



## 转录组测序筛选克氏原螯虾卵巢发育、免疫和生长相关基因

江红霞, 刘慧芬, 马 晓, 吴利敏, 孔祥会\*, 李学军\*

(河南师范大学水产学院, 水产动物疾病控制河南省工程实验室,  
河南省水产动物养殖工程技术研究中心, 河南 新乡 453007)

**摘要:** 为发掘克氏原螯虾卵巢发育、免疫和肌肉生长的重要功能基因, 采用 Illumina HiSeq™ 2500 高通量测序平台对克氏原螯虾的卵巢、肝胰腺和肌肉组织进行了转录组测序。所得序列经质控、组装后比对到 NR、Swiss-Prot、pfam、COG、GO 和 KEGG 数据库中注释, 并进行差异基因聚类分析。结果显示, 测序共获得了 53 006 个 unigene, 平均长度为 1 194 bp。对 3 个组织样品的测序文库进行两两比较, 发现在卵巢 vs. 肝胰腺中有差异表达基因 (differentially expressed gene, DEG) 20 382 个, 在肝胰腺 vs. 肌肉中有 DEG 12 753 个, 在肌肉 vs. 卵巢中有 DEG 21 629 个。GO 功能分类分析发现, 部分 DEG 被注释到繁殖 (reproduction)、繁殖过程 (reproduction process)、免疫系统过程 (immune system process) 和生长 (growth) GO 条目。KEGG pathway 分析显示, 一部分 DEG 在卵巢发育、免疫和肌肉生长相关的信号通路中得到了富集。根据 GO 功能分类和 KEGG 信号通路分析筛选出了大量与克氏原螯虾卵巢发育、免疫和肌肉生长相关的候选基因, 如卵黄蛋白原、卵黄蛋白原受体、Toll 样受体 2、Toll 样受体相互作用蛋白、肌肉生长抑制素和 5-羟色胺受体等。本研究结果丰富了克氏原螯虾的基因资源, 可为克氏原螯虾的遗传育种和免疫研究提供基础数据。

**关键词:** 克氏原螯虾; 转录组测序; 差异表达基因; 功能注释

**中图分类号:** S 966.12

**文献标志码:** A

克氏原螯虾 (*Procambarus clarkii*) 隶属于节肢动物门 (Arthropoda) 甲壳纲 (Crustacea) 软甲亚纲 (Malacostraca) 十足目 (Decapoda) 螯虾科 (Cambaridae) 原螯虾属 (*Procambarus*), 俗称淡水龙虾、小龙虾。克氏原螯虾原产于北美洲, 在 1929 年由日本引入我国江苏南京地区。该虾具有适应性广、繁殖力强、生长快、肉质鲜美及营养丰富等特点, 广受国内外养殖者和消费者青睐, 为我国主要经济虾类之一。中国渔业统计年鉴 (2019) 数据显示<sup>[1]</sup>, 2018 年全国克氏原螯虾养殖面积已超过 8 000 km<sup>2</sup>, 产量突破 163 万 t, 经济

总产值达到 2 685 亿元, 在淡水甲壳动物养殖中位居第一位。然而, 长期的高密度养殖和近亲繁殖也给克氏原螯虾的养殖业带来了许多负面的影响, 如生长速度慢、个体小、病害频发, 再加上性腺发育的不同步性给人工繁育带来的困难, 这严重制约着克氏原螯虾养殖业的进一步发展。

卵巢发育调控是水产甲壳动物繁育研究的热点。摘除眼柄是现如今克氏原螯虾人工繁育中最常用的卵巢促熟技术, 但摘除眼柄会造成激素紊乱, 卵巢成熟过快, 卵子质量低及亲虾

收稿日期: 2020-06-03 修回日期: 2020-08-22

资助项目: 国家自然科学基金 (31972812); 河南省重点科技攻关项目 (192102110081); 河南师范大学博士科研启动费支持项目 (qd17142); 河南师范大学个人科研项目结余经费资助专项 (20180531)

第一作者: 江红霞 (照片), 从事水生生物遗传资源研究, E-mail: jianghongxia2007@126.com

通信作者: 孔祥会, E-mail: xhkong@htu.edu.cn; 李学军, E-mail: xjli@htu.cn



死亡率高现象<sup>[2]</sup>, 因此, 寻找摘除眼柄之外的卵巢促熟技术是克氏原螯虾产业亟待解决的问题之一。而对卵巢发育相关基因进行鉴别, 从基因水平上调控克氏原螯虾卵巢发育则更具发展潜力。克氏原螯虾的免疫相关基因和免疫机制的研究对提高其抗病能力及抗病育种具有重要的实践意义。目前, 已有多种克氏原螯虾的免疫相关基因被鉴别出来, 例如组织蛋白酶 L 基因 (*cathepsin L*)<sup>[3]</sup>、肿瘤易感基因 101 (*TSG101*)<sup>[4]</sup>、前酚氧化酶基因 (*Po*)<sup>[5]</sup>、多巴脱羧酶基因 (*DDC*)<sup>[6]</sup> 和 *ficolin-like* 蛋白基因 (*FLP*)<sup>[7]</sup> 等。但这些基因尚不能使人们全面了解克氏原螯虾的免疫分子机制, 因此, 更多的免疫相关基因等待进一步挖掘。生长是经济水产动物的重要性能, 与水产养殖企业的生产率和利润率密切相关。甲壳动物的生长与蜕皮密切相关, 蜕皮诱导甲壳动物不同部位的肌肉生长。与甲壳动物蜕皮过程中肌肉生长的调控有关的物质主要包括蜕皮激素和蜕皮抑制激素、肌肉生长抑制素、泛素和热激蛋白等分子伴侣等<sup>[8]</sup>。找到影响肌肉生长发育的相关基因有利于提高甲壳动物的生产性能, 许多水产经济鱼类的肌肉生长发育相关基因的研究已经取得了长足进展<sup>[9-12]</sup>, 但是有关克氏原螯虾这一领域的研究还很薄弱。

近些年来, 转录组学技术广泛应用于水产动物繁殖、发育和免疫等各个方向的研究。转录组测序 (transcriptome sequencing) 是对某一物种的 mRNA 进行的高通量测序, 具有通量高、成本低、灵敏度高、无需基因组序列信息等优势, 可用于检测新的转录本、研究基因转录水平、非编码区域功能、转录本结构变异, 以及开发单核苷酸多态性 (single nucleotide polymorphism, SNP) 位点和简单重复序列 (simple sequence repeats, SSR) 等<sup>[13-14]</sup>。克氏原螯虾不同组织的转录组分析近几年已有报道, 如 Manfrin 等<sup>[15]</sup> 对克氏原螯虾的眼柄, Du 等<sup>[16]</sup> 对白斑综合征病毒 (WSSV) 感染的克氏原螯虾的肠道, Zhang 等<sup>[17]</sup> 对镉暴露下的克氏原螯虾的肝胰腺以及 Kang 等<sup>[18]</sup> 对克氏原螯虾的卵巢均进行了转录组分析, 为克氏原螯虾的神经内分泌、免疫、解毒和性腺成熟调控等研究提供了基础资料。本实验利用转录组测序技术构建了克氏原螯虾卵巢、肝胰腺和肌肉的转录组文库, 旨在进一步筛选出与克氏原螯虾卵巢发育、免疫和肌肉生长相关的信号通路基因, 为其繁育、抗病以及生长性能提高的研究

奠定理论基础。

## 1 材料与方法

### 1.1 实验材料

实验用克氏原螯虾购自新乡市水产品市场, 挑选 3 只卵巢发育处于卵黄合成期 (Ⅲ期) 的健康雌虾, 体质量约 20 g, 快速分离其卵巢、肝胰腺和腹部肌肉组织, 取样规格 100 mg, 贮存于液氮中备用。卵巢发育分期参考宋光同等<sup>[19]</sup>。

### 1.2 文库构建与高通量测序

转录组文库的制备及测序由上海美吉生物医药科技有限公司完成。从组织样品中提取总 RNA, 利用 Nanodrop 2000 对所提 RNA 的浓度和纯度进行检测, 琼脂糖凝胶电泳检测 RNA 完整性, Agilent 2100 测定 RNA 完整值 (RNA integrity number, RIN)。用带有 Oligo(dT) 的磁珠富集 mRNA, 并加入 fragmentation buffer 将 mRNA 处理成大小约 300 bp 的短片段。在逆转录酶的作用下, 加入六碱基随机引物 (random hexamers), 以 mRNA 为模板反转合成一链 cDNA, 随后二链合成, 形成稳定的双链结构。双链的 cDNA 结构为黏性末端, 加入 End Repair Mix 将其补成平末端, 随后在 3' 末端加上一“*A*”碱基, 用于连接 Y 字形的接头。加接头后进行 PCR 扩增, 最后用 AMPure XP beads 纯化扩增产物, 构建克氏原螯虾卵巢、肝胰腺和肌肉 cDNA 文库各 3 个。文库质检合格后, 上机进行 Illumina HiSeq™ 2500 测序, 测序读长为双端 150 bp。

### 1.3 从头组装和功能注释

Illumina HiSeq 测序得到原始数据 (raw reads) 通过去除低质量序列、去接头等质控过程得到高质量的质控数据 (clean reads), 最后将获得的 clean reads 用 Trinity<sup>[20]</sup> 软件对无参转录组进行从头组装, 使用 TransDecoder<sup>[21]</sup> 软件进行 unigene 的编码区序列及其对应氨基酸序列的预测。为获得全面的基因功能信息, 使用 BLAST<sup>[22]</sup> 软件将 unigene 序列与 NR、Swiss-Prot、Pfam、COG、GO 和 KEGG 数据库进行比对, 获得 unigene 的注释信息。

### 1.4 差异表达基因

采用 Bowtie<sup>[23]</sup> 将各样品测序得到的 reads 与 unigene 库进行比对, 根据比对结果, 结合 RSEM<sup>[24]</sup>

进行表达量水平估计。基因表达量的计算采用 FPKM(fragments per kilobase per million) 值<sup>[25]</sup>, 可直接用于比较不同样品间的基因表达差异。采用 DEGSeq<sup>[26]</sup> 进行样品组间的差异表达分析, 获得不同组织间的差异表达基因。在筛选过程中, 使用 IDEG6 软件进行卡方检验, 得到的  $P$  值通过多重假设检验 (false discovery ratio, FDR) 进行校正, 校正后取  $FDR < 0.01$  的, 并且在样品间 FPKM 比值 2 倍以上 ( $|Fold\ Change| \geq 2$ ) 的基因作为差异表达基因。随后对差异表达基因进行 KEGG 通路富集分析, GO 功能及 pfam 结构域分析, 获得差异表达基因相对应的各种注释信息。

1.5 实时荧光定量 PCR(qRT-PCR) 分析

分别从卵巢 vs. 肝胰腺、肝胰腺 vs. 肌肉和肌肉 vs. 卵巢比对组中各随机挑选出 5 个差异基因, 包括 3 个上调基因和 2 个下调基因, 用 Primer

premier 5 软件设计引物 (表 1), 进行 qRT-PCR 分析。各样品均有 3 个生物学重复, 按试剂盒 Total RNA kit II (Omega 公司) 的说明书提取 RNA 后采用 PrimeScript<sup>®</sup> RT reagent Kit with gDNA Eraser (TaKaRa) 试剂盒反转录为 cDNA, 以反转录产物为模板。以  $\beta$ -肌动蛋白 ( $\beta$ -actin) 为内参, 采用和 SYBR<sup>®</sup> Premix Ex Taq<sup>™</sup> II (TaKaRa, 大连) 试剂盒, 利用 CFX96 qRT-PCR 检测系统 (Bio-Rad, USA) 进行荧光定量检测。每个基因重复 3 次, 按  $2^{-\Delta\Delta C_t}$  法<sup>[27]</sup> 计算基因相对表达量, 使用 SPSS 20.0 软件对结果进行单因素方差分析,  $P < 0.05$  表示差异具有统计学意义。

