes

2.3 GO 功能分类

GO是一个标准化的基因功能分类体系, 对基因和蛋白功能进行统一的限定和描述。利用GO数据库, 可以对1个或1组基因按照其参与的生物学过程(Biological process, BP), 基因的分子功能(Molecular

function, MF)和细胞组分(Cellular component, CC)三方面进行分类注释。对全部序列进行GO功能分类, 共有6422条unigenes被归类到59个功能分类中(图4), 这些功能主要涉及代谢相关、生长发育调控、凋亡、生物合成、免疫防御、分子加工、信号转导及转录调控等。

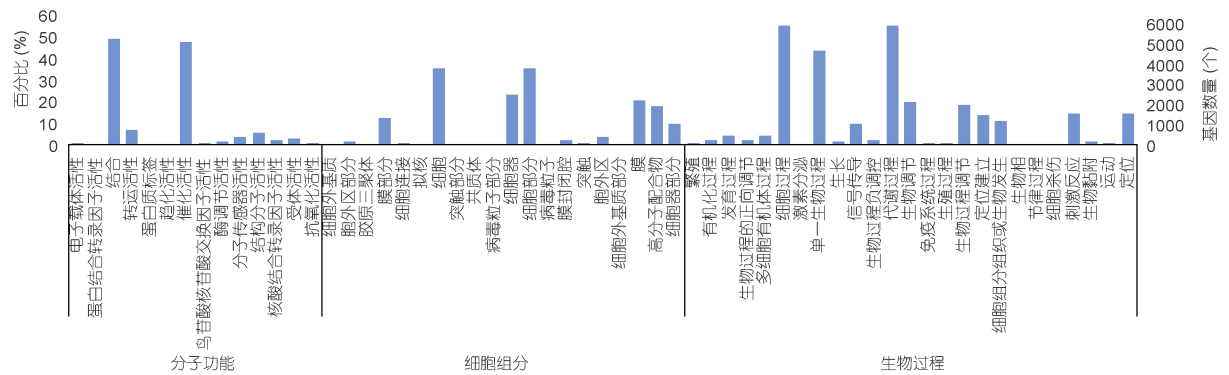


图 4 罗氏沼虾转录组 Unigenes 的 GO 功能分类

Fig.4 Gene Ontology classification of assembled unigenes in *M. rosenbergii*

按生物学过程(Biological processes, BP)进行差异基因分类注释, 结果 6 个比较组分别有 82、117、51、15、94、113 个 GO 类别出现富集, 按细胞组分(Cellular components, CC)进行分类注释, 结果 6 个比较组分别有 13、14、1、1、19、25 个 GO 类别出现富集, 按分子功能(Molecular functions, MF)进行差异基因分类注释, 结果 6 个比较组, 分别有 55、79、15、5、73、99 个 GO 类别出现富集。差异基因在糖类转运、代谢、氨基酸合成等生物学过程有显著变化, 影响分子功能主要体现在酶活性, 影响细胞组分主要涉及细胞膜的组成(表 3)。

表 3 差异基因富集前 5 的 GO 分类

Tab.3 Top 5 enrichment GO categories in the DGEs

比较组	GO 层级	GO 分类	集群频率(%)	P 值	类型
II 期 VS III 期	0015772	寡糖转运	0.16	0.000163	生物学过程
	0015766	二糖转运	0.16	0.000163	
	0030203	糖胺聚糖代谢过程	0.25	0.00123	
	0010992	泛素稳态	0.16	0.00159	
	0008150	生物学过程	6.98	0.00177	
	0043205	原纤维	0.08	0.0128	细胞组分
	0016010	肌营养不良相关糖蛋白复合物	0.08	0.0128	
	0001527	微纤维	0.08	0.0128	
	0019866	细胞器内膜	0.25	0.0144	
	0005743	线粒体内膜	0.25	0.0144	
	0016787	水解酶活性	2.88	1.69E-05	分子功能
	0003824	催化活性	5.42	3.55E-05	
	0003674	分子功能	8.38	5.11E-05	
	0046872	金属离子结合	1.81	0.00302	
	0043169	阳离子结合	1.81	0.00325	
II 期 VS IV 期	0008150	生物学过程	7.46	3.81E-05	生物学过程
	0044699	单生物过程	4.25	4.88E-05	
	0015772	寡糖转运	0.14	0.000226	
	0015766	二糖转运	0.14	0.000226	
	0051258	蛋白质聚合	0.35	0.00155	
	0005874	微管	0.35	0.00846	细胞组分
	0016010	肌营养不良相关糖蛋白复合物	0.07	0.015	
	0005576	胞外区	0.42	0.0174	
	0019866	细胞器内膜	0.21	0.0222	
	0005743	线粒体内膜	0.21	0.0222	分子功能
	0003674	分子功能	8.71	9.41E-07	

