

# 马氏珠母贝转录组数据的重新组装\*

陈 琼 顾志峰 王爱民 战 欣 石耀华<sup>①</sup>

(海南大学海洋学院 热带生物资源教育部重点实验室 海南省热带水生生物技术重点实验室 海口 570228)

**摘要** 对本实验室和其他学者已经发表的马氏珠母贝外套膜和珍珠囊的 454 焦磷酸测序转录组数据以及 NCBI 中的 EST 数据进行了重新组装、同源序列比对和 GO 注释分析。重新组装获得了 30266 个 contig, 平均长度 522 bp, 其中最长的 contig 为 4144 bp; NR 数据库 BLAST 分析获取了 2310 个相似性 contig ( $E \leq 10^{-5}$ ), 其中 1902 个在模式动物(人、小鼠、海胆、线虫、斑马鱼、果蝇或者太平洋牡蛎和珠母贝)中可以查找到同源序列。基因注释(GO)结果表明, 注释到生物学过程的 3846 个 contig 可分为 23 个亚类, 其中代谢过程蛋白、细胞过程蛋白和生物学调控蛋白为 contig 数量最多的 3 个亚类; 注释到分子功能的 4601 个 contig 共分为 11 个亚类, 其中结合蛋白的 contig 数量最多, 其次是具有催化活性的蛋白和结构分子活性蛋白; 注释到细胞成分的 2992 个 contig 分为 17 个亚类, 其中细胞和细胞组成部分的 contig 数量最多, 其次分别是细胞器和大分子复合体 contig 数量。本研究结果有助于开展外套膜和珍珠囊特异性表达基因的大规模筛选。

**关键词** 马氏珠母贝; 外套膜; 珍珠囊; 转录组; 454 焦磷酸测序

**中图分类号** S917 **文献标识码** A **文章编号** 1000-7075(2014)06-0097-06

马氏珠母贝又称合浦珠母贝(*Pinctada fucata martensii* Dunker, *P. fucata*, *P. martensii*), 是我国培育海水珍珠的主要贝类(Guo *et al.*, 1999; 王爱民等, 2010; Wang *et al.*, 2011; 顾志峰等, 2012)。虽然马氏珠母贝贝壳内部的珍珠质层和外部的棱柱层均由碳酸钙组成, 但是棱柱层是无光泽的方解石, 而紧邻内部的珍珠质层则是有光泽的文石(Zhan *et al.*, 2013)。马氏珠母贝可以培育海水游离珍珠, 因而是研究生物矿化的理想模式生物。马氏珠母贝贝壳的文石和方解石中均含有不超过 5%的以蛋白质为主的生物大分子, 其在贝壳的形成中发挥着关键性作用(Belcher *et al.*, 1996; Joubert *et al.*, 2010)。因此, 这些决定贝壳形成的蛋白质基因即矿化基因的筛选是阐明贝壳生物矿化形成机理的前提与基础。

采用贝壳和珍珠的蛋白质成分分析、抑制差减杂

交筛选等技术已经鉴定了许多参与贝壳棱柱层和珍珠质层形成的矿化基因(Sudo *et al.*, 1997; Fang *et al.*, 2011)。尽管如此, 贝壳的形成和生物矿化是一个多基因控制的过程, 筛选鉴定更多的矿化基因仍然是很有必要的基础性工作。为此, 应用第二代测序技术对参与马氏珠母贝贝壳和珍珠形成的外套膜及珍珠囊开展了较多的大规模转录组测序(Kinoshita *et al.*, 2011; Zhao *et al.*, 2012; Shi *et al.*, 2013; Zhan *et al.*, 2013)。基于经费等方面的制约, 虽然测序数据量已经很多, 但并不是很完整, 每个报道中均有较多的 singleton, 组装后的 contig 序列长度也较短, 而且没有采用他人已经发表的数据。因此, 本研究对已经发表的 454 测序技术获取的马氏珠母贝的珍珠囊和外套膜的转录组数据以及 NCBI 中马氏珠母贝 EST 数据进行重新组装和分析, 为进一步开展芯片杂交筛选矿

\* 国家自然科学基金(41366003)和国家重点基础研究发展计划(973 计划)(2010CB126405)共同资助。陈 琼, E-mail: joise234@sina.com