2 结果

2.1 Unigene 的组装

利用 Illumina HiSeq<sup>™</sup> 2500 分别对克氏原螯

表 1 实验所用的引物

Tab. 1 primers used in the experiments

| 对照组<br>group                           | 差异表达基因<br>DEGs                                                          | 正向引物序列(5'-3')<br>forward(5'-3') | 反向引物序列(5'-3')<br>reverse(5'-3') |
|----------------------------------------|-------------------------------------------------------------------------|---------------------------------|---------------------------------|
| 卵巢vs. 肝胰腺<br>ovary vs. hepatopancreas  | 细胞周期蛋白A基因<br>cyclin A                                                   | CGGCTTCTGCTGTTGCTTT             | CCGTGCCACTTTGAACCTTTTG          |
|                                        | 前列腺素E <sub>2</sub> 合成酶2基因<br>prostaglandin E <sub>2</sub> synthase 2    | TGGTGGGGTTGGGGTTT               | CAGAAGAGTTTAGTATTGGAGTGCC       |
|                                        | 前列腺素E <sub>2</sub> 受体EP4基因<br>prostaglandin E <sub>2</sub> receptor EP4 | GCTGGTGGAACCTGTGCATC            | CCTCATACTCGTGGGCTCGT            |
|                                        | 金属硫蛋白基因<br>metallothionein                                              | GATTAGACCACCGCTTCACCC           | CCAGTCCAATACCAAACCAACA          |
|                                        | 拟血蓝蛋白1基因<br>cryptocyanin 1                                              | CAGGAAATCTGTTGAGTCGTCG          | CGTTGCCCTTGGGTATGAG             |
|                                        | Toll样受体相互作用蛋白基因<br>Toll-interacting protein                             | ACACCTCCCTGTTGCTACTCC           | GGCAGCTTCTTTGTTCCCTC            |
| 肝胰腺vs. 肌肉<br>hepatopancreas vs. muscle | 半胱天冬酶7基因<br>caspase 7                                                   | GGCTATGGAATACCTGTGGACC          | AACAAAGATGACGGCAAGACC           |
|                                        | 核因子 $\kappa$ B基因<br>nuclear factor kappa-B                              | CCCTCGACTATCTTCTCGCTTTA         | TCACCCACATCTGCCATCC             |
|                                        | 肌钙蛋白T基因<br>troponin T                                                   | AAGTTTGTCTCCAAGGGAGGTG          | CCAGTTTGATGATGAGGTTCCA          |
|                                        | 原肌球蛋白基因<br>tropomyosin                                                  | GCGGATACCTTGGAACAGCA            | CCTCACCTCAGCGTTAGATAGA          |
|                                        | 肌肉抑制素基因<br>myostatin                                                    | GCGGATACCTTGGAACAGCA            | CCTCACCTCAGCGTTAGATAGA          |
|                                        | 肌球蛋白重链基因<br>myosin heavy chain                                          | GGCTTCCAAGGAGATGAGTGA           | TGTTGCCGAGACGATAGAGC            |
| 肌肉vs. 卵巢<br>muscle vs. ovary           | 肌球蛋白轻链激酶基因<br>myosin light chain kinase                                 | CCCATCCCAGGCTCATACA             | AACAGTTCCCCACCATCCAC            |
|                                        | vasa样蛋白基因<br>vasa-like protein                                          | AGGAAAGAATCCGTCCACCTA           | TTGGCTCCGTCACAGTTAC             |
|                                        | 泛素-2基因<br>ubiquitin-2                                                   | GCACGAAAGATTAGGCAAGAAA          | CTCATCCGTTACAAAGGGGTC           |
|                                        | $\beta$ -肌动蛋白基因<br>$\beta$ -actin                                       | AGTAGCCGCCCTGGTTGTAGAC          | TTCTCCATGTCGTCCAGT              |
|                                        |                                                                         |                                 |                                 |
|                                        |                                                                         |                                 |                                 |

虾的卵巢、肝胰腺和肌肉进行转录组测序，分别获得 27.20、23.62 和 24.30 Gb 的测序数据。经过质控后，分别得到 26.50、23.34 和 23.63 Gb 的 clean data。碱基质量及组成分析显示，每组 GC 含量为 44.87%~49.59%，各样品 Q30 的碱基质量值比例均不小于 92.12%(表 2)。将测序获得的 clean reads 用 Trinity 软件从头组装后得到 53 006 个 unigenes，平均长度为 1 194 bp，N50 的长度(按照长度将组装 unigene 从大到小排序，累加 uni-

gene 的长度到总长度的一半时，对应 unigene 的长度)为 2 159 bp。其中长度为 0~500 bp 的 unigenes 数量最多，为 23 093 个，占总 unigenes 的 44%，长度为 0~500 bp 的 unigenes，为 12 543 个，占总 unigenes 的 24%，长度在 1 kb 以上的 unigene 有 17 370 个，占总 unigenes 的 32%(图 1)。将每个样本的 clean reads 与 Trinity 组装获得的参考序列进行比对(表 3)，6 个样本中能定位到组装转录本上的 clean reads 占两边拉平 63.87%~76.30%。

表 2 测序数据统计表  
Tab. 2 statistics table of sequencing data

| 样本名<br>sample         | 原始数据/个<br>raw reads | 原始碱基/bp<br>raw bases | 质控数据/个<br>clean reads | 质控碱基/bp<br>clean bases | 测序错误率/%<br>error rate | 测序质量值<br>大于20的<br>碱基比例/%<br>Q20 | 测序质量值<br>大于30的<br>碱基比例/%<br>Q30 | GC含量/%<br>GC content |
|-----------------------|---------------------|----------------------|-----------------------|------------------------|-----------------------|---------------------------------|---------------------------------|----------------------|
| 卵巢1 ovary 1           | 64 410 966          | 9 726 055 866        | 63 874 734            | 9 469 706 847          | 0.0268                | 97.36                           | 92.39                           | 44.87                |
| 卵巢2 ovary 2           | 56 573 504          | 8 542 599 104        | 56 053 912            | 8 312 129 208          | 0.0268                | 97.35                           | 92.36                           | 45.26                |
| 卵巢3 ovary 3           | 59 172 452          | 8 935 040 252        | 58 688 446            | 8 720 641 700          | 0.0269                | 97.31                           | 92.28                           | 45.69                |
| 肝胰腺1 hepatopancreas 1 | 56 497 996          | 8 531 197 396        | 56 003 840            | 8 290 893 885          | 0.0266                | 97.42                           | 92.57                           | 48.9                 |
| 肝胰腺2 hepatopancreas 2 | 50 247 732          | 7 587 407 532        | 49 675 676            | 7 342 454 386          | 0.0268                | 97.35                           | 92.44                           | 48.82                |
| 肝胰腺3 hepatopancreas 3 | 49 678 812          | 7 501 500 612        | 49 152 828            | 7 266 943 045          | 0.0271                | 97.22                           | 92.12                           | 47.38                |
| 肌肉1 muscle 1          | 52 456 792          | 7 920 975 592        | 51 958 444            | 7 706 330 814          | 0.027                 | 97.27                           | 92.25                           | 49.4                 |
| 肌肉2 muscle 2          | 52 859 342          | 7 981 760 642        | 52 263 398            | 7 747 633 436          | 0.0271                | 97.23                           | 92.16                           | 49.59                |
| 肌肉3 muscle 3          | 55 633 048          | 8 400 590 248        | 55 119 380            | 8 179 927 423          | 0.0269                | 97.33                           | 92.34                           | 49.04                |

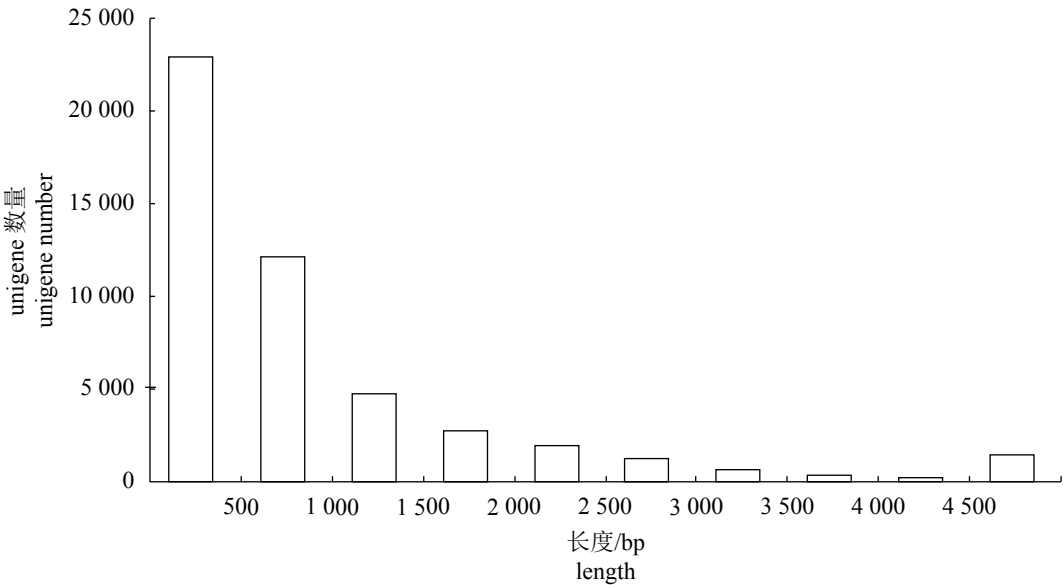


图 1 克氏原螯虾 unigenes 的长度分布  
Fig. 1 Length distribution of unigene sequence in *P. clarkii*



表 3 测序数据与组装结果比对统计表

Tab. 3 Comparison table of sequencing data and assembly results

| 样本名<br>samples        | 质控数据/个<br>clean reads | 比对到组装转录本上的<br>质控数据/个<br>mapped reads | 比对到组装转录本上的<br>质控数据的比例/%<br>mapped ratio |
|-----------------------|-----------------------|--------------------------------------|-----------------------------------------|
| 卵巢1 ovary 1           | 31 937 367            | 20 629 324                           | 64.59                                   |
| 卵巢2 ovary 2           | 28 026 956            | 18 234 386                           | 65.06                                   |
| 卵巢3 ovary 3           | 29 344 223            | 18 804 929                           | 64.08                                   |
| 肝胰腺1 hepatopancreas 1 | 28 001 920            | 17 956 182                           | 64.12                                   |
| 肝胰腺2 hepatopancreas 2 | 24 837 838            | 15 863 109                           | 63.87                                   |
| 肝胰腺3 hepatopancreas 3 | 24 576 414            | 16 003 756                           | 65.12                                   |
| 肌肉1 muscle 1          | 25 979 222            | 19 823 041                           | 76.30                                   |
| 肌肉2 muscle 2          | 26 131 699            | 19 839 284                           | 75.92                                   |
| 肌肉3 muscle 3          | 27 559 690            | 20 474 866                           | 74.29                                   |

2.2 unigene 功能注释

将 unigene 与 NR、Swiss-Prot、pfam、COG、GO 和 KEGG 共 6 个数据库进行比对, 得到注释的基因数总共为 15 956 个, 注释率为 30.1%。其中比对到 NR 库的基因数为 15 029 个, 注释率为 28.35%; 比对到 Swiss Prot 库的基因数为 11 571 个, 注释率为 21.83%; 比对到 pfam 库的基因数为 12 150 个, 注释率为 22.92%; 比对到 COG 库的基因数为 2 244 个, 注释率为 4.23%; 比对到 GO 库的基因数为 8 594 个, 注释率为 16.21%; 比对到 KEGG 库基因数为 8 462 个, 注释率为 15.96% (图 2)。拼接所得的 unigene 与 Nr 蛋白数据库进行同源性比对, 有 15 029 条与已知基因

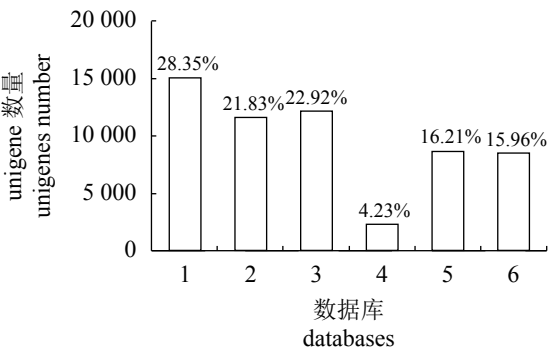


图 2 基因注释数量统计

1. NR 数据库, 2. Swiss-Prot 数据库, 3. Pfam 数据库, 4. COG 数据库, 5. GO 数据库, 6. KEGG 数据库

Fig. 2 Statistics of annotation gene number

1. annotated in NR, 2. annotated in Swiss-Prot, 3. annotated in Pfam, 4. annotated in COG, 5. annotated in GO, 6. annotated in KEGG

同源。相似序列所占比例最高的前 10 种物种分别是美洲钩虾 (*Hyalella azteca*, 31.28%), 澳大利亚干木白蚁 (*Cryptotermes secundus*, 5.11%)、内华达古白蚁 (*Zootermopsis nevadensis*, 3.83%)、雕纹刺尾蝎 (*Centruroides sculpturatus*, 2.91%)、美洲鲎 (*Limulus polyphemus*, 2.47%)、克氏原螯虾 (2.11%)、水螅 (*Hydra vulgaris*, 1.66%)、星状海葵 (*Nematostella vectensis*, 1.58%)、刺参 (*Apostichopus japonicus*, 1.50%) 和文昌鱼 (*Branchiostoma belcheri*, 1.47%) (图 3)。