续表

比较组	GO 层级	GO 分类	集群频率(%)	P 值	类型
II 期 VS V 期	0003824	催化活性	5.64	1.02E-06	生物学过程
	0016787	水解酶活性	2.79	9.12E-06	
	0016788	酯酶作用下的水解酶活性	0.77	0.000353	
	0005200	细胞骨架的结构成分	0.35	0.00131	
	0044036	细胞壁大分子代谢过程	1.45	4.77E-05	
	0006542	谷氨酰胺生物合成过程	0.97	0.000255	生物学过程
	0009252	肽聚糖生物合成过程	0.97	0.000255	
	0000270	肽聚糖代谢过程	0.97	0.000255	
	0006022	氨基糖苷代谢过程	1.45	0.000299	
	0005576	胞外区	0.97	0.0356	细胞组分
	0016211	氨基糖苷代谢过程	0.97	0.000255	
	0016880	酰胺连接酶活性	0.97	0.000255	
	0004356	谷氨酸氨连接酶活性	0.97	0.000255	
	0003824	催化活性	7.73	0.00106	
III 期 VS IV 期	0016787	水解酶活性	4.35	0.00165	生物学过程
	0006687	鞘糖脂代谢过程	0.57	0.0022	
	0006664	糖脂代谢过程	0.57	0.00244	
	0008209	雄激素代谢过程	0.29	0.00367	
	0008210	雌激素代谢过程	0.29	0.00367	
	0006665	鞘脂代谢过程	0.57	0.00381	细胞组分
	0005874	微管	0.57	0.0376	
	0004065	芳基硫酸酯酶活性	0.29	0.00367	
	0004563	β -N-乙酰己糖胺酶活性	0.29	0.00733	
	0015929	己糖胺酶活性	0.29	0.011	分子功能
	0005200	细胞骨架的结构成分	0.57	0.0169	
	0008484	硫酸酯水解酶活性	0.28	0.0182	
	0030203	糖胺聚糖代谢过程	0.33	5.18E-06	
	0000270	肽聚糖代谢过程	0.27	1.88E-05	
III 期 VS V 期	0009252	肽聚糖生物合成过程	0.27	1.88E-05	生物学过程
	0006542	谷氨酰胺生物合成过程	0.27	1.88E-05	
	0044038	细胞壁大分子生物合成过程	0.27	3.97E-05	
	0019866	细胞器内膜	0.27	0.00362	
	0005743	线粒体内膜	0.27	0.00362	细胞组分
	0016020	内膜	1.39	0.00421	
	0045277	呼吸链复合物IV	0.2	0.0117	
	0031966	线粒体膜	0.27	0.0127	
	0016211	氨连接酶活性	0.27	1.88E-05	
	0004356	谷氨酸氨连接酶活性	0.27	1.88E-05	分子功能
	0016880	酰胺连接酶活性	0.27	1.88E-05	
	0005215	转运活性	1.12	0.000573	
	0015929	己糖胺酶活性	0.13	0.000742	
	0009084	谷氨酰胺家族氨基酸生物合成过程	0.26	2.09E-05	
IV 期 VS V 期	0000270	肽聚糖代谢过程	0.21	4.52E-05	生物学过程
	0009252	肽聚糖生物合成过程	0.21	4.52E-05	
	0006542	谷氨酰胺生物合成过程	0.21	4.52E-05	