① 通讯作者: 石耀华, 教授, E-mail: stone70@126.com

收稿日期: 2013-12-25, 收修改稿日期: 2014-05-02

化基因等奠定基础。

## 1 材料与方法

### 1.1 数据来源

Shi 等(2013) 454 焦磷酸测序获取的马氏珠母贝外套膜组织 220824 个 reads, Zhan 等(2013)焦磷酸测序获取的马氏珠母贝珍珠囊 86121 个 reads, Kinoshita 等(2011)焦磷酸测序获取的马氏珠母贝外套膜外缘组织 32388 个 reads、外套膜制备细胞小片相应部位组织 174872 个 reads 和珍珠囊组织 183265 个 reads (<http://trace.ddbj.nig.ac.jp/DRAsearch/submission?acc=DRA000393>), 以及 2012 年 10 月 31 日从 NCBI 下载的 19073 个马氏珠母贝 EST 序列。

### 1.2 数据组装与分析

采用软件 CLC Genomics Workbench 5(默认参数)对以上转录组数据进行序列拼接形成 contig, 并对照 NR 库(Non-redundant GenBank CDS, Database: All translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects, Date: Apr 8, 2012, 12:23 AM)进行同源性查找比对, BLAST 参数为: blastx-db/nr-out result.txt-evalue 1e-5-max\_target\_seqs1-outfmt "6 qseqid qlen sseqid slen evalue score length"。

查找到同源性的 contig 进行基因注释(Gene ontology annotation, GO), 将其分别归类于生物学过程(Biological process, BP)、分子功能(Molecular function, MF)和细胞成分(Cellular component, CC) 3 大类。由于同一个 contig 可能同时归类于几个亚类别, 因此每一大类的百分数之和均可能超过 100% (Vizcaino *et al.*, 2006)。

## 2 结果

### 2.1 序列拼接组装

用于组装的 816543 个 reads 或者 EST 序列中, 有 129131 个序列不能与其他任何序列组装(称为 singleton), 其余的 687412 个可以比对拼接上, 共组装成 30266 个 contig, 其中的 29666 个 contig 序列超过 200 bp。contig 最长的为 4144 bp, 平均长度为 522 bp (表 1)。

Contig 中各种核苷酸的数量不相同, 其中腺嘌呤(Adenine, A)和胸腺嘧啶(Thymine, T)的含量较高, 分别占 32.80%和 31.80%; 鸟嘌呤(Guanine, G)显著低于腺嘌呤和胸腺嘧啶, 仅占 contig 中核苷酸的 18.40%; 胞嘧啶(Cytosine, C)的含量最低, 仅占 16.50%; 此外

表 1 拼接组装序列信息  
Tab.1 Assembly information

	组装特征 Characters	数量 Number
用于拼接的序列 Assembled reads or EST	拼接组装上的序列 Matched reads or EST	687412
	未拼接组装上的序列 Not matched reads or EST	129131
	小计 Total	816543
拼接后的 contigs Assembled contigs	contig 总数 contig number	30266
	最长序列 Maximum length of contigs (bp)	4144
	序列平均长度 Average length of contigs (bp)	522

还有约 0.50%难以区分的核苷酸(N)。

### 2.2 序列相似性

NR 库 BLAST 分析结果显示, 有 7597 个 contig 可以找到相似性序列, 其中  $E$  值小于或者等于  $10^{-5}$  的 contig 有 2310 个。2310 个具有较高序列相似性的 contig 中有 1902 个在模式动物的人(*Homo sapiens*)、小鼠(*Mus musculus*)、紫海胆(*Strongylocentrotus purpuratus*)、线虫(*Caenorhabditis elegans*)、斑马鱼(*Danio rerio*)、果蝇(*Drosophila melanogaster*)和太平洋牡蛎(*Crassostrea gigas*)或者珠母贝属(*Pinctada*)中可以查找到同源序列。 $E$  值小于或者等于  $10^{-5}$  的水平上, contig 在人、斑马鱼和小鼠中的同源序列最多, 约有 1500 个 contig; 其次是果蝇, 有 1082 个 contig; 线虫和海胆的同源序列比果蝇少, 约 700 个 contig; 珠母贝属和太平洋牡蛎同源序列更少, 分别仅有 278 和 184 个 contig。在  $E$  值小于或者等于  $10^{-10}$  的水平上, contig 在上述模式动物和珠母贝属中的同源序列数量变动趋势与  $E$  值小于或者等于  $10^{-5}$  基本相似, 斑马鱼、人和小鼠中的同源序列仍然最多, 约为 1200 个 contig; 其次是果蝇, 893 个 contig; 线虫和海胆紧随果蝇之后, 约 550 个 contig; 珠母贝属和太平洋牡蛎则分别只有 220 和 158 个 contig 可以查找到同源序列(图 1)。