2.3 差异表达基因

利用 DEseq 分析基因在克氏原螯虾卵巢、肝胰腺和肌肉中的差异表达, 发现 3 738 个基因在肝胰腺中特异表达, 11 003 个基因在肌肉中特异表达, 7 101 个基因在卵巢中特异表达 (图 4-a)。在卵巢 vs. 肝胰腺组中的差异表达基因 (DEG) 有 20 382 个, 其中上调 DEG 10 353 个 (50.79%), 下调 DEG 10 029 个 (49.21%); 肝胰腺 vs. 肌肉组中的 DEG 有 12 753 个, 其中上调 DEG 5 877 个 (46.08%), 下调 DEG 6 876 个 (53.92%); 肌肉 vs. 卵巢组中的 DEG 有 21 629 个, 其中上调 DEG 10 260 个 (47.44%), 下调 DEG 11 369 个 (42.56%) (图 4-b)。

2.4 差异表达基因的 GO 功能分类分析

差异表达基因的 GO 功能分类结果显示, 卵巢 vs. 肝胰腺组、肌肉 vs. 卵巢组和肝胰腺 vs. 肌肉组中的 DEG 均为注释到结合 (binding) GO 条目的最多, 分别有 2 222、2 156 和 1 251 个 DEG, 分

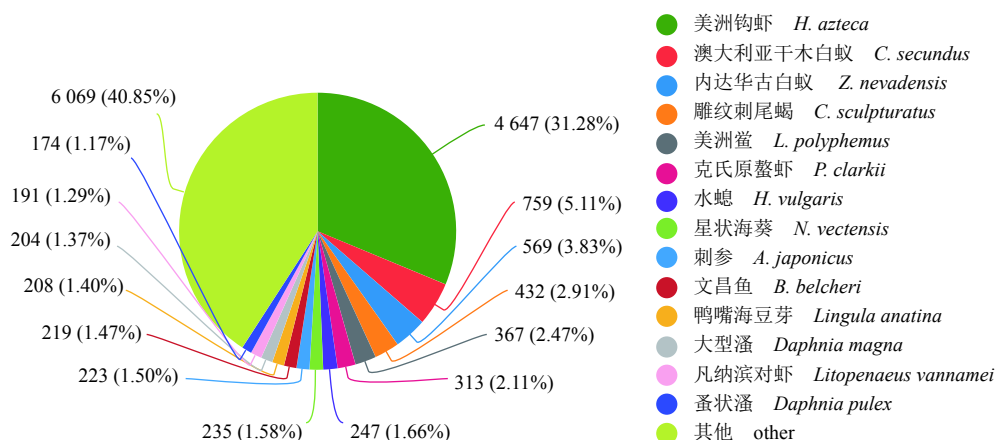


图 3 NR 注释物种分布

Fig. 3 Pie chart of NR annotated species distribution

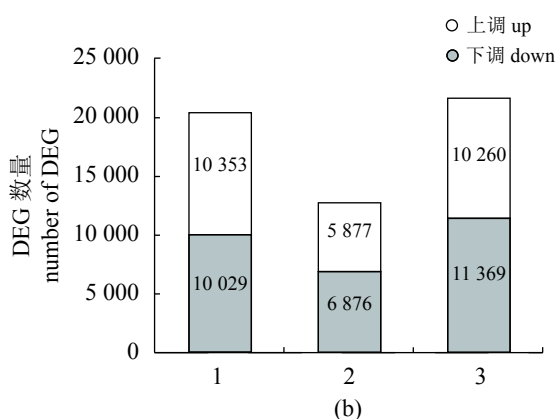
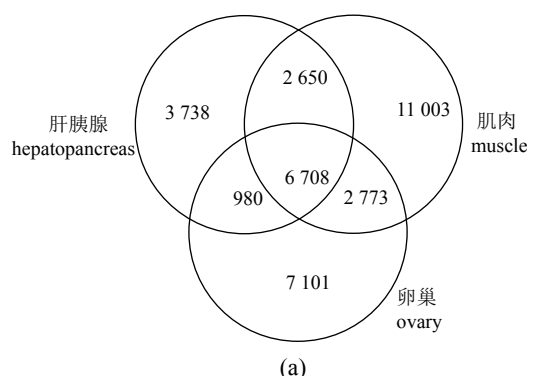


图 4 DEGs 统计图

(a) 样本间基因表达量 Venn 图; (b) 不同组织间 DEGs 统计图; 1. 卵巢 vs. 肝胰腺, 2. 肝胰腺 vs. 肌肉, 3. 肌肉 vs. 卵巢, 括号中数字为上下调 DEGs 的百分比

Fig. 4 DEGs statistics

(a) Venn diagram of gene expression between samples; (b) statistics of DEGs in different tissues; 1. ovary vs. hepatopancreas, 2. hepatopancreas vs. muscle, 3. muscle vs. ovary, the number in brackets are the percentages of up and down DEGs

别占该组总 DEG 数的 15.11%、14.25% 和 13.29%; 其次注释到催化活性 (catalytic activity) GO 条目的 DEG 分别有 1 795、1 772 和 1 244 个, 分别占该组总 DEG 数的 12.21%、11.71% 和 13.21%; 注释到细胞过程 (cellular process) GO 条目的 DEG 分别有 1 527、1 526 和 947 个, 分别占该组总 DEG 数的 10.38%、10.08% 和 10.06% (图 5)。另外, 卵巢 vs. 肝胰腺组、肌肉 vs. 卵巢组和肝胰腺 vs. 肌肉组中还分别有 18、14 和 10 个 DEG 注释到繁殖 (reproduction) 和繁殖过程 (reproduction process) GO 条目, 分别有 18、14 和 6 个 DEG 注释到生长 (growth) GO 条目, 分别有 19、21 和 10 个 DEG 注释到免疫系统过程 (immune system process) GO 条目。

## 2.5 差异表达基因 KEGG 信号通路富集分析

3 个比较组的差异基因在 KEGG 代谢通路中出现不同程度的富集。卵巢 vs. 肝胰腺组的 DEGs 富集到 311 条代谢通路, 发生显著富集的 KEGG 通路为 DNA 复制 (DNA replication)、核酸切除修复 (nucleotide excision repair)、泛素介导的蛋白水解 (ubiquitin mediated proteolysis)、孕酮介导的卵母细胞成熟 (progesterone-mediated oocyte maturation) 和错配修复 (mismatch repair) (图 6-a)。部分差异表达基因注释到卵母细胞减数分裂 (oocyte meiosis)、内分泌抵抗 (endocrine resistance)、卵巢类固醇生成 (ovarian steroidogenesis)、胰岛素信号通路 (insulin signaling pathway)、Wnt 信号通路 (Wnt signaling pathway)、类固醇激素合成 (steroid hormone biosynthesis)、胰岛素抵抗 (insulin resistance)、前列腺素合成通路 (prostaglandin synthesis)

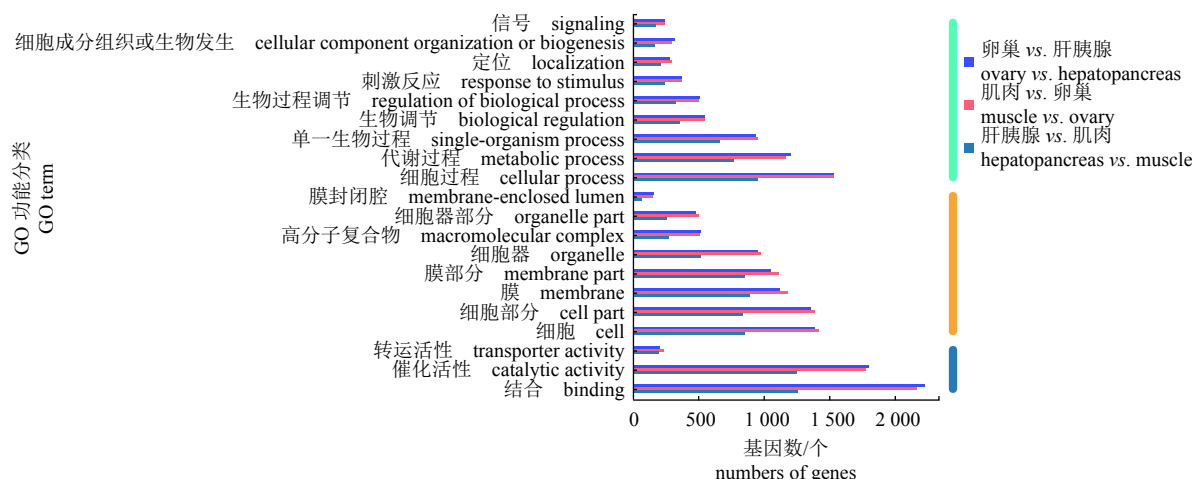


图 5 差异表达基因的 GO 功能分类 (二级)

右边浅蓝、橙和深蓝三种颜色分别表示 GO 分类的三大分支 (即生物过程、细胞组分和分子功能)

Fig. 5 Gene ontology classification (level 2) of putative functions of differentially expressed genes

The three colors of light blue, orange and dark blue on the right represent the three branches of the GO classification (ie. biological processes, cellular components and molecular functions)

pathway)、促性腺激素释放激素信号通路 (GnRH signaling pathway)、胰岛素分泌 (insulin secretion)、雌激素信号通路 (estrogen signaling pathway) 和催产素信号通路 (oxytocin signaling pathway) 等与卵巢发育相关的信号通路。

肝胰腺 vs. 肌肉组的 DEGs 富集到 313 条 KEGG 通路上, 主要显著富集到溶酶体 (lysosome)、蛋白酶体 (proteasome)、胞吞作用 (endocytosis)、过氧化物酶体 (peroxisome)、幽门螺杆菌 (*Helicobacter pylori*) 感染中的上皮细胞信号传导 (epithelial cell signaling in *Helicobacter pylori* infection)、甘油磷脂代谢 (glycerophospholipid metabolism) 等 KEGG 通路 (图 6-b)。部分 DEG 富集到 Toll 和 Imd 信号通路 (Toll and Imd signaling pathway)、Jak-STAT 信号通路 (Jak-STAT signaling pathway)、Rap1 信号通路 (Rap1 signaling pathway)、肿瘤坏死因子信号通路 (TNF signaling pathway)、趋化因子信号通路 (chemokine signaling pathway)、FcγR 介导的吞噬作用 (Fc gamma R-mediated phagocytosis)、NF-κB 信号通路 (NF-kappa B signaling pathway)、转化生长因子 β 信号通路 (TGF-beta signaling pathway)、细胞因子与细胞因子受体的相互作用 (cytokine-cytokine receptor interaction)、TRP 通道的炎性介质调节 (inflammatory mediator regulation of TRP channels)、Th1 和 Th2 细胞分化 (Th1 and Th2 cell differentiation)、吞噬体 (phagosome)、

RIG-I 样受体信号通路 (RIG-I-like receptor signaling pathway)、产生 IgA 的肠道免疫网络 (intestinal immune network for IgA production)、IL-17 信号通路 (IL-17 signaling pathway)、细菌侵袭上皮细胞 (bacterial invasion of epithelial cells) 和抗原处理和展示 (antigen processing and presentation) 等免疫相关信号通路。

肌肉 vs. 卵巢组的 DEGs 也富集到 313 条 KEGG 通路上, 发生显著富集的 KEGG 通路为 DNA 复制 (DNA replication)、核酸切除修复 (nucleotide excision repair) 和错配修复 (mismatch repair) (图 6-c)。部分 DEG 富集到胰岛素信号通路 (insulin signaling pathway)、胰岛素分泌 (insulin secretion)、肌动蛋白细胞骨架的调节 (regulation of actin cytoskeleton)、钙信号通路 (calcium signaling pathway)、PI3K-Akt 信号通路 (PI3K-Akt signaling pathway)、细胞周期 (cell cycle)、Notch 信号通路 (Notch signaling pathway)、细胞黏附分子 (cell adhesion molecules)、黏附连接 (adherens junction)、TGF-β 信号通路 (TGF-beta signaling pathway)、Hedgehog 信号通路 (Hedgehog signaling pathway)、内质网中的蛋白质加工 (protein processing in endoplasmic reticulum) 等与肌肉生长有关的信号通路。