					续表
比较组	GO 层级	GO 分类	集群频率(%)	P 值	类型
	0044038	细胞壁大分子生物合成过程	0.21	9.49E-05	细胞组分
	0005874	微管	0.32	0.00617	
	0019866	细胞器内膜	0.21	0.00794	
	0005743	线粒体内膜	0.21	0.00794	
	0044421	胞外区部分	0.32	0.0117	
	0031090	细胞器膜	0.32	0.0134	
	0003674	molecular_function	7.94	1.50E-05	分子功能
	0016211	氨连接酶活性	0.21	4.52E-05	
	0004356	谷氨酸氨连接酶活性	0.21	4.52E-05	
	0016880	酰胺连接酶活性	0.21	4.52E-05	
	0005200	细胞骨架的结构成分	0.37	9.25E-05	

2.4 KEGG 通路富集分析

KEGG是基因组破译方面的公共数据库, 该数据库从分子水平信息, 尤其是大型分子数据集生成的基因组测序和其他高通量实验技术的实用程序数据库资源中系统分析基因功能、联系基因组信息和功能信息的大型知识库。通过KEGG通路注释, 罗氏沼虾卵巢发育4个时期共有8423(8.83%)条unigenes被注释, 共涉及184种代谢途径, 基因数量排名前20位的KEGG途径如图5所显示, 其中unigenes数量最多的为核糖体(ribosome)(308条, 3.66%), 其余依次为RNA转运(RNA transport)(171条, 2.03%)、帕金森氏病

(Parkinson's disease)(165条, 1.96%)、溶酶体(Lysosome) (156条, 1.85%)、吞噬体(Phagosome)(151条, 1.79%)等, 参与基因数最少的为钙离子信号通路(calcium signaling pathway)(94条, 1.12%)。在184个代谢通路中, 有7个代谢通路可能与卵巢发育相关(表4), 分别为视黄醇代谢(Retinol metabolism)、孕酮介导卵母细胞成熟(Progesterone-mediated oocyte maturation)、胰岛素信号转导(Insulin signaling pathway)、细胞周期(Cell cycle)、Wnt信号通路(Wnt signaling pathway)、卵巢类固醇激素分泌(Ovarian steroidogenesis)、mTOR信号通路(mTOR signaling pathway)。

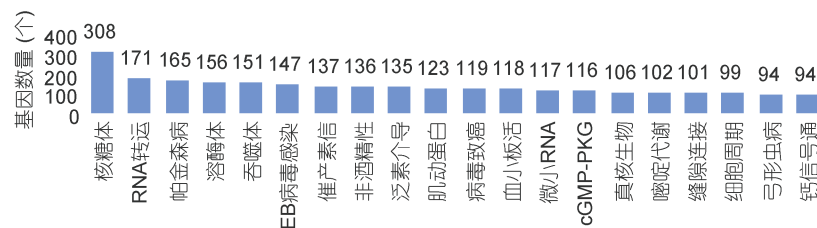


图 5 基因数目最多的前 20 个代谢通路

Fig.5 Top 20 metabolic pathways involving *M. rosenbergii* unigenes

将6个比较组的差异基因进行KEGG Pathway通路富集分析, 结果显示, 6个比较组分别有78、210、67、34、220、223个通路出现富集, 6个比较组中富集水平最高的通路分别为药物代谢细胞色素P450通路(Drug metabolism-cytochrome P450)、药物代谢细胞色素P450通路(Drug metabolism-cytochrome P450)、氨基丁酸神经元突触(GABAergic synapse)、鞘糖脂生物合成(Glycosphingolipid biosynthesis-glyco series)、氮素代谢(Nitrogen metabolism)、药物代谢细胞色素P450通路(Drug metabolism-cytochrome P450), 此外, 还有

细胞色素P450异物代谢(Metabolism of xenobiotics by cytochrome P450)、视黄醇代谢(Retinol metabolism)、谷氨酸能突触(Glutamatergic synapse)、缝管隙连接(Gap junction)等通路出现在富集前5的通路中(表5)。