### 2.3 GO 注释

3846 个具有同源性的 contig 可以 GO 注释至生物学过程上, 并细分为多达 23 个亚类。其中, 归于代谢过程和细胞过程的 contig 最多, 分别为 75.35%和 69.99%; 归于生物学调控、生物学过程调控、对刺激的反应、定位和建立定位的 contig 也较多, 在 20.20%-11.28%之间变动; 信号蛋白、细胞的成分组

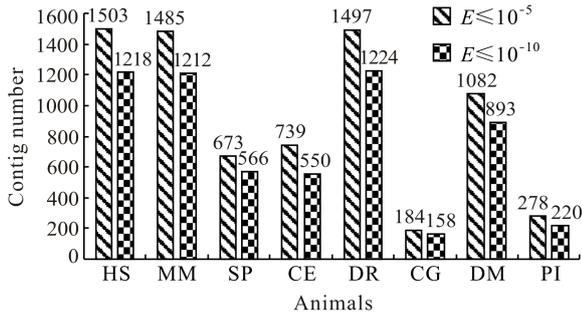


图 1 与模式动物以及珠母贝具有相似性的 contig 数量  
Fig.1 Number of contigs with sequence similarity to several model animals or *Pinctada*

HS: 人; MM: 小鼠; SP: 紫海胆; CE: 秀丽隐杆线虫;  
DR: 斑马鱼; CG: 太平洋牡蛎; DM: 果蝇; PI: 珠母贝  
HS: *H. sapiens*; MM: *M. musculus*; SP: *S. purpuratus*;  
CE: *C. elegans*; DR: *D. rerio*; CG: *C. gigas*;  
DM: *D. melanogaster*; PI: *Pinctada*

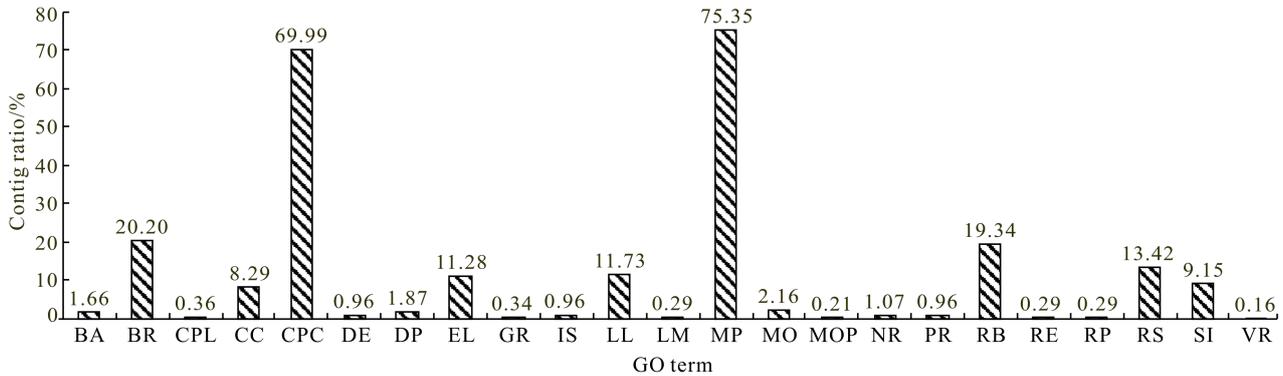


图 2 contig 的生物学过程 GO 注释  
Fig.2 Gene ontology annotation of biological process for contig

BA: 生物学附着; BR: 生物学调控; CPL: 细胞扩增; CC: 细胞成分的组织或者生物学发生; CPC: 细胞过程; DE: 死亡;  
DP: 发育过程; EL: 建立定位; GR: 生长; IS: 免疫系统过程; LL: 定位; LM: 运动; MP: 代谢过程; MO: 多细胞有机体过程;  
MOP: 多器官进程; NR: 生物学过程负调控; PR: 生物学过程正调控; RB: 生物学过程调控; RE: 繁殖; RP: 繁殖过程;  
RS: 对刺激的反应; SI: 信号; VR: 病毒繁殖