## 2.6 卵巢发育、免疫和肌肉生长相关基因筛选

根据差异表达基因的 GO 功能分类和 KEGG

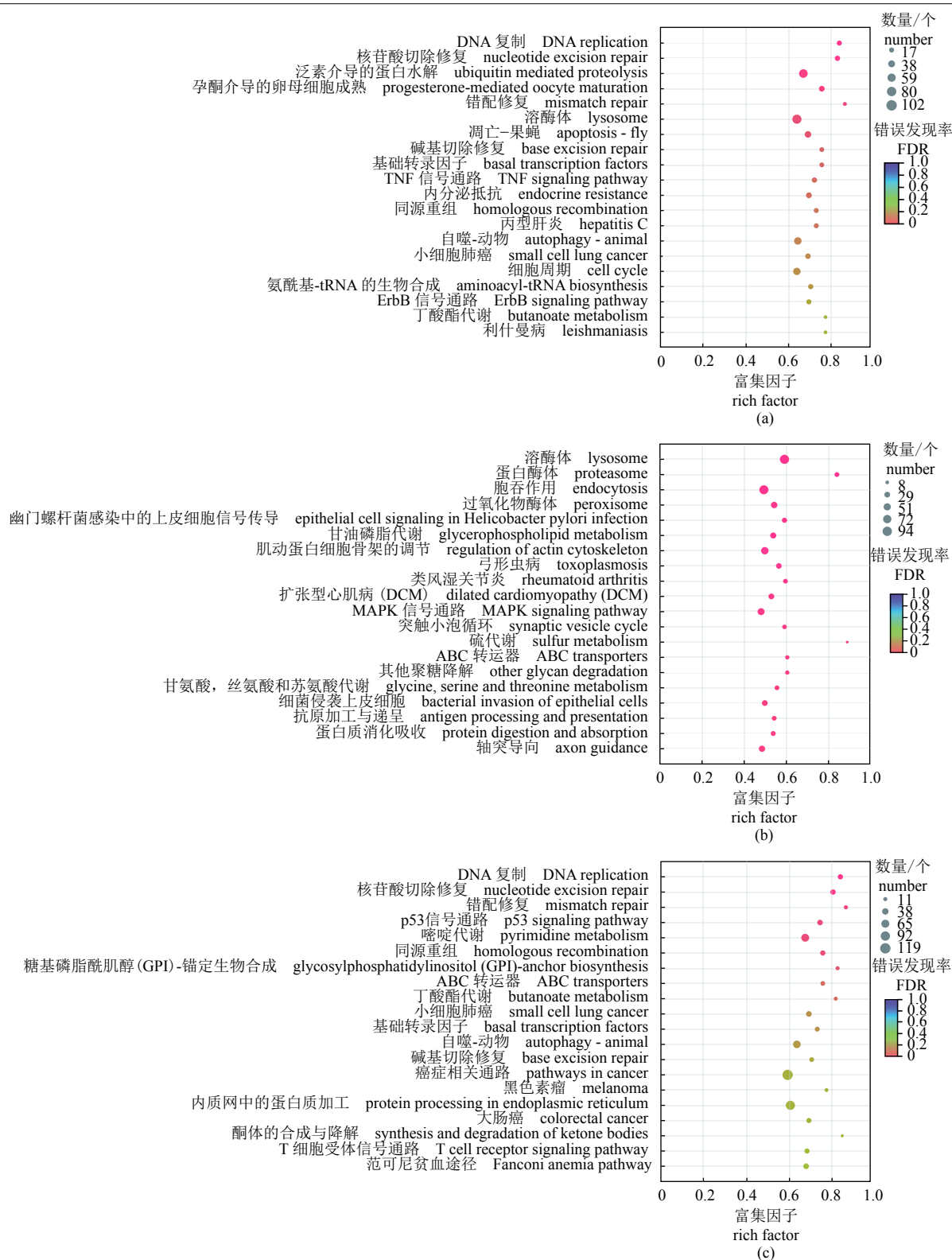


图 6 差异表达基因的 KEGG 富集分类

(a) 卵巢 vs. 肝胰腺; (b) 肝胰腺 vs. 肌肉; (c) 肌肉 vs. 卵巢

Fig. 6 KEGG classification of putative functions of differentially expressed genes

(a) ovary vs. hepatopancreas; (b) hepatopancreas vs. muscle; (c) muscle vs. hepatopancreas



信号通路富集分析, 筛选到主要的卵巢发育相关基因: 卵黄蛋白原 (vitellogenin) 基因、卵黄蛋白原受体 (vitellogenin receptor)、环氧合酶 (cyclooxygenase) 基因、前列腺素  $E_2$  合成酶 2 (prostaglandin  $E_2$  synthase 2)、前列腺素  $E_2$  受体 EP4 (prostaglandin  $E_2$  receptor EP4)、细胞色素 P450 (Cytochrome P450)、 $3\beta$ -羟类固醇脱氢酶 7 型 (3 beta-hydroxysteroid dehydrogenase type 7-like)、胰岛素样肽受体 (insulin-like peptide receptor) 和细胞分裂周期蛋白 (cell division cycle) 基因等, 其中的色素分散激素 A 肽 (pigment-dispersing hormone A peptides)、甲壳动物雌性特异荷尔蒙 (crustacean female specific hormone) 和卵黄膜外层蛋白 1 (vitelline membrane outer layer protein 1) 在 3 个组织的比对中是卵巢特异基因 (表 4)。

筛选到免疫相关基因主要有 Toll 样受体 2 (Toll-like receptor 2)、Toll 样受体相互作用蛋白 (toll-interacting protein)、cactus 蛋白 (cactus protein)、tube 蛋白 (tube)、pelle 蛋白 (pelle)、dorsal 蛋白 (dorsal)、domeless 蛋白 (domeless)、TAK1 结合蛋白 1 (TAK1 binding protein 1)、核因子  $\kappa$ B 激酶  $\epsilon$  亚型 2 抑制剂 (inhibitor of nuclear factor kappa-B kinase epsilon isoform 2)、骨髓分化因子 (myeloid differentiation factor)、凋亡蛋白抑制因子 (inhibitor of apoptosis protein)、泛素结合酶 E2 b (ubiquitin-conjugating enzyme E2 b)、转化生长因子  $\beta$  激活激酶 1 (transforming growth factor beta-activated kinase 1) 和抑制  $\kappa$ B 激酶  $\beta$  (IKK-beta) 基因等, 其中的内切葡聚糖酶 E-4-样蛋白 (endoglucanase E-4-like)、肝羧酸酯酶 1 样蛋白 (liver carboxylesterase 1-like) 和  $\alpha$ -维生素 E 转移蛋白样亚型 X4 (alpha-tocopherol transfer protein-like isoform X4) 在 3 个组织的比对中是肝胰腺特异基因 (表 5)。

筛选到的主要肌肉生长相关基因有肌球蛋白重链 (myosin heavy chain)、肌动蛋白 (actin)、肌球蛋白轻链激酶 (myosin light chain kinase)、肌肉特异性钙蛋白酶 (muscle-specific calpain)、肌钙蛋白 T (troponin T)、肌肉生长抑制素 (myostatin)、生长激素 (growth hormone)、胰岛素样生长因子 2 (insulin-like growth factor 2)、胰岛素样肽受体 (insulin-like peptide receptor)、表皮生长因子受体 (epidermal growth factor receptor)、5-羟色胺受体 (5-hydroxytryptamine receptor)、泛素结合酶 (ubiquitin-conjugating enzyme) 和苏氨酸/丝氨酸蛋白激酶

(threonine/serine protein kinase) 基因等, 其中的联结素样蛋白 (titin-like) 和肌膜相关蛋白样亚型 X2 (sarcolemmal membrane-associated protein-like isoform X2) 在 3 个组织的比对中是肌肉特异基因 (表 6)。

## 2.7 qRT-PCR 结果

为进一步验证高通量测序结果的可靠性, 通过 qRT-PCR 分别验证了卵巢 vs. 肝胰腺、肝胰腺 vs. 肌肉和肌肉 vs. 卵巢比对照组中各 5 个差异表达基因的表达, 将这些基因的差异表达与转录组测序结果相比较 (图 7), 结果显示, 被检测的基因表达趋势与转录组表达分析结果基本一致, 即证明转录组测序的结果可靠。

## 3 讨论

### 3.1 转录组测序分析

近年来, 转录组技术已广泛应用于甲壳动物研究中, 包括凡纳滨对虾<sup>[28]</sup>、中华绒螯蟹<sup>[29]</sup>、岸蟹 (*Carcinus maenas*)<sup>[30]</sup>、银色小长臂虾 (*Palaeomonetes argentinusfan*)<sup>[31]</sup>、墨吉明对虾<sup>[32]</sup>、三疣梭子蟹<sup>[33]</sup>等, 鉴别出了不同甲壳动物的繁殖、生长发育、免疫和环境胁迫等方面的相关基因。本研究对克氏原螯虾的卵巢、肝胰腺和肌肉组织进行了无参转录组测序, 得到 53 006 个 unigenes, 其中得到注释的基因数总共为 15 956 个, 注释率为 30.1%, 这种注释率低的现象可能是由于与克氏原螯虾相近的物种的基因注释信息在公共基因数据库中收录比较少, 缺乏充分可靠的参考依据。有 69.9% 的序列没有注释成功, 这些基因序列与已有公共数据库中的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, 这些基因也有可能参与重要的生物学过程, 但有待于进一步研究。其他甲壳动物的转录组测序也有类似注释率低的报道, 例如日本沼虾 (*Macrobrachium nipponense*) 转录组分析中只有 23.89% 的序列被注释<sup>[34]</sup>, 银色小长臂虾中有 24% 的序列被注释<sup>[30]</sup>, 凡纳滨对虾中有 37.28% 的序列被注释<sup>[35]</sup>。本研究得到的 unigenes 与 Nr 数据库比对结果显示: 与在公共数据库中已注册的克氏原螯虾序列匹配的 unigenes 仅有 2.11%, 这说明目前公共数据库中仍然缺乏克氏原螯虾的基因信息, 本研究是对克氏原螯虾基因信息库的一个有力补充。