2.5 SSR 与 SNP 多态性分析

见图 6 和图 7。共查找到 18592 个由 1—6 个核苷酸重复序列组成的 SSR 位点, 占 Unigenes 总数的比例为 19.49%。单核苷酸重复类型比例最高, 达 52.71%, 随后依次为二、三、四、五、六核苷酸重复类型, 占比分别为 32.61%、15.90%、0.80%、0.05%、

表 4 卵巢发育相关代谢通路注释统计		
Tab.4 The Pathways involved in ovary development of <i>M. rosenbergii</i>		
代谢通路	基因数量	代谢通路编号
视黄醇代谢	37	ko00830
孕酮介导卵母细胞成熟	70	ko04914
胰岛素信号通路	86	ko04910
细胞周期	99	ko04110
Wnt 信号通路	79	ko04310
卵巢类固醇生成	40	ko04913
mTOR 信号通路	36	ko04150

0.04%。在单、二、三核苷酸重复类型中, 数量最多的重复基元分别为 A/T (9407 个)、AG/GA (2244 个)、AAT/ATA/TAA (344 个), 数量最少的重复基元分别为 C/G (392 个)、CG/GC (12 个)、CGG/GGC/GCG (15 个)。

在整个转录组序列中, 共寻找到 111163 个候选 SNP 位点, 碱基转换位点 80379 个, 占 72.31%, 其中 A/G 转换 40126 个, A/G 转换 40253 个; 碱基颠换位点 30784 个, 占 27.69%, A/T、A/C、T/G、C/G 颠换位点分别为 8149、6661、6397、9577 个。

表 5 差异基因富集前 5 的 KEGG 通路列表					
Tab.5 Top 5 enrichment KEGG pathways in the DGEs					
比较组	通路 ID	信号通路	差异基因数	总基因数	P 值
II 期 VS III 期	ko00982	药物代谢-细胞色素 P450	6	44	0.001358385
	ko00980	细胞色素 P450 对外源化合物代谢的影响	6	49	0.002235939
	ko00830	视黄醇代谢	5	37	0.003534841
	ko05204	化学致癌	6	56	0.004107436
	ko04975	脂肪消化吸收	4	26	0.005959205
II 期 VS IV 期	ko00982	药物代谢-细胞色素 P450	7	44	0.000824461
	ko00980	细胞色素 P450 对外源化合物代谢的影响	7	49	0.001450139
	ko04975	脂肪消化吸收	5	26	0.002223842
	ko04640	造血细胞谱系	4	16	0.002749694
	ko05204	化学致癌	7	56	0.002885248
II 期 VS V 期	ko04727	γ-氨基丁酸能突触	3	66	0.005340201
	ko04724	谷氨酸能突触	3	78	0.008322176
	ko00910	氮代谢	2	25	0.008387551
	ko00220	精氨酸生物合成	2	40	0.019579873
	ko02020	双组分体系	2	44	0.023228944
III 期 VS IV 期	ko00603	球系列鞘糖脂生物合成	2	22	0.001602012
	ko00531	糖胺聚糖降解	2	24	0.001880508
	ko02010	ABC 转运蛋白	2	46	0.006305471
	ko04142	溶酶体	3	157	0.007261514
	ko00533	糖胺聚糖生物合成-硫酸角质素	1	8	0.022113416
III 期 VS V 期	ko00910	氮代谢	5	25	0.002238343
	ko05204	化学致癌	7	56	0.003520999
	ko04726	5-羟色胺突触	6	43	0.004166298
	ko02020	双组分体系	6	44	0.004611427
	ko00982	药物代谢-细胞色素 P450	6	44	0.004611427
IV 期 VS V 期	ko00982	药物代谢-细胞色素 P450	8	44	0.000472991
	ko00980	细胞色素 P450 对外源化合物代谢的影响	8	49	0.000888444
	ko05204	化学致癌	8	56	0.001913735
	ko00910	氮代谢	5	25	0.003872671
	ko04540	缝隙连接	10	101	0.006353131