BA: biological adhesion; BR: biological regulation; CPL: cell proliferation; CC: cellular component organization or biogenesis;  
CPC: cellular process; DE: death; DP: developmental process; EL: establishment of localization; GR: growth; IS: immune system  
process; LL: localization; LM: locomotion; MP: metabolic process; MO: multicellular organismal process; MOP: multi-organism  
process; NR: negative regulation of biological process; PR: positive regulation of biological process; RB: regulation of biological  
process; RE: reproduction; RP: reproductive process; RS: response to stimulus; SI: signaling; VR: viral reproduction

可以 GO 注释到细胞成分的 contig 较少,为 2992 个,可分为 17 个亚类。归于细胞和细胞组成部分的 contig 最多而且数量相同,均是 68.22%; 其次是归于细胞器的 contig, 约占 47.46%; 大分子复合体、膜和膜组成部分的 contig 也较多,分别为 30.48%、30.21%和 22.09%; 此外,细胞器组成部分和胞外区的 contig 数量也不少,分别为 19.65%和 15.37%; 其余的胞外区组成部分、膜内包腔、胞外基质、细胞连接、细胞外基质组成部分、突触、突触组成部分、病毒粒子、

织或者生物学发生、多细胞有机体过程、发育过程、生物学附着和生物学过程负调控的 contig 虽然超过 1%,但均不足 10%; 其余与死亡、免疫系统过程、生物学过程正调控等 10 个亚类相关的 contig 均不足 1% (图 2)。

4601 个具有同源性的 contig 可以 GO 注释至分子功能上,共分为 11 个亚类。contig 数量最多的是结合蛋白和具有催化活性的蛋白,分别有 2902 和 2232 个 contig,约占 63.07%和 48.51%; 其余依次为结构分子活性蛋白、转运子活性蛋白、酶调控活性蛋白、电子载体活性蛋白、受体活性蛋白、分子转导活性蛋白和结合核酸的转录因子活性,分别占 contig 的 9.89%–1.80%不等; 比例最小的则是抗氧化活性蛋白和结合蛋白质的转录因子活性,分别为 0.93%和 0.70% (图 3)。

病毒粒子组成部分等 9 个亚类均较少,其 contig 数量都不足 4% (图 4)。

### 3 讨论

第二代测序技术主要有 454 焦磷酸测序、Illumina solexa 测序、Solid 等。虽然 Illumina 和 Solid 每个反应可以产生更多的数据,但 Illumina 和 Solid 测序获取的每个 read 序列长度非常短,容易导致进行序列

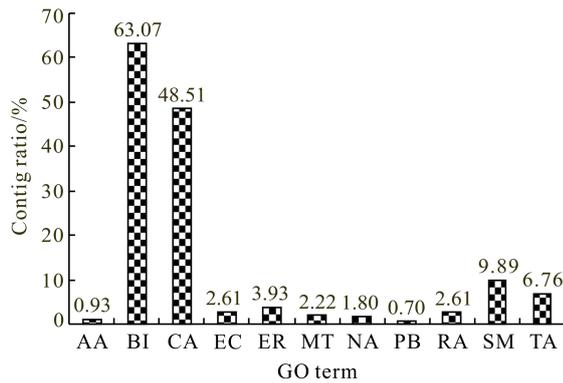


图3 contig 分子功能的 GO 注释

Fig.3 Gene ontology annotation of molecular function for contig

AA: 抗氧化活性; BI: 结合; CA: 催化活性; EC: 电子载体活性; ER: 酶调控活性; MT: 分子转导活性; NA: 结合核酸的转录因子活性; PB: 结合蛋白质的转录因子活性; RA: 受体活性; SM: 结构分子活性; TA: 转运子活性

AA: antioxidant activity; BI: binding; CA: catalytic activity; EC: electron carrier activity; ER: enzyme regulator activity; MT: molecular transducer activity; NA: nucleic acid binding transcription factor activity; PB: protein binding transcription factor activity; RA: receptor activity; SM: structural molecule activity; TA: transporter activity

拼接时同一 read 拼接位置的不确定性, 对于没有全基因组信息的物种而言这种拼接组装困难更加突出 (Zhao *et al*, 2012; Shi *et al*, 2013)。454 焦磷酸测序每个 read 序列平均长度多在 350 bp 以上, 明显长于 Illumina 和 Solid 测序, 拼接结果的可靠性显著高于 Illumina 和 Solid 测序, 并且每个反应也可以获取大