表 4 卵巢发育相关基因

Tab. 4 Ovarian development related genes

| 基因编号<br>gene number   | Nr注释<br>Nr annotation                                                    | 匹配物种<br>matched species                  | Log <sub>2</sub> (卵巢/肝胰腺)<br>Log <sub>2</sub> (ovary/hepatop<br>ancreas) |
|-----------------------|--------------------------------------------------------------------------|------------------------------------------|--------------------------------------------------------------------------|
| TRINITY_DN10714_c2_g1 | 色素分散激素A肽 pigment-dispersing hormone A peptides                           | 内华达古白蚁<br><i>Z. nevadensis</i>           | 10.615                                                                   |
| TRINITY_DN9774_c3_g8  | 甲壳动物雌性特异荷尔蒙 crustacean female specific hormone                           | 日本囊对虾<br><i>Marsupenaeus japonicu</i>    | 9.779                                                                    |
| TRINITY_DN10215_c0_g1 | 细胞周期蛋白 B cyclin B                                                        | 克氏原螯虾<br><i>P. clarkii</i>               | 9.683                                                                    |
| TRINITY_DN14942_c1_g2 | 细胞周期蛋白 A cyclin A                                                        | 斑节对虾<br><i>P. monodon</i>                | 9.498                                                                    |
| TRINITY_DN10183_c4_g1 | 细胞分裂周期2蛋白 cell division cycle 2                                          | 克氏原螯虾<br><i>P. clarkii</i>               | 8.415                                                                    |
| TRINITY_DN9894_c7_g1  | 卵黄蛋白原受体 vitellogenin receptor                                            | 斑节对虾<br><i>Penaeus monodon</i>           | 8.110                                                                    |
| TRINITY_DN10156_c3_g2 | 卵黄膜外层蛋白1 vitelline membrane outer layer protein 1                        | 环带长角长蚬<br><i>Orchesella cincta</i>       | 7.324                                                                    |
| TRINITY_DN14708_c0_g1 | 细胞分裂周期蛋白20同源异构体X2<br>cell division cycle protein 20 homolog isoform X2   | 美洲钩虾<br><i>H. azteca</i>                 | 7.009                                                                    |
| TRINITY_DN13684_c3_g2 | 胰岛素样肽受体 insulin-like peptide receptor                                    | 烟草草虱<br><i>Bemisia tabaci</i>            | 6.989                                                                    |
| TRINITY_DN15119_c1_g2 | 细胞周期蛋白 E cyclin E                                                        | 斑节对虾<br><i>P. monodon</i>                | 6.933                                                                    |
| TRINITY_DN15482_c3_g4 | aurora激酶样蛋白 aurora kinase-like                                           | 美洲钩虾<br><i>H. azteca</i>                 | 6.500                                                                    |
| TRINITY_DN13872_c0_g2 | 前列腺素E <sub>2</sub> 受体 EP4 prostaglandin E <sub>2</sub> receptor EP4      | 美洲钩虾<br><i>H. azteca</i>                 | 5.854                                                                    |
| TRINITY_DN10450_c2_g7 | 金属磷酸酯酶MPPED 2样蛋白<br>metallophosphoesterase MPPED 2-like                  | 美洲钩虾<br><i>H. azteca</i>                 | 5.602                                                                    |
| TRINITY_DN14325_c0_g2 | G2/有丝分裂特异性细胞周期蛋白B3样亚型X2<br>G2/mitotic-specific cyclin-B3-like isoform X2 | 美洲钩虾<br><i>H. azteca</i>                 | 5.200                                                                    |
| TRINITY_DN10334_c1_g4 | 鸟嘌呤脱氨酶样蛋白 guanine deaminase-like                                         | 美洲钩虾<br><i>H. azteca</i>                 | 4.430                                                                    |
| TRINITY_DN9857_c2_g3  | Myt1蛋白 Myt1 protein                                                      | 三疣梭子蟹<br><i>Portunus trituberculatus</i> | 3.712                                                                    |
| TRINITY_DN13731_c9_g1 | 后期促进复合体亚基7样异构体X2<br>anaphase-promoting complex subunit 7-like isoform X2 | 雕纹刺尾蝎<br><i>C. sculpturatus</i>          | 3.277                                                                    |
| TRINITY_DN10144_c1_g2 | tolloid样蛋白1 tolloid-like protein 1                                       | 堆砂白蚁<br><i>Cryptotermes secundus</i>     | 3.177                                                                    |
| TRINITY_DN13254_c2_g1 | polo样激酶1 polo-like kinase 1                                              | 斑节对虾<br><i>P. monodon</i>                | 2.942                                                                    |
| TRINITY_DN11324_c1_g3 | 细胞分裂周期蛋白16同源物<br>cell division cycle protein 16 homolog                  | 澳大利亚干木白蚁<br><i>C. secundus</i>           | 2.775                                                                    |
| TRINITY_DN9415_c1_g5  | 前列腺素E <sub>2</sub> 合成酶2 prostaglandin E <sub>2</sub> synthase 2          | 斑节对虾<br><i>P. monodon</i>                | 2.704                                                                    |
| TRINITY_DN11782_c1_g1 | 丝裂原激活的蛋白激酶激酶<br>mitogen-activated protein kinase kinase                  | 拟穴青蟹<br><i>Scylla paramamosain</i>       | 2.382                                                                    |
| TRINITY_DN13560_c1_g8 | 有丝分裂纺锤体组装检查点蛋白MAD2B<br>mitotic spindle assembly checkpoint protein MAD2B | 澳大利亚干木白蚁<br><i>C. secundus</i>           | 2.047                                                                    |
| TRINITY_DN9257_c0_g1  | 细胞分裂周期蛋白23同源物<br>cell division cycle protein 23 homolog                  | 雕纹刺尾蝎<br><i>C. sculpturatus</i>          | 2.042                                                                    |
| TRINITY_DN15848_c1_g1 | 卵黄蛋白原 vitellogenin                                                       | 红螯螯虾<br><i>Cherax quadricarinatus</i>    | -11.056                                                                  |
| TRINITY_DN11333_c2_g1 | 3 $\beta$ -羟基类固醇脱氢酶7型<br>3 beta-hydroxysteroid dehydrogenase type 7-like | 褐飞虱<br><i>Nilaparvata lugens</i>         | -7.183                                                                   |
| TRINITY_DN12969_c2_g1 | 环氧化酶 cyclooxygenase                                                      | 美国龙虾<br><i>Homarus americanus</i>        | -2.177                                                                   |
| TRINITY_DN11648_c0_g3 | 丝氨酸/苏氨酸蛋白磷酸酶2A<br>serine/threonine-protein phosphatase 2A                | 斑节对虾<br><i>P. monodon</i>                | -2.052                                                                   |
| TRINITY_DN10613_c2_g1 | 细胞色素P450 Cytochrome P450                                                 | 多刺面颊小龙虾<br><i>Faxonius limosus</i>       | -2.052                                                                   |

表 5 免疫相关基因

Tab. 5 Immune related genes

| 基因编号<br>gene number   | Nr注释<br>Nr annotation                                                                               | 匹配物种<br>matched species               | Log <sub>2</sub> (肝胰腺/<br>肌肉)<br>Log <sub>2</sub><br>(hepatopancreas<br>/muscle) |
|-----------------------|-----------------------------------------------------------------------------------------------------|---------------------------------------|----------------------------------------------------------------------------------|
| TRINITY_DN12423_c1_g1 | 内切葡聚糖酶E-4-样蛋白 endoglucanase E-4-like                                                                | 美洲钩虾<br><i>H. azteca</i>              | 13.933                                                                           |
| TRINITY_DN12265_c0_g1 | $\alpha$ -维生素E转移蛋白样亚型X4<br>alpha-tocopherol transfer protein-like isoform X4                        | 美洲钩虾<br><i>H. azteca</i>              | 13.429                                                                           |
| TRINITY_DN14992_c1_g1 | 半胱天冬酶7 Caspase 7                                                                                    | 日本囊对虾<br><i>M. japonicus</i>          | 13.049                                                                           |
| TRINITY_DN12116_c3_g1 | 肝羧酸酯酶1样蛋白 liver carboxylesterase 1-like                                                             | 美洲钩虾<br><i>H. azteca</i>              | 13.063                                                                           |
| TRINITY_DN12599_c3_g1 | 核因子-kappa B NF- $\kappa$ B                                                                          | 克氏原螯虾<br><i>P. clarkii</i>            | 3.918                                                                            |
| TRINITY_DN15627_c0_g3 | tube蛋白 tube                                                                                         | 中华绒螯蟹<br><i>E. sinensis</i>           | 3.311                                                                            |
| TRINITY_DN14699_c0_g1 | 酪蛋白激酶II亚基 $\beta$ 同工型X3 casein kinase II subunit beta isoform X3                                    | 克氏原螯虾<br><i>P. clarkii</i>            | 2.887                                                                            |
| TRINITY_DN11409_c3_g1 | 骨髓分化因子 myeloid differentiation factor                                                               | 中华绒螯蟹<br><i>E. sinensis</i>           | 2.845                                                                            |
| TRINITY_DN15539_c0_g1 | Toll样受体相互作用蛋白 Toll-interacting protein                                                              | 中华绒螯蟹<br><i>Eriochelone sinensis</i>  | 2.659                                                                            |
| TRINITY_DN12174_c1_g4 | 凋亡抑制因子 inhibitor of apoptosis protein                                                               | 凡纳滨对虾<br><i>L. vannamei</i>           | 2.197                                                                            |
| TRINITY_DN14308_c1_g1 | 激活蛋白1 AP-1                                                                                          | 凡纳滨对虾<br><i>L. vannamei</i>           | 2.099                                                                            |
| TRINITY_DN12174_c1_g4 | 凋亡蛋白抑制剂 inhibitor of apoptosis protein                                                              | 凡纳滨对虾<br><i>L. vannamei</i>           | 2.097                                                                            |
| TRINITY_DN10085_c3_g1 | 核因子 $\kappa$ B激酶 $\epsilon$ 亚型2抑制因子<br>inhibitor of nuclear factor kappa-B kinase epsilon isoform 2 | 拟穴青蟹<br><i>S. paramamosain</i>        | 2.083                                                                            |
| TRINITY_DN11645_c1_g1 | Ras样GTP结合蛋白 ras-like gtp-binding protein                                                            | 黑蚁<br><i>Lasius niger</i>             | 2.073                                                                            |
| TRINITY_DN11992_c0_g1 | Bcl2样蛋白 Bcl2-like protein                                                                           | 欧氏沼虾<br><i>Macrobrachium olfersii</i> | 2.062                                                                            |
| TRINITY_DN15356_c2_g2 | TAK1结合蛋白1 TAK1 binding protein 1                                                                    | 凡纳滨对虾<br><i>L. vannamei</i>           | 2.003                                                                            |
| TRINITY_DN15627_c0_g8 | pelle蛋白 pelle                                                                                       | 凡纳滨对虾<br><i>L. vannamei</i>           | 2.001                                                                            |
| TRINITY_DN12707_c3_g1 | 转化生长因子 $\beta$ 激活激酶1<br>transforming growth factor beta-activated kinase 1                          | 凡纳滨对虾<br><i>L. vannamei</i>           | -5.926                                                                           |
| TRINITY_DN9077_c0_g1  | dorsal蛋白 dorsal                                                                                     | 日本囊对虾<br><i>M. japonicus</i>          | -5.157                                                                           |
| TRINITY_DN15049_c1_g7 | Myc蛋白 Myc                                                                                           | 大型蚤<br><i>Daphnia magna</i>           | -4.707                                                                           |
| TRINITY_DN13288_c1_g1 | domeless蛋白 domeless                                                                                 | 凡纳滨对虾<br><i>L. vannamei</i>           | -4.201                                                                           |
| TRINITY_DN11102_c0_g1 | cactus蛋白 cactus protein                                                                             | 凡纳滨对虾<br><i>Litopenaeus vannamei</i>  | -3.903                                                                           |
| TRINITY_DN12381_c2_g3 | p38丝裂原活化蛋白激酶 p38 MAPK                                                                               | 脊尾白虾<br><i>Palaemon carinicauda</i>   | -3.756                                                                           |
| TRINITY_DN12062_c2_g2 | Toll样受体2 Toll-like receptor 2                                                                       | 克氏原螯虾<br><i>P. clarkii</i>            | -2.681                                                                           |
| TRINITY_DN12176_c0_g3 | 泛素结合酶E2 b ubiquitin-conjugating enzyme E2 b                                                         | 中华绒螯蟹<br><i>E. sinensis</i>           | -2.021                                                                           |
| TRINITY_DN12933_c0_g5 | 抑制 $\kappa$ B激酶 $\beta$ IKK-beta                                                                    | 凡纳滨对虾<br><i>L. vannamei</i>           | -2.013                                                                           |
| TRINITY_DN14064_c2_g2 | $\beta$ -TrCP样蛋白 beta-TrCP-like                                                                     | 美洲钩虾<br><i>H. azteca</i>              | -2.061                                                                           |
| TRINITY_DN10016_c2_g1 | 丝裂原激活的蛋白激酶7<br>mitogen-activated protein kinase 7                                                   | 凡纳滨对虾<br><i>L. vannamei</i>           | -2.083                                                                           |