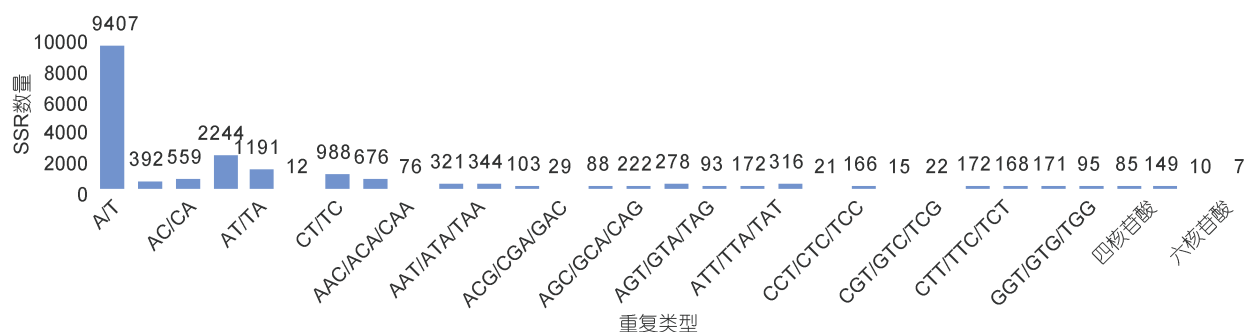


图6 罗氏沼虾转录组中 Unigenes 的 SSR 分析结果

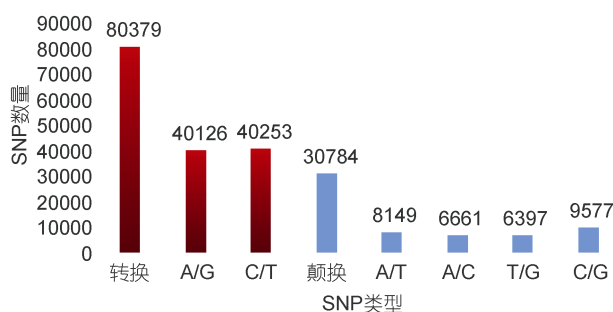
Fig.6 Summary of simple sequence repeats (SSRs) of unigenes in *M. rosenbergii* ovary transcriptome

图7 罗氏沼虾转录组中 Unigenes 的 SNP 分析结果

Fig.7 Summary of Single Nucleotide Polymorphisms (SNPs) of unigenes in *M. rosenbergii* ovary transcriptome

3 讨论

3.1 罗氏沼虾转录组测序

转录组测序技术作为一种快捷、高效的分子生物学研究手段,已经被广泛应用于重要经济水产甲壳动物的研究中(Liao *et al.*, 2013; 刘金磊等, 2018)。有关罗氏沼虾的转录组测序研究工作也已陆续开展, Jung等(2016)对罗氏沼虾肌肉、卵巢、精巢等组织制成混合样品进行了转录组测序,获得了787731个 reads, 经过拼接组装得到123534个 Unigenes, 并对参与肌肉生长的相关功能基因进行了挖掘; Rao等(2015)对弧菌感染与未感染的罗氏沼虾肝胰腺组织进行了转录组测序分析, 组装后, 对照组得到59050个 Unigenes, 感染组得到73946个 Unigenes, 显著表达差异基因14569个, 其中显著上调基因11446个, 显著下调基因3103个; Thanh等(2009)对养成规格大小具有明显差异的2个养殖场的罗氏沼虾的肝胰腺进行了转录组测序, 组装获得66953个 isogene, 2组样本鉴定得到了147个差异表达基因。本研究对4个时期卵巢样品进行转录组测序, 总共组装得到平均长度为766.29bp

的 Unigenes 序列95379条, 长度范围在1—1000bp 占80.98%, 长度范围在1001—2000bp 的占9.87%, 长度范围在2001—4000bp 的占6.59%, 长度大于4000bp 的占2.56%, 随着序列长度的增加, Unigenes 分布数量呈逐级递减, 没有明显中断, 表明本次转录组测序质量较高。本研究首次对罗氏沼虾卵巢发育4个时期的卵巢样品进行了转录组测序, 组装后, 进行了GO、KEEG、差异基因筛选等分析, 进一步丰富了罗氏沼虾转录组序列数据库, 为研究罗氏沼虾卵巢发育的分子机理提供了非常重要的信息。