量数据, 因而是许多转录组分析的首选第二代测序技术 (Kinoshita *et al*, 2011; Yúfera *et al*, 2012)。马氏珠母贝的外套膜和珍珠囊组织转录组虽然有 Illumina 测序数据 (Huang *et al*, 2013; Zhao *et al*, 2012), 但是采用 454 焦磷酸测序已经获得了大量数据, 因此, 本研究在进行数据分析时没有采用 Illumina 测序的数据, 而是将更有效可靠的 454 焦磷酸测序数据与 NCBI 中经典测序获取的 EST 数据整合进行分析, 确保数据的一致性和可靠性。

组装得到 30266 个 contig, 其中 29666 个 contig 的序列长度大于 200 bp。Shi 等 (2013) 外套膜转录组数据的组装只是获得了 3574 个长度大于 100 bp 的 contig, Zhan 等 (2013) 关于珍珠囊的转录组数据的组装也只得到 1554 个长度大于 100 bp 的 contig, 他们组装的 unique 基因中 90% 以上都是单一 read 构成的 singleton。Kinoshita 等 (2011) 虽然在结果中列出了共 29682 个 contig, 但是其每个 contig 的 read 数量为 1-2460, 很显然也将很多单个 read 列为 contig, 而且, 该研究 contig 最小的只有 118 bp, 组装的 contig 的平均长度为 495 bp, 显著小于本次重组 contig 的平均长度 522 bp。本次重组的 contig 最长为 4144 bp, 也明显长于其他报道。因而, 本次重组更好地利用了测序获取的 read 数据, 为进一步用于大规模的基因芯片杂交筛选组织特异性转录子和功能基因鉴定奠定了初步基础。

组装的 contig 中腺嘌呤和胸腺嘧啶的含量较高, 而鸟嘌呤和胞嘧啶含量较低, 这与组装的是转录组数据有关。真核生物转录的 mRNA 在加工成熟过程中

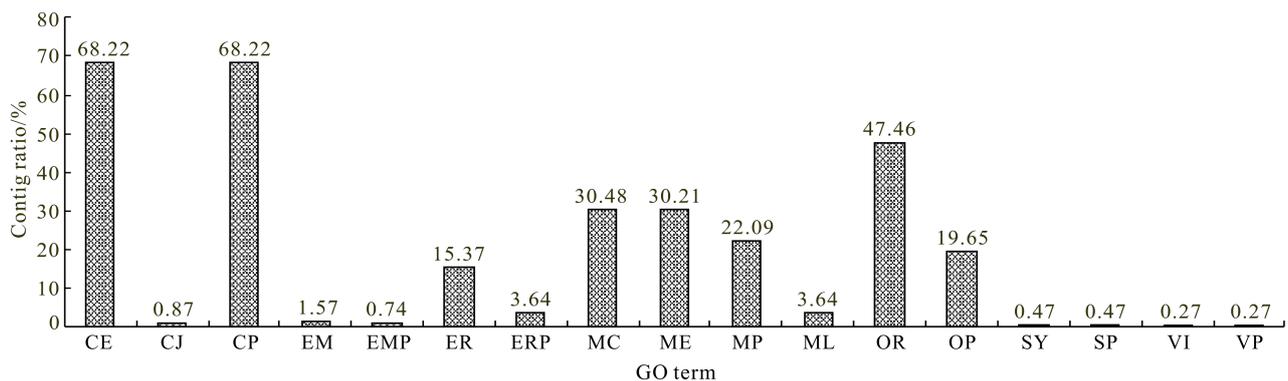


图4 contig 的细胞成分 GO 注释

Fig.4 Gene ontology annotation of cellular component for contig

CE: 细胞; CJ: 细胞连接; CP: 细胞组分; EM: 胞外基质; EMP: 细胞外基质的组成部分; ER: 胞外区; ERP: 胞外区组成部分; MC: 大分子复合体; ME: 膜; MP: 膜组成部分; ML: 膜内包腔; OR: 细胞器; OP: 细胞器组成部分; SY: 突触; SP: 突触组成部分; VI: 病毒粒子; VP: 病毒粒子组成部分

CE: cell; CJ: cell junction; CP: cell part; EM: extracellular matrix; EMP: extracellular matrix part; ER: extracellular region; ERP: extracellular region part; MC: macromolecular complex; ME: membrane; MP: membrane part; ML: membrane-enclosed lumen; OR: organelle; OP: organelle part; SY: synapse; SP: synapse part; VI: virion; VP: virion part