表 6 肌肉生长相关基因

Tab. 6 Muscle growth related genes

| 基因编号<br>gene number   | Nr注释<br>Nr annotation                                                          | 匹配物种<br>matched species                   | Log <sub>2</sub> (肌肉/卵巢)<br>Log <sub>2</sub> (muscle/ovary) |
|-----------------------|--------------------------------------------------------------------------------|-------------------------------------------|-------------------------------------------------------------|
| TRINITY_DN15747_c0_g3 | 肌球蛋白重链 myosin heavy chain                                                      | 美洲钩虾<br><i>H. azteca</i>                  | 16.182                                                      |
| TRINITY_DN14677_c4_g1 | 肌钙蛋白 Ttroponin T                                                               | 天空蓝魔虾<br><i>Cherax destructor</i>         | 16.119                                                      |
| TRINITY_DN15292_c1_g1 | 骨骼肌肌动蛋白6 skeletal muscle actin 6                                               | 美国龙虾<br><i>H. americanus</i>              | 14.816                                                      |
| TRINITY_DN10361_c4_g1 | 肌球蛋白轻链激酶 myosin light chain kinase                                             | 内华达古白蚁<br><i>Z. nevadensis</i>            | 12.539                                                      |
| TRINITY_DN14479_c1_g1 | 肌膜相关蛋白样亚型X2<br>sarcolemmal membrane-associated protein-like isoform X2         | 美洲钩虾<br><i>H. azteca</i>                  | 10.507                                                      |
| TRINITY_DN10765_c0_g2 | anoctamin 10样蛋白 anoctamin-10-like                                              | 美洲钩虾<br><i>H. azteca</i>                  | 10.221                                                      |
| TRINITY_DN9839_c0_g2  | 肌动蛋白 actin                                                                     | 玻璃海鞘<br><i>Ciona intestinalis</i>         | 6.218                                                       |
| TRINITY_DN13424_c0_g2 | 肌肉特异性钙蛋白酶 muscle-specific calpain                                              | 美国龙虾<br><i>H. americanus</i>              | 5.993                                                       |
| TRINITY_DN13797_c0_g3 | 联结素样蛋白 titin-like                                                              | 美洲钩虾<br><i>H. azteca</i>                  | 5.562                                                       |
| TRINITY_DN14746_c4_g1 | 肌肉生长抑制素 myostatin                                                              | 墨吉明对虾<br><i>Fenneropenaeus merguensis</i> | 5.103                                                       |
| TRINITY_DN10224_c1_g1 | 磷脂酰肌醇3-激酶调节亚基α样<br>phosphatidylinositol 3-kinase regulatory subunit alpha-like | 美洲钩虾<br><i>H. azteca</i>                  | 4.230                                                       |
| TRINITY_DN15048_c0_g1 | cGMP依赖性蛋白激酶 cGMP dependent protein kinase                                      | 澳大利亚干木白蚁<br><i>C. secundus</i>            | 3.893                                                       |
| TRINITY_DN10348_c2_g2 | spalt-major-样同源蛋白 homeotic protein spalt-major-like                            | 美洲钩虾<br><i>H. aztec</i>                   | 3.219                                                       |
| TRINITY_DN12293_c1_g1 | 泛素结合酶 ubiquitin-conjugating enzyme                                             | 凡纳滨对虾<br><i>L. vannamei</i>               | 2.591                                                       |
| TRINITY_DN10870_c2_g3 | S期激酶相关蛋白1 S-phase kinase-associated protein 1                                  | 中华绒螯蟹<br><i>E. sinensis</i>               | 2.537                                                       |
| TRINITY_DN14038_c1_g1 | 5-羟色胺受体 5-hydroxytryptamine receptor                                           | 吊丝虫<br><i>Plutella xylostella</i>         | 2.411                                                       |
| TRINITY_DN10292_c1_g5 | 锌指MYM型蛋白1 zinc finger MYM-type protein 1                                       | 隆头蛛<br><i>Stegodyphus mimosarum</i>       | 2.094                                                       |
| TRINITY_DN10215_c0_g1 | 细胞周期蛋白 B cyclin B                                                              | 克氏原螯虾<br><i>P. clarkii</i>                | -10.411                                                     |
| TRINITY_DN13684_c3_g2 | 胰岛素样肽受体 insulin-like peptide receptor                                          | 烟草粉虱<br><i>B. tabaci</i>                  | -10.089                                                     |
| TRINITY_DN10191_c1_g3 | 苏氨酸/丝氨酸蛋白激酶 threonine/serine protein kinase                                    | 美洲钩虾<br><i>H. azteca</i>                  | -8.849                                                      |
| TRINITY_DN10183_c4_g1 | 细胞分裂周期2蛋白 cell division cycle 2                                                | 克氏原螯虾<br><i>P. clarkii</i>                | -5.988                                                      |
| TRINITY_DN10980_c0_g6 | 细胞周期蛋白 H cyclin H                                                              | 拟穴青蟹<br><i>S. paramamosain</i>            | -4.220                                                      |
| TRINITY_DN13449_c1_g2 | 胰岛素样生长因子2 insulin-like growth factor 2                                         | 澳大利亚干木白蚁<br><i>C. secundus</i>            | -3.771                                                      |
| TRINITY_DN11337_c2_g3 | 肌肉特异性蛋白20 muscle-specific protein 20                                           | 茧蜂<br><i>Microplitis demolitor</i>        | -3.345                                                      |
| TRINITY_DN10962_c2_g1 | Abl酪氨酸蛋白激酶 tyrosine-protein kinase Abl-like                                    | 美洲钩虾<br><i>H. azteca</i>                  | -3.066                                                      |
| TRINITY_DN13353_c1_g4 | 生长激素 growth hormone                                                            | 美洲钩虾<br><i>H. azteca</i>                  | -2.962                                                      |
| TRINITY_DN13174_c1_g2 | 单酸甘油酯脂肪酶 monoglyceride lipase                                                  | 大型蚤<br><i>D. magna</i>                    | -2.851                                                      |
| TRINITY_DN10841_c2_g2 | 转录因子Dp-1 transcription factor Dp-1                                             | 澳大利亚干木白蚁<br><i>C. secundus</i>            | -2.775                                                      |
| TRINITY_DN11626_c2_g2 | 表皮生长因子受体 epidermal growth factor receptor                                      | 美洲钩虾<br><i>H. azteca</i>                  | -2.687                                                      |
| TRINITY_DN11531_c4_g3 | 腺苷酸环化酶 adenylyl cyclase                                                        | 地蟹<br><i>Gecarcinus lateralis</i>         | -2.076                                                      |



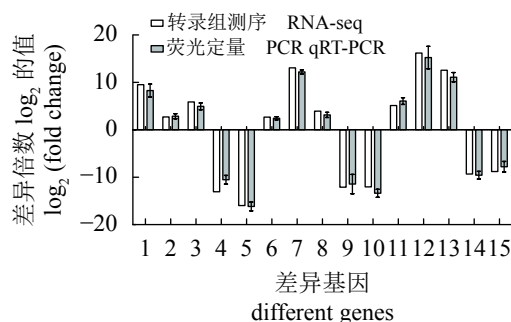


图7 15个差异表达基因的qRT-PCR与转录组的比较分析

1. 细胞周期蛋白A基因, 2. 前列腺素E<sub>2</sub>合成酶2基因, 3. 前列腺素E<sub>2</sub>受体EP4基因, 4. 金属硫蛋白基因, 5. 拟血蓝蛋白1基因, 6. Toll样受体相互作用蛋白基因, 7. 半胱天冬酶7基因, 8. 核因子κB基因, 9. 肌钙蛋白T基因, 10. 原肌球蛋白基因, 11. 肌肉抑制素基因, 12. 肌球蛋白重链基因, 13. 肌球蛋白轻链激酶基因, 14. vasa样蛋白基因, 15. 泛素-2基因; 1、2、3、4和5为卵巢vs.肝胰腺组的差异基因; 6、7、8、9和10为肝胰腺组vs.肌肉组的差异基因; 11、12、13、14和15为肌肉vs.卵巢组的差异基因

Fig. 7 Comparison of fifteen DEGs by qRT-PCR and transcriptome analysis

1. cyclin A, 2. prostaglandin E<sub>2</sub> synthase 2, 3. prostaglandin E<sub>2</sub> receptor EP4, 4. metallothionein, 5. cryptocyanin 1, 6. toll-interacting protein, 7. caspase 7, 8. nuclear factor kappa-B, 9. troponin T, 10. tropomyosin, 11. myostatin, 12. myosin heavy chain, 13. myosin light chain kinase, 14. vasa-like protein, 15. ubiquitin-2; 1, 2, 3, 4 and 5 are the DEGs in ovary vs. hepatopancreas group; 6, 7, 8, 9 and 10 are the DEGs in hepatopancreas vs. muscle group; 11, 12, 13, 14 and 15 are the DEGs in muscle vs. ovary group

### 3.2 部分卵巢发育相关通路及基因分析

卵巢发育调控机制一直是经济甲壳动物研究的热点, 已有研究证实, 甲壳动物的卵巢发育受类固醇激素、神经多肽、神经递质以及外源环境因子的综合调控<sup>[36-40]</sup>。本研究的卵巢vs.肝胰腺组中的DEGs显著富集至孕酮介导的卵母细胞成熟KEGG通路, 在这个通路中筛选到了卵黄蛋白原(Vg)和卵黄蛋白原受体(Vgr)基因。Vg是卵黄素的前体, 在卵巢内外合成后, 通过血淋巴运输到卵母细胞膜, 然后通过Vgr介导的内吞作用进入卵母细胞<sup>[40]</sup>, 卵母细胞中Vg的迅速积累是甲壳动物卵巢成熟的重要标志<sup>[41]</sup>。目前, Vg和Vgr基因cDNA序列全长已在多种甲壳动物中鉴别出来, 如红螯螯虾<sup>[42]</sup>、罗氏沼虾(*Macrobrachium rosenbergii*)<sup>[43-44]</sup>、短沟对虾(*Penaeus semisulcatus*)<sup>[45]</sup>和斑节对虾<sup>[46]</sup>等, 在克氏原螯虾中Vg和Vgr基因的部分序列片段已克隆出来,

但其cDNA序列全长仍需进一步鉴别。卵巢vs.肝胰腺组中的DEGs还富集到卵巢类固醇生成、前列腺素合成通路、类固醇激素合成、促性腺激素释放激素信号通路、胰岛素分泌、雌激素信号通路等KEGG通路, 这些通路均与各种激素的合成和分泌有关。在这些通路中筛选出了一些卵巢发育的候选基因, 如环氧合酶、前列腺素E<sub>2</sub>合成酶2、前列腺素E<sub>2</sub>受体EP4、3β-羟基类固醇脱氢酶7型和胰岛素样肽受体等。其中的环氧合酶(COX)和前列腺素E<sub>2</sub>合成酶2(PGES2)是前列腺素E<sub>2</sub>合成的2个关键酶, 而前列腺素E<sub>2</sub>受体EP4(EP4)是细胞膜上G蛋白偶联的前列腺素E<sub>2</sub>受体(EP)之一。PGE<sub>2</sub>与许多雌性甲壳类动物的卵巢和卵母细胞的发育有关, 例如PGE<sub>2</sub>量的增加可以促进日本囊对虾卵巢的发育<sup>[47]</sup>, 给斑节对虾饲喂含有PGE<sub>2</sub>的多毛类动物(*Perinereis* sp.)具有增强其卵母细胞发育和排卵的作用<sup>[48]</sup>, 而克氏原螯虾前列腺素E<sub>2</sub>合成通路及其受体基因对卵巢发育的调控作用的研究尚未见报道。另外, 在卵巢发育过程中, 卵巢中的卵原细胞先进行有丝分裂进行增殖, 然后行减数分裂才能最后形成成熟的卵子。在这个过程中, 调控细胞分裂周期的各种因子也与卵巢的发育和成熟密切相关。我们在孕酮介导的卵母细胞成熟和卵母细胞减数分裂KEGG通路中筛选到细胞分裂周期(cdc)蛋白16、20和23同源物、细胞周期蛋白A(cyclin A)和细胞周期蛋白E(cyclin E)等细胞周期调控因子基因, 这些候选基因在克氏原螯虾卵巢发育过程中的调控作用都有待于进一步研究。

### 3.3 部分免疫相关通路及基因分析

近年来, 各类病毒和细菌性疾病的暴发严重制约了克氏原螯虾养殖业的健康发展, 研究其免疫系统、探索其抗病机制对克氏原螯虾的健康养殖具有重要意义。肝胰腺是虾蟹类重要的免疫防御器官, 本研究的肝胰腺vs.肌肉组中的DEGs富集的免疫相关KEGG通路较多, 其中重要的有Toll和Imd信号通路、Jak-STAT信号通路、NF-κB信号通路、溶酶体、肿瘤坏死因子信号通路、趋化因子信号通路、转化生长因子β信号通路、IL-17信号通路和HIF-1信号通路等。Toll和Imd信号传导途径是甲壳动物的2条重要的免疫信号通路, Toll信号传导途径主要介导对

真菌和大部分革兰氏阳性菌的先天性免疫反应, 而对革兰氏阴性菌的免疫应答则主要由 Imd 途径介导<sup>[49]</sup>。这 2 条信号传导途径都能激活 NF- $\kappa$ B 核转录因子, 通过转录因子 NF- $\kappa$ B 作用于靶标基因  $\kappa$ B 位点, 从而调节抗菌活性物质的表达。本研究从 Toll 和 Imd 信号通路中筛选出了 Toll 样受体 2(Toll2)、Toll 样受体相互作用蛋白、*cactus*、*tube*、*pelle*、*dorsal*、*caspase* 和 TAK1 结合蛋白 1 等免疫相关基因。从 NF- $\kappa$ B 信号通路中筛选到 NF- $\kappa$ B 基因。Toll2 基因全长 cDNA 序列已在克氏原螯虾中克隆出来, 其抗革兰氏阳性菌的免疫功能和机制也已被阐明<sup>[50]</sup>。然而克氏原螯虾的 Toll 样受体相互作用蛋白及 Toll 信号通路中的 *cactus*、*tube*、*pelle*、*dorsal* 和 *caspase* 内在元件, 以及 TAK1 结合蛋白 1 和 NF- $\kappa$ B 基因的鉴定和免疫功能研究尚未见报道。Jak-STAT 信号通路由细胞因子激活, 在动物中高度保守, 参与抗细菌和抗病毒免疫<sup>[51]</sup>, 其主要由 3 个成分组成, 即酪氨酸激酶相关受体、酪氨酸激酶 JAK 和转录因子 STAT。本研究在 Jak-STAT 信号通路中筛选到 *domeless* 基因, *domeless* 是一种酪氨酸激酶受体, 其接收来自膜外细胞因子 Upds 家族蛋白的信号而被激活, 使细胞膜内的 Jak 发生磷酸化, 然后再通过激活转录因子 STAT 而启动免疫反应。目前, 在甲壳动物中仅凡纳滨对虾<sup>[52]</sup>和中华绒螯蟹<sup>[53]</sup>已完成 *domeless* 基因的克隆和抗菌免疫功能的鉴定。本研究还从其他免疫相关 KEGG 通路中筛选到其他大量的免疫相关候选基因(表 2), 其在克氏原螯虾体内的免疫功能均需进一步研究。