3.2 转录组文库注释分析

本研究对罗氏沼虾卵巢组织 Unigenes 进行生物学功能注释, 有52029个 Unigenes 与已知基因同源, 注释率超过一半, 占整个组装转录本的54.55%, 与 NR 数据库比对结果显示, 有接近20%(17.21%)的序列同源性, 相对而言, 本文测定的转录组序列注释率较高, 原因可能是由于相近物种的基因注释信息在现有的国际公共数据库中收录较多, 尤其是罗氏沼虾自身的转录组测序研究已经得到开展, 从而导致了较高的注释率。同样, 尚有45.45%的序列没有注释成功, 这些基因序列与已有数据库中基因序列相似性较低, 可能为非编码RNA, 或是无功能结构域序列, 这些基因可能参与重要的生物学过程, 有待进一步研究。GO、KEEG 等数据库是为系统、深入了解基因的功能、基因序列同源性及其功能基因的代谢通路信息而开发的。本研究测定的转录组序列经 COG 功能分类, 共有2719个 unigenes 被归类到预测功能、蛋白质翻译、核糖体结构和生物合成相关基因等25个功能类别中; 通过 GO 功能分类, 共有6422条 unigenes 被归类到代谢相关、生长发育调控、凋亡、生物合成等59个功能分类中; KEGG 通路注释, 共有8423(8.83%)条 unigenes 被注释, 共涉及核糖体、RNA 转运、帕金森

氏病等184种代谢途径。这些功能基因及重要代谢通路途径的发现为后期在分子水平上对罗氏沼虾重要生物学功能的机制机理研究提供了较为全面的基础数据。

3.3 差异表达基因分析

经济甲壳动物的卵巢发育调控机制一直是研究热点,已有研究证实,罗氏沼虾的卵巢发育受神经多肽、类固醇激素、神经递质以及外源环境因子的综合调控(姚泊,1991;张艳,1999;黎东,2013;吴维福等,2014),本研究共检测到6605个差异表达基因,其中显著上调基因2410个,显著下调基因4195个,II期与III期比较有1217个差异表达基因,III期与IV期比较有350个差异基因,IV期与V期比较有1888个差异表达基因,这表明参与卵巢发育过程的调控基因较多,多种因素的共同作用影响控制卵巢发育的相关基因的表达与抑制,从而推动整个卵巢发育的生物学过程(陈变变,2014)。差异基因GO功能分析表明,糖类转运、代谢、氨基酸合成等生物学过程有显著变化,影响分子功能主要体现在酶活性,影响细胞组分主要涉及细胞膜的组成。差异基因KEGG富集结果显示药物代谢细胞色素P450通路、氨基丁酸神经元突触、鞘糖脂生物合成、氮素代谢等通路富集水平较高,这些代谢通路与蛋白质合成、脂类合成及神经递质转导有关,在卵巢发育过程中的重要作用有待进一步研究。

3.4 卵巢发育相关通路

本次转录组测序中,视黄醇代谢、孕酮介导卵母细胞成熟、胰岛素信号转导、细胞周期、Wnt信号通路、卵巢类固醇激素分泌、mTOR信号通路等7个与卵巢发育相关代谢通路被发现,其中细胞周期通路中的一些关键基因家族如Cdk基因家族、Cyclin基因家族调控卵巢发育在其他虾蟹类的研究中已有报道,Cyclin基因家族通过磷酸化Cdk基因家族的基因,从而调节细胞周期,郭松等(2014)研究表明,Cdk8基因对斑节对虾的卵巢发育启动具有重要作用,贾复龙等(2016)研究发现在三疣梭子蟹卵巢发育过程中,Cyclin H在卵巢发育I期与II期的表达量显著高于其他各期,房君江(2008)分析了罗氏沼虾cyclin B基因在卵子发生过程中的表达情况,表明了cyclin B基因可能与卵原细胞增殖和卵母细胞减数分裂成熟存在密切关系。此外,孕酮介导卵母细胞成熟通路中的cdc2、cdc25等基因在虾蟹类卵巢发育中的作用研究也有报道(周俊等,2012;Chen *et al.*, 2013)。视黄醇代谢、胰岛素信号转导、Wnt信号通路、卵巢类固醇激素分泌、