有 3'端加尾修饰,显著增加了腺嘌呤的数量,在测序过程中虽然只有部分转录子测序到了 3'末端,仍然会显著增加腺嘌呤的数量。因此,腺嘌呤和胸腺嘧啶的含量较高是合理的。

contig 与模式动物人、斑马鱼和小鼠的同源序列最多,其他的果蝇、线虫和海胆中同源的 contig 虽然明显比这 3 种动物低,但是大大高于珠母贝和太平洋牡蛎。虽然珠母贝与马氏珠母贝是同一属生物,而与太平洋牡蛎均属于软体动物的双壳纲,理论上应该与马氏珠母贝的同源序列更多且同源性也更高,而且太平洋牡蛎最近完成了全基因组序列的测定(Zhang *et al.*, 2012),但是,这些生物不是模式生物,与模式生物的基因组和转录组研究水平相去甚远,很多序列没有进行注释,而且注释也往往是参照模式生物,很少是直接进行功能鉴定分析后注释的。因此,很多非模式生物进行同源性查找时,与人、小鼠和斑马鱼等模式生物的同源序列最多,这里马氏珠母贝的外套膜和珍珠囊转录组分析结果也不例外。

Kinoshita 等(2011)分析的生物学过程 GO 注释中所占百分比最高的 3 个亚类与 Shi 等(2013)关于外套膜转录组的注释结果相同,依次均为细胞过程蛋白、代谢过程蛋白和生物学调控蛋白;重组装后的注释中前二者的顺序发生了互换,即代谢过程蛋白高于细胞过程蛋白的 contig 数量。Zhan 等(2013)关于珍珠囊的生物学过程 GO 注释则是定位蛋白取代细胞过程蛋白进入 contig 数量第 3。

重组装后的分子功能 GO 注释中所占百分比最高的 3 个亚类与 Shi 等(2013)关于外套膜转录组的注释和 Zhan 等(2013)关于珍珠囊的注释结果相同,分别是结合蛋白、催化活性蛋白和结构分子活性蛋白。Kinoshita 等(2011)的分子功能 GO 注释结果也显示,这 3 类蛋白 contig 数量最多,而且其中的结合蛋白 contig 数量也是最高的,但其结构分子活性蛋白比催化活性蛋白的 contig 数量要少。

Shi 等(2013)和 Zhan 等(2013)关于外套膜和珍珠囊的转录组的细胞成分 GO 注释与重组装后的注释结果均显示细胞或细胞组成部分、细胞器蛋白以及大分子复合体蛋白依次是 contig 数量最多的 3 大亚类,与 Kinoshita 等(2011)的结果既有相似之处,即细胞或细胞组成部分是 contig 数量最多的,但彼此间又存在一定差异,即 Kinoshita 等(2011)注释时没有大分子复合体蛋白亚类,而是细胞器蛋白细分为非膜结合细胞器亚类和膜结合细胞器亚类,而且这两个亚类的 contig 数量分别居于第 2、3 位。