### 3.4 部分肌肉生长相关通路与基因分析

在虾类组织中, 腹部的肌肉组织占到自身重量的很大比例, 并且是最主要的食用部位, 也是生长调控的靶器官, 虾类增产的主要目标也是腹部肌肉的快速增加。目前, 关于虾类肌肉生长的研究处于起步阶段, 对于肌肉的生长规律、调控机制的研究很少。本研究的肌肉 vs. 卵巢组中的 DEGs 富集的肌肉生长相关 KEGG 通路比较多, 其中重要的有肌动蛋白细胞骨架的调节、钙信号通路、PI3K-Akt 信号通路、细胞周期、胰岛素信号通路、胰岛素分泌等。在这些通路中筛选到了重要肌肉生长相关基因: 肌动蛋白、肌球蛋白重链、肌肉生长抑制素、生长激素、5-羟色胺受体、胰岛素样生长因子 2、胰

岛素样肽受体、表皮生长因子受体、肌钙蛋白 T、泛素结合酶等。早期在分子水平上对甲壳动物肌肉生长的研究主要集中在不同物种肌纤维蛋白基因如肌动蛋白基因(*actin*)、肌球蛋白重链基因(*MHC*)及原肌球蛋白基因(*Tropomyosins*)等基因的克隆与表达分析。在对黑色陆地蟹(*Gecarcinus lateralis*)<sup>[54]</sup>腿部肌肉的研究中发现, *actin* mRNA 水平在蜕壳前期及后期相对于蜕壳间期的表达水平增加。另外研究还发现, 黑色陆地蟹在蜕壳前期虽然其蛋白合成大概增加了 9 倍, 而 *MHC* 的表达水平却下降, 表明其转录及翻译过程均参与调控黑色陆地蟹肌肉的生长<sup>[55]</sup>。肌肉生长抑制素(MSTN)也称为生长分化因子 8(GDF-8), 是一个哺乳动物关键的骨骼肌生长的自分泌或旁分泌抑制剂, 在无脊椎动物肌肉生长和发育中, *MSTN* 基因在进化中被保留。*MSTN* 基因在黑色陆地蟹<sup>[56]</sup>和中华绒螯蟹<sup>[57]</sup>中的克隆和功能分析证实了该基因与哺乳动物一样在甲壳动物的肌肉生长和发育中扮演了重要的角色。在甲壳动物中, 假定的生长激素(*GH*)和生长激素受体(*GHR*)基因已经被报道<sup>[58]</sup>, 但缺乏直接的证据证明其具有促进生长的作用, 本研究筛选出的克氏原螯虾 *GH* 基因的促生长作用还有待于进一步研究证实。在甲壳类动物中, 5-羟色胺(5-HT)已被证明可以刺激神经激素的释放。甲壳动物高血糖激素(CHH)、神经抑制激素和蜕皮抑制激素(MIH)通过与各种 5-HT 受体亚型相互作用调节生长<sup>[59]</sup>。研究证实在凡纳滨对虾中, 5-HT 受体具有两个同义的单核苷酸多态性(SNP)位点均与其体质量的增高有关<sup>[60]</sup>。本实验从克氏原螯虾肌肉 vs. 卵巢组中筛选到的这些差异基因中, *actin* 基因已克隆获得 cDNA 序列全长, 并已证实了其促生长的作用<sup>[61]</sup>, 而 *MHC*、*MSTN*、*GH*、5-HT 受体及表 3 中大部分基因均未获得 cDNA 序列全长, 其对克氏原螯虾的肌肉生长调控作用还需进一步研究证实。

### 3.5 组织特异基因与新基因的挖掘

本研究发现了大量的卵巢、肝胰腺和肌肉组织的特异基因(图 4-a), 但只有少量的基因得到了注释, 通过 GO 功能分类和 KEGG 信号通路富集分析发现其中的色素分散激素 A 肽、甲壳动物雌性特异荷尔蒙和卵黄膜外层蛋白 1 基因可能与卵巢发育有关(表 4), 内切葡聚糖酶 E-4 样蛋白、肝羧酸酯酶 1 样蛋白和泛素羧基末端水解

酶 8 样蛋白基因可能与免疫相关 (表 5), 联结素样蛋白和肌膜相关蛋白样亚型 X2 基因可能与肌肉生长相关 (表 6)。色素分散激素是甲壳动物眼柄基部的 X 器-窦腺 (XO-SG) 复合体分泌的神经激素, 具有抑制性腺发育的功能<sup>[62]</sup>, 但该基因和其他几种组织特异基因对克氏原螯虾卵巢发育、免疫和肌肉生长的作用还需要进一步研究证明。本研究通过 GO 功能分类和 KEGG 信号通路富集分析还发现了一些可能与卵巢发育、免疫和肌肉生长相关的新的功能基因, 如卵巢发育相关基因中的鸟嘌呤脱氨酶样蛋白、tolloid 样蛋白 1、金属磷酸酯酶 MPPED2 样蛋白等基因 (表 4), 免疫相关基因中的内切葡聚糖酶 E-4-样蛋白、肝羧酸酯酶 1 样蛋白和  $\alpha$ -维生素 E 转移蛋白样亚型 X4 基因 (表 5), 以及肌肉生长相关基因中的 spalt-major-样同源蛋白、锌指 MYM 型蛋白 1 和 anoctamin 10 样蛋白基因 (表 6) 等, 这些基因的相关功能还需要进一步验证。另外, 本研究在组织的两两比对中还发现了大量尚未得到注释的组织高表达基因, 它们和未注释的组织特异基因一起构成了一个潜在的可能与克氏原螯虾卵巢发育、免疫和肌肉生长相关的新基因库, 这些新基因的序列特征和功能尚待进一步挖掘。

## 4 结论

采用 Illumina HiSeq<sup>TM</sup> 2500 高通量测序平台对克氏原螯虾卵巢、肝胰腺和肌肉 3 个组织进行转录组测序, 共获得 53 006 个 unigenes。分别对 3 个样品的测序文库进行两两比较, 在卵巢 vs. 肝胰腺、肝胰腺 vs. 肌肉和肌肉 vs. 卵巢组中分别检测到差异表达基因 20 382、12 753 和 21 629 个。对差异表达基因进行了 GO 功能分类及 KEGG 通路富集分析, 获得大量的卵巢发育、免疫和肌肉生长候选基因。研究结果为克氏原螯虾的繁殖、生长发育和免疫机制的进一步探索提供了基础信息。

## 参考文献 (References):

- [1] 农业农村部渔业渔政管理局. 中国渔业统计年鉴 2019[M]. 北京: 中国农业出版社, 2019: 24.  
Fisheries and Fisheries Administration Bureau of the Ministry of Agriculture and Rural Affairs. China Fishery Statistical Yearbook 2019[M]. Beijing: China Agriculture Press, 2019: 24 (in Chinese).
- [2] Vaca A A, Alfaro J. Ovarian maturation and spawning in the white shrimp, *Penaeus vannamei*, by serotonin injection[J]. *Aquaculture*, 2000, 182(3-4): 373-385.
- [3] Li S D, Sheng H C, Yu X M, *et al.* A role of cathepsin L gene in innate immune response of crayfish (*Procambarus clarkii*)[J]. *Fish & Shellfish Immunology*, 2017, 71: 246-254.
- [4] Sun Y X, Tang L, Gao J, *et al.* A role of tumor susceptibility gene 101 (TSG101) in innate immune response of crayfish *Procambarus clarkii*[J]. *Developmental & Comparative Immunology*, 2017, 76: 268-273.
- [5] Qin Z D, Babu V S, Lin H Z, *et al.* The immune function of prophenoloxidase from red swamp crayfish (*Procambarus clarkii*) in response to bacterial infection[J]. *Fish & Shellfish Immunology*, 2019, 92: 83-90.
- [6] Wang K, Ren Q, Shen X L, *et al.* Molecular characterization and expression analysis of dopa decarboxylase involved in the antibacterial innate immunity of the freshwater crayfish, *Procambarus clarkii*[J]. *Fish & Shellfish Immunology*, 2019, 91: 19-28.
- [7] Dai Y J, Wang Y Q, Zhang Y H, *et al.* The role of ficolin-like protein (PcFLP1) in the antibacterial immunity of red swamp crayfish (*Procambarus clarkii*)[J]. *Molecular Immunology*, 2017, 81: 26-34.
- [8] 田志环, 焦传珍. 甲壳动物蜕皮诱导的肌肉萎缩与生长[J]. *水产科学*, 2016, 35(5): 603-606.  
Tian Z H, Jiao C Z. Muscle atrophy and growth induced by molting in crustacean: A review[J]. *Fisheries Science*, 2016, 35(5): 603-606(in Chinese).
- [9] 吴奎, 梁宏伟, 李忠, 等. 尼罗罗非鱼雌雄鱼肌肉组织差异表达基因的筛选[J]. *水产学报*, 2014, 38(3): 316-324.  
Wu K, Liang H W, Li Z, *et al.* Screen of the differentially expressed genes in muscle tissue of the male and female Nile tilapia (*Oreochromis niloticus*)[J]. *Journal of Fisheries of China*, 2014, 38(3): 316-324(in Chinese).
- [10] Gregory D J, Waldbieser G C, Bosworth B G. Cloning and characterization of myogenic regulatory genes in three Ictalurid species[J]. *Animal Genetics*, 2004, 35(6): 425-430.
- [11] Peng G G, Zhao W, Shi Z G, *et al.* Cloning HSP70 and HSP90 genes of kaluga (*Huso dauricus*) and the effects of temperature and salinity stress on their gene expression[J]. *Cell Stress and Chaperones*, 2016, 21(2): 349-359.



- 359.
- [12] 傅萃长, 朱作言. 生长激素及其基因转移对鱼类生长和渗透压的调节作用[J]. *动物学报*, 2000, 46(4): 407-415.
- Fu C Z, Zhu Z Y. Impacts on growth and osmoregulation in fishes by growth hormone (GH) and GH-transgene[J]. *Acta Zoologica Sinica*, 2000, 46(4): 407-415(in Chinese).
- [13] Maher C A, Kumar-Sinha C, Cao X H, *et al.* Transcriptome sequencing to detect gene fusions in cancer[J]. *Nature*, 2009, 458(7234): 97-101.
- [14] 王晓锋, 何卫龙, 蔡卫佳, 等. 马尾松转录组测序和分析[J]. *分子植物育种*, 2013, 11(3): 385-392.
- Wang X F, He W L, Cai W J, *et al.* Analysis on transcriptome sequenced for *Pinus massoniana*[J]. *Molecular Plant Breeding*, 2013, 11(3): 385-392(in Chinese).
- [15] Manfrin C, Tom M, De Moro G, *et al.* The eyestalk transcriptome of red swamp crayfish *Procambarus clarkii*[J]. *Gene*, 2015, 557(1): 28-34.
- [16] Du Z Q, Jin Y H, Ren D M. In-depth comparative transcriptome analysis of intestines of red swamp crayfish, *Procambarus clarkii*, infected with WSSV[J]. *Scientific Reports*, 2016, 6(1): 26780.
- [17] Zhang Y, Li Z Y, Kholodkevich S, *et al.* Cadmium-induced oxidative stress, histopathology, and transcriptome changes in the hepatopancreas of freshwater crayfish (*Procambarus clarkii*)[J]. *Science of the Total Environment*, 2019, 666: 944-955.
- [18] Kang P F, Mao B, Fan C, *et al.* Transcriptomic information from the ovaries of red swamp crayfish (*Procambarus clarkii*) provides new insights into development of ovaries and embryos[J]. *Aquaculture*, 2019, 505: 333-343.
- [19] 宋光同, 何吉祥, 吴本丽, 等. 克氏原螯虾雌性生殖系统发育及组织结构观察[J]. *江西农业学报*, 2018, 30(2): 68-75.
- Song G T, He J X, Wu B L, *et al.* Study on development of female reproductive system and histological structure of *Procambarus clarkii*[J]. *Acta Agriculturae Jiangxi*, 2018, 30(2): 68-75(in Chinese).
- [20] Grabherr M G, Haas B J, Yassour M, *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome[J]. *Nature Biotechnology*, 2011, 29(7): 644-652.
- [21] Haas B J, Papanicolaou A. TransDecoder (Find Coding Regions Within Transcripts)[EB/OL]. [2016-05-16]. <http://transdecoder.github.io>.
- [22] Altschul S F, Madden T L, Schäffer A A, *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs[J]. *Nucleic Acids Research*, 1997, 25(17): 3389-3402.
- [23] 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.
- [24] Li B, Dewey C N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome[J]. *BMC Bioinformatics*, 2011, 12(1): 323.
- [25] Mortazavi A, Williams B A, McCue K, *et al.* Mapping and quantifying mammalian transcriptomes by RNA-Seq[J]. *Nature Methods*, 2008, 5(7): 621-628.
- [26] Anders S, Huber W. Differential expression analysis for sequence count data[J]. *Genome Biology*, 2010, 11(10): R106.
- [27] Schmittgen T D, Livak K J. Analyzing real-time PCR data by the comparative  $C_T$  method[J]. *Nature Protocols*, 2008, 3(6): 1101-1108.
- [28] Ghaffari N, Sanchez-Flores A, Doan R, *et al.* Novel transcriptome assembly and improved annotation of the whiteleg shrimp (*Litopenaeus vannamei*), a dominant crustacean in global seafood mariculture[J]. *Scientific Reports*, 2014, 4(1): 7081.
- [29] Tian Z H, Jiao C Z. Molt-dependent transcriptome analysis of claw muscles in Chinese mitten crab *Eriocheir sinensis*[J]. *Genes & Genomics*, 2019, 41: 515-528.
- [30] Oliphant A, Alexander J L, Swain M T, *et al.* Transcriptomic analysis of crustacean neuropeptide signaling during the moult cycle in the green shore crab, *Carcinus maenas*[J]. *BMC Genomics*, 2018, 19(1): 711.
- [31] Garcia C F, Pedrini N, Sánchez-Paz A, *et al.* De Novo assembly and transcriptome characterization of the freshwater prawn *Palaemonetes argentinus*: Implications for a detoxification response[J]. *Marine Genomics*, 2018, 37: 74-81.
- [32] Saetan U, Sangket U, Deachamag P, *et al.* Ovarian transcriptome analysis of vitellogenic and non-vitellogenic female banana shrimp (*Fenneropenaeus merguensis*)[J]. *PLoS One*, 2016, 11(10): e0164724.
- [33] Wang Z F, Sun L X, Guan W B, *et al.* De novo transcriptome sequencing and analysis of male and female swim-