mTOR信号通路等的关键基因对卵巢发育的调控在其他生物中已有较多研究报道,这些通路中的基因在罗氏沼虾卵巢发育过程中如何调控关键生物学过程有待进一步的研究。

4 结论

采用 Illumina HiSeqTM 4000 高通量测序平台对罗氏沼虾卵巢发育四个时期卵巢组织进行转录组测序,获得 95379 个 Unigenes, 分别对四个样品测序文库进行两两比较,共检测到 6605 个差异表达基因,对差异表达基因进行了GO功能分类及KEGG通路注释,同时对测序序列中的SSR与SNP分子标记进行了鉴定。研究结果为进一步探索罗氏沼虾卵巢发育机制提供了基础信息。

参 考 文 献

- 王 跃,毛开云,王恒哲等,2017. 转录组学测序技术应用与市场分析. 生物产业技术, (5): 11—17
- 邓泱泱, 荔建琦, 吴松峰等, 2006. nr 数据库分析及其本地化. 计算机工程, 32(5): 71—73, 76
- 刘金磊, 邓思平, 江东能等, 2018. ERR-dsRNA 对罗氏沼虾卵巢中 ERR 及生殖相关基因表达的影响. 广东海洋大学学报, 38(3): 8—16
- 吴维福, 陈变变, 李郁娇等, 2014. 三丁基锡对罗氏沼虾血清中免疫酶活力的影响. 广东海洋大学学报, 34(3): 17—21
- 张 艳, 1999. 罗氏沼虾卵巢发育的内分泌调控. 广州: 中山大学硕士学位论文
- 陆 洋, 吴旭干, 柳梅梅等, 2018. 雌激素拮抗剂对三疣梭子蟹卵巢发育及相关基因表达的影响. 水产学报, 42(5): 653—662
- 陈变变, 2014. 壬基酚在罗氏沼虾体内代谢及其对性腺发育相关基因表达的影响. 湛江: 广东海洋大学硕士学位论文
- 陈雪峰, 杨国梁, 孔 杰等, 2012. 人工养殖与选育对罗氏沼虾遗传多样性的影响. 水生生物学报, 36(5): 866—873
- 周 俊, 周发林, 杨丽诗等, 2012. 斑节对虾 cdc25 基因的表达分析. 广东农业科学, 39(12): 148—150
- 房君江, 2008. 中华绒螯蟹、罗氏沼虾 CyclinB 基因 cDNA 克隆及其在卵子发生过程中表达分析. 上海: 上海海洋大学硕士学位论文
- 赵维信, 魏 华, 汪志强等, 1995. 人工诱导罗氏沼虾同步产卵与卵巢组织学研究. 水产学报, 19(4): 289—296
- 柳军涛, 2017. 基于高通量 RNA-seq 数据转录组拼接的算法研究. 济南: 山东大学博士学位论文
- 姚 泊, 1991. 不同温泉水水温对罗氏沼虾卵巢发育的影响. 广州师院学报(自然科学版), (1): 68—72
- 贾复龙, 孟宪亮, 刘 萍等, 2016. 三疣梭子蟹细胞周期蛋白 H 基因克隆及其在卵巢发育中的表达分析. 中国海洋大学学报, 46(7): 62—69
- 郭 松, 傅明骏, 赵 超等, 2014. 斑节对虾 CDK8 基因的克隆表达和原核表达及蛋白纯化. 见: 2014 年中国水产学会学术年会论文摘要集. 长沙: 中国水产学会