## 参 考 文 献

- 王爱民, 石耀华, 王嫣, 等. 马氏珠母贝生物学与养殖新技术. 北京: 中国农业科技出版社, 2010
- 顾志峰, 黄锋绍, 王海, 等. 不同海区养殖的马氏珠母贝珍珠质颜色比较. 渔业科学进展, 2012, 33(5): 91-94
- Belcher AM, Wu XH, Christensen RJ, *et al.* Control of crystal phase switching and orientation by soluble mollusc-shell proteins. *Nature*, 1996, 381(6577): 56-58
- Fang D, Xu G, Hu Y, *et al.* Identification of genes directly involved in shell formation and their functions in pearl oyster, *Pinctada fucata*. *PLoS ONE*, 2011, 6(7): e21860
- Guo X, Ford SE, Zhang F. Molluscan aquaculture in China. *J Shellfish Res*, 1999, 18(1): 19-31
- Huang XD, Zhao M, Liu WG, *et al.* Gigabase-scale transcriptome analysis on four species of pearl oysters. *Mar Biotechnol*, 2013, 15(3): 253-264
- Joubert C, Piquemal D, Marie B, *et al.* Transcriptome and proteome analysis of *Pinctada margaritifera* calcifying mantle and shell: focus on biomineralization. *BMC Genomics*, 2010, 11(1): 613
- Kinoshita S, Wang N, Inoue H, *et al.* Deep sequencing of ESTs from nacreous and prismatic layer producing tissues and a screen for novel shell formation-related genes in the pearl oyster. *PLoS One*, 2011, 6(6): e21238
- Shi YH, Yu CC, Gu ZF, *et al.* Characterization of the pearl oyster (*Pinctada martensii*) mantle transcriptome unravels biomineralization genes. *Mar Biotechnol*, 2013, 15(2): 175-187
- Sudo S, Fujikawa T, Nagakura T, *et al.* Structures of mollusc shell framework proteins. *Nature*, 1997, 387(6633): 563-564
- Vizcaino JA, Gonzalez FJ, Suarez MB, *et al.* Generation, annotation and analysis of ESTs from *Trichoderma harzianum* CECT 2413. *BMC Genomics*, 2006, 7: 193
- Wang AM, Wang Y, Gu ZF, *et al.* Development of expressed sequence tags from the pearl oyster, *Pinctada martensii* Dunker. *Mar Biotechnol*, 2011, 13(2): 275-283
- Yúfera M, Halm S, Beltran S, *et al.* Transcriptomic characterization of the larval stage in gilthead seabream (*Sparus aurata*) by 454 pyrosequencing. *Mar Biotechnol*, 2012, 14(4): 423-435
- Zhan X, Gu ZF, Yu CC, *et al.* Expressed sequence tags 454 sequencing and biomineralization gene expression for pearl sac of the pearl oyster, *Pinctada fucata martensii*. *Aquaculture Research*, 2013, DOI: 10.1111/are.12227
- Zhang G, Fang X, Guo X, *et al.* The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, 2012, 490(7418): 49-54
- Zhao X, Wang Q, Jiao Y, *et al.* Identification of genes potentially related to biomineralization and immunity by transcriptome analysis of pearl sac in pearl oyster *Pinctada martensii*. *Mar Biotechnol*, 2012, 14(6): 730-739

(编辑 冯小花)

## Data Reassembling Reveals More Information of Transcriptome for *Pinctada fucata martensii* Dunker

CHEN Qiong, GU Zhifeng, WANG Aimin, ZHAN Xin, SHI Yaohua<sup>①</sup>

(Key Laboratory of Tropic Biological Resources, Ministry of Education, Hainan Key Laboratory of Tropical Hydrobiological Technology, Ocean College, Hainan University, Haikou 570228)

**Abstract** The pearl oyster, *Pinctada fucata martensii*, was an ideal animal for studies of biomineralization. The mantle and pearl sac were the two important tissues involved in biomineralization of the shell and pearls. In this study we reassembled the transcriptome data of 454 pyrosequencing of the mantle and pearl sac. The sources of the data included studies in our own laboratory, reports from other researchers, and ESTs from NCBI. We conducted BLAST and Gene Ontology analysis, and analyzed the contig reassembling with CLC Genomics Workbench 5. The reassembling generated 30266 contigs with an average length of 522 bp, and the largest length of 4144 bp. The Blast analysis against non-redundant (NR) database revealed 2310 contigs with the *E*-values smaller than  $10^{-5}$ , among which 1902 contigs had homologous sequences in at least one of the following model animals of *Homo sapiens*, *Mus musculus*, *Strongylocentrotus purpuratus*, *Caenorhabditis elegans*, *Danio rerio*, *Drosophila melanogaster*, or *Crassostrea gigas* and *Pinctada*. The GO annotation showed that 3846 contigs could be characterized into 23 subcategories of biological processes, of which the top three subcategories with most contigs were metabolic process, cellular process and biological regulation; 4601 contigs could be characterized into 11 subcategories of molecular functions, of which the top three with most contigs were binding, catalytic activity, and structural molecular activity; 2992 contigs could be characterized into 17 subcategories of cellular components, of which the top three were cell part, organelle, and macromolecular complex. These results suggested that there were a great number of transcriptome contigs of *P. fucata martensii*. Our study will be helpful in the large-scale screening of specially-expressed genes in the mantle and pearl sac of *P. fucata martensii*.

**Key words** *Pinctada fucata martensii*; Mantle; Pearl sac; Transcriptom; 454 Pyrosequencing

<sup>①</sup> Corresponding author: SHI Yaohua, E-mail: stone70@126.com