- ming crab (*Portunus trituberculatus*) reproductive systems during mating embrace (stage II)[J]. *BMC Genet-ics*, 2018, 19(1): 3.
- [34] Jiang H X, Li X L, Sun Y H, *et al.* Insights into sexual precocity of female oriental river prawn *Macrobrachium nipponense* through transcriptome analysis[J]. *PLoS One*, 2016, 11(6): e0157173.
- [35] 董丽君, 孟宪红, 孔杰, 等. 基于转录组分析筛选凡纳滨对虾低温胁迫下的差异表达基因[J]. *中国水产科学*, 2019, 26(1): 161-171.
- Dong L J, Meng X H, Kong J, *et al.* Screening of differentially expressed genes related to the cold tolerance in *Litopenaeus vannamei* based on high-throughput transcriptome sequencing[J]. *Journal of Fishery Sciences of China*, 2019, 26(1): 161-171(in Chinese).
- [36] Muhd-Farouk H, Jasmani S, Ikhwanuddin M. Effect of vertebrate steroid hormones on the ovarian maturation stages of orange mud crab, *Scylla olivacea* (Herbst, 1796)[J]. *Aquaculture*, 2016, 451: 78-86.
- [37] Tomy S, Saikrithi P, James N, *et al.* Serotonin induced changes in the expression of ovarian gene network in the Indian white shrimp, *Penaeus indicus*[J]. *Aquaculture*, 2016, 452: 239-246.
- [38] Qiao H, Xiong Y W, Zhang W Y, *et al.* Characterization, expression, and function analysis of gonad-inhibiting hormone in Oriental River prawn, *Macrobrachium nipponense* and its induced expression by temperature[J]. *Comparative Biochemistry and Physiology A: Molecular & Integrative Physiology*, 2015, 185: 1-8.
- [39] 吴维福, 陈奕奕, 李郁娇, 等. 三丁基锡对罗氏沼虾血清中免疫酶活力的影响[J]. *广东海洋大学学报*, 2014, 34(3): 17-21.
- Wu W F, Chen L L, Li Y J, *et al.* Effects of tributyltin on the activities of immunologic enzyme in blood serum of the *Macrobrachium rosenbergii*[J]. *Journal of Guangdong Ocean University*, 2014, 34(3): 17-21(in Chinese).
- [40] Warriar S, Subramoniam T. Receptor mediated yolk protein uptake in the crab *Scylla serrata*: Crustacean vitellogenin receptor recognizes related mammalian serum lipoproteins[J]. *Molecular Reproduction and Development*, 2002, 61(4): 536-548.
- [41] Buaklin A, Sittikankaew K, Khamnamtong B, *et al.* Characterization and expression analysis of the *Broad-complex (Br-c)* gene of the giant tiger shrimp *Penaeus monodon*[J]. *Comparative Biochemistry and Physiology* Part B: Biochemistry and Molecular Biology, 2013, 164(4): 280-289.
- [42] Abdu U, Davis C, Khalaila L, *et al.* The vitellogenin cDNA of *Cherax quadricarinatus* encodes a lipoprotein with calcium binding ability, and its expression is induced following the removal of the androgenic gland in a sexually plastic system[J]. *General and Comparative Endocrinology*, 2002, 127(3): 263-272.
- [43] Okuno A, Yang W J, Jayasankar V, *et al.* Deduced primary structure of vitellogenin in the giant freshwater Prawn, *Macrobrachium rosenbergii*, and yolk processing during ovarian maturation[J]. *Journal of Experimental Zoology*, 2002, 292(5): 417-429.
- [44] Roth Z, Khalaila I. Identification and characterization of the vitellogenin receptor in *Macrobrachium rosenbergii* and its expression during vitellogenesis[J]. *Molecular Reproduction and Development*, 2012, 79(7): 478-487.
- [45] Avarre J C, Michelis R, Tiet Z A, *et al.* Relationship between vitellogenin and vitellin in a marine shrimp (*Penaeus semisulcatus*) and molecular characterization of vitellogenin complementary DNAs[J]. *Biology of Reproduction*, 2003, 69(1): 355-364.
- [46] Tiu S H K, Benzie J, Chan S M, *et al.* From hepatopancreas to ovary: Molecular characterization of a shrimp vitellogenin receptor involved in the processing of vitellogenin[J]. *Biology of Reproduction*, 2008, 79(1): 66-74.
- [47] Tahara D, Yano I. Development of hemolymph prostaglandins assay systems and their concentration variations during ovarian maturation in the kuruma prawn, *Penaeus japonicus*[J]. *Aquaculture*, 2003, 220(1-4): 791-800.
- [48] Meunpol O, Duangjai E, Yoonpun R, *et al.* Detection of prostaglandin E<sub>2</sub> in polychaete *Perinereis* sp. and its effect on *Penaeus monodon* oocyte development in vitro[J]. *Fisheries Science*, 2010, 76(2): 281-286.
- [49] Dostert C, Jouanguy E, Irving P, *et al.* The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of drosophila[J]. *Nature Immunology*, 2005, 6(9): 946-953.
- [50] Lan J F, Zhao L J, Wei S, *et al.* PcToll2 positively regulates the expression of antimicrobial peptides by promoting PcATF4 translocation into the nucleus[J]. *Fish & Shellfish Immunology*, 2016, 58: 59-66.
- [51] O'Shea J J, Plenge R. JAK and STAT signaling molecules in immunoregulation and immune-mediated

- disease[J]. *Immunity*, 2012, 36(4): 542-550.
- [52] Yan M T, Li C Z, Su Z Q, *et al.* Identification of a JAK/STAT pathway receptor domeless from Pacific white shrimp *Litopenaeus vannamei*[J]. *Fish & Shellfish Immunology*, 2015, 44(1): 26-32.
- [53] Ruan Z C, Wan Z C, Yang L, *et al.* JAK/STAT signalling regulates antimicrobial activities in *Eriocheir sinensis*[J]. *Fish & Shellfish Immunology*, 2019, 84: 491-501.
- [54] Varadaraj K S, Kumari S S, Skinner D M. Actin-encoding cDNAs and gene expression during the intermolt cycle of the Bermuda land crab *Gecarcinus lateralis*[J]. *Gene*, 1996, 171(2): 177-184.
- [55] Mykles D L. Crustacean muscle plasticity: Molecular mechanisms determining mass and contractile properties[J]. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 1997, 117(3): 367-378.
- [56] Covi J A, Kim H W, Mykles D L. Expression of alternatively spliced transcripts for a myostatin-like protein in the blackback land crab, *Gecarcinus lateralis*[J]. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 2008, 150(4): 423-430.
- [57] Kim K S, Jeon J M, Kim H W. A myostatin-like gene expressed highly in the muscle tissue of Chinese mitten crab, *Eriocheir sinensis*[J]. *Fisheries and Aquatic Sciences*, 2009, 12(3): 185-193.
- [58] Toullec J Y, Van Wormhoudt A. Quantitative variations during the intermolt cycle of peptides related to human growth hormone in *Palaemon serratus* (Crustacea; Decapoda)[J]. *Comptes Rendus de l'Académie des Sciences-Series III-Sciences de la Vie*, 1987, 305(7): 265-269.
- [59] Ongvarrasopone C, Roshorm Y, Somyong S, *et al.* Molecular cloning and functional expression of the *Penaeus monodon* 5-HT receptor[J]. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 2006, 1759(7): 328-339.
- [60] Marti S M, Onteru S K, Du Z Q, *et al.* Short communication. SNP analyses of the 5HT1R and STAT genes in Pacific white shrimp, *Litopenaeus vannamei*[J]. *Spanish Journal of Agricultural Research*, 2010, 8(1): 53-55.
- [61] Kang W K, Naya Y. Sequence of the cDNA encoding an actin homolog in the crayfish *Procambarus clarkii*[J]. *Gene*, 1993, 133(2): 303-304.
- [62] 宁黔冀. 甲壳动物激素及其应用 [M]. 北京: 化学工业出版社, 2006.
- Ning Q J. Crustacean Hormone and Its Application[M]. Beijing: Chemical Industry Press, 2006 (in Chinese).

## Transcriptome analysis of *Procambarus clarkii* to screen genes related to ovary development, immunity and growth

JIANG Hongxia, LIU Huifen, MA Xiao, WU Limin, KONG Xianghui\*, LI Xuejun\*

(Engineering Lab of Henan Province for Aquatic Animal Disease Control, Engineering Technology Research Center of Henan Province for Aquatic Animal Cultivation, College of Fisheries, Henan Normal University, Xinxiang 453007, China)

**Abstract:** *Procambarus clarkii* is one of the freshwater lobsters which has important economic value in China. However, long-term high-density farming and inbreeding have brought many negative effects on the cultivation of *P. clarkii*, such as slow growth, small size, frequent diseases, and the difficulty of artificial breeding caused by the asynchrony of gonadal development, which have severely restricted the further development of *P. clarkii* aquaculture. In order to find candidate genes involved in ovary development, immunity and muscle growth of *P. clarkii* for improving the production performance of this species, the transcriptomes of ovary, hepatopancreas and muscle of *P. clarkii* were sequenced by a new generation of high throughput sequencing technology. After quality control and assembly, the sequences acquired were blasted against NR, Swiss-Prot, pfam, COG, GO and KEGG databases, and then cluster analyses were performed. In total, 53 006 final unigenes with an average length of 1 194 bp were obtained. Pairwise comparison of sequencing libraries of 3 tissue samples revealed that 20 382 differentially expressed genes (DEGs) were found in ovary vs. hepatopancreas group, 12 753 DEGs were found in hepatopancreas vs. muscle group, 21 629 DEGs were found in muscle vs. ovary group. Gene ontology analysis indicated that some DEGs were annotated into “reproduction”, “reproduction process”, “immune system process” and “growth” GO terms. KEGG pathway analysis showed that some DEGs were enriched in signaling pathways related to ovarian development, immunity and muscle growth. Based on GO functional classification and KEGG pathway analysis, a large number of candidate genes related to ovarian development, immunity and muscle growth of *P. clarkii* were screened, such as *vitellogenin*, *vitellogenin receptor*, *Toll-like receptor 2*, *Toll-interacting protein*, *myostatin*, *5-hydroxytryptamine receptor*, etc. The results of this study enriched the gene resources of *P. clarkii* and could provide basic data for the further genetic breeding and immunity research of this commercial species.

**Key words:** *Procambarus clarkii*; transcriptome sequencing; differentially expressed gene; functional annotation

**Corresponding authors:** KONG Xianghui. E-mail: xhkong@htu.edu.cn;

LI Xuejun. E-mail: xjli@htu.cn

**Funding projects:** National Natural Science Foundation of China (31972812); Key Technology Research Project of Henan Province (192102110081); Fund for the Doctoral Program of Henan Normal University (qd17142); Special Project of Surplus Funds from Personal Scientific Research Projects of Henan Normal University (20180531)