- 谢波波, 2017. 斑节对虾卵巢发育过程中 RBL 和下游 Chk1 基因的生物功能分析. 上海: 上海海洋大学硕士学位论文
- 黎 东, 2013. Cu^{2+} 、 Zn^{2+} 对罗氏沼虾生长以及卵巢发育的影响. 湛江: 广东海洋大学硕士学位论文
- 颜素芬, 姜永华, 2004. 南美白对虾卵巢结构及发育的组织学研究. 海洋湖沼通报, (2): 52—58
- Apweiler R, Bairoch A, Wu C H *et al*, 2004. UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, 32(S1): D115—D119
- Ashburner M, Ball C A, Blake J A *et al*, 2000. Gene ontology: tool for the unification of biology. *Nature Genetics*, 25(1): 25—29
- Chen J, Liu P, Li Z *et al*, 2013. The cloning of the cdk2 transcript and the localization of its expression during gametogenesis in the freshwater giant prawn, *Macrobrachium rosenbergii*. *Molecular Biology Reports*, 40(8): 4781—4790
- Eddy S R, 1998. Profile hidden Markov models. *Bioinformatics*, 14(9): 755—763
- Jung H, Yoon B H, Kim W J *et al*, 2016. Optimizing hybrid de novo transcriptome assembly and extending genomic resources for giant freshwater prawns (*Macrobrachium rosenbergii*): the identification of genes and markers associated with reproduction. *International Journal of Molecular Sciences*, 17(5): 690
- Kanehisa M, Goto S, Kawashima S *et al*, 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Research*, 32(S1): D277—D280
- Langmead B, Trapnell C, Pop M *et al*, 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3): R25
- Liao X L, Cheng L, Xu P *et al*, 2013. Transcriptome analysis of crucian carp (*Carassius auratus*), an important aquaculture and hypoxia-tolerant species. *PLoS One*, 8(4): e62308
- Rao R M, Zhu Y B, Alinejad T *et al*, 2015. RNA-seq analysis of *Macrobrachium rosenbergii* hepatopancreas in response to *Vibrio parahaemolyticus* infection. *Gut Pathogens*, 7(1): 6
- Thanh N M, Ponzone R W, Nguyen N H *et al*, 2009. Evaluation of growth performance in a diallel cross of three strains of giant freshwater prawn (*Macrobrachium rosenbergii*) in Vietnam. *Aquaculture*, 287(1—2): 75—83

TRANSCRIPTOME ANALYSIS OF *MACROBRACHIUM ROSENBERGII* OVARY IN FOUR DEVELOPMENT STAGES

CHEN Xue-Feng^{1,2}, WANG Chun-Lin¹, GU Zhi-Min², XU Bin-Peng², ZHANG Yu-Fei², SHEN Pei-Jing², CHENG Hai-Hua², PENG Fei², LI Xi-Lian², HUANG Zhen-Yuan²

(1. School of Marine Science, Ningbo University, Ningbo 315211, China; 2. National Genetic Breeding Center for *Macrobrachium rosenbergii*, Zhejiang Institute of Freshwater Fisheries, Huzhou 313001, China)

Abstract To analyze important functional genes of *Macrobrachium rosenbergii* during ovarian development, we conducted the transcriptome sequencing by Illumina HiSeqTM 4000. After quality control and assembly, the sequences acquired were blasted against NR, String, Swiss-prot, Pfam, GO and KEGG databases, and then the cluster analysis was performed. In total, 221580604 clean reads and 89529 Unigenes were obtained, and 6605 genes were differentially expressed, among which 2410 were up-regulated and 4195 down-regulated. According to the results of gene ontology functional enrichment analysis, 6422 Unigenes in the transcriptome library could be divided into 3 categories (molecular function, cellular component and biological process) with 59 branches. Significant differently-expressed genes were mainly distributed in the pathways like carbohydrate transport, amino acid synthesis, enzyme activity and the composition of the cell membrane. KEGG pathway analysis showed that 8423 Unigenes were given the pathway annotations and related to 184 pathways, and 7 pathways were associated with ovarian development. Analysis of differentially expressed genes revealed that several pathways, such as Drug metabolism—cytochrome P450, GABAergic synapse, Glycosphingolipid biosynthesis—globo series and Nitrogen metabolism, were significantly enriched. In addition, SSR and SNP molecular markers were identified. The results provide basic information for further exploration of the ovary development mechanism of *M. rosenbergii*.

Key words *Macrobrachium rosenbergii*; transcriptome sequencing; ovarian development; differentially expressed genes