

笠贝幼虫变态相关基因筛选及分子网络构建

尹 诚^{1,2}, 张丽莉^{1,2}, 王国栋^{1,2}, 王艺磊^{1,2}

(1. 集美大学水产学院, 福建 厦门 361021; 2. 农业部东海海水健康重点实验室, 福建 厦门 361021)

[摘要] 利用NCBI在线数字化差异显示工具(Digital Differential Display)对笠贝(*Lottia gigantea*) EST数据库进行筛选, 获得笠贝幼虫时期显著差异表达基因203个, 其中124个基因有Blast2Go注释结果, 这124个基因可以归类成细胞组分、分子功能和生物学过程三大类; 参与MAPK、Ras、PI3K-Akt、Rap1、cAMP等信号通路; 涉及能量代谢、免疫应激、细胞分化、神经发育、细胞骨架、信号传递等生物学过程。将获得的转录子运用Cytoscape软件构建分子相互作用网络, 最终获得7个核心转录子。涉及到神经发育、次生壳形成等相关基因调控关系。

[关键词] 笠贝; 数字化差异显示; 变态相关基因; 生物分子网络

[中图分类号] Q 959.212⁺.2; Q 789

[文献标志码] A

Identification of Metamorphosis Related Genes and Development of Molecular Networks of *Lottia gigantea* Larval

YIN Cheng^{1,2}, ZHANG Li-li^{1,2}, WANG Guo-dong^{1,2}, WANG Yi-lei^{1,2}

(1. Fisheries College, Jimei University, Xiamen 361021, China; 2. Key Laboratory of Healthy Mariculture for the East China Sea, Ministry of Agriculture, Xiamen 361021, China)

Abstract: By using the NCBI online Digital Differential Display tool, 203 differentially expressed genes in metamorphosis of *Lottia gigantea* larvae were identified from ESTs database of *L. gigantea*. About 124 genes had Blast2Go annotations which were classified into cellular components, molecular functions and biological processes by GO (Gene Ontology). These genes are members of some signal pathways, such as MAPK, Ras, PI3K-Akt, Rap1, cAMP, et al, and refer to many biological processes, such as energy metabolism, immunological stress, cell differentiation, neurodevelopment, cytoskeleton, signal transmission. These 124 genes were used to build molecular interaction network by Cytoscape software. A core network, including seven core genes, was showed. The network involved in nerve development, tissue differentiation, secondary shell formation and energy metabolism, and revealed a possible molecular mechanism of metamorphosis.

Key words: *Lottia gigantea*; Digital Differential Display; metamorphosis related genes; molecular interaction network

0 引言

贝类的生活史中有浮游幼虫阶段,其幼虫和成体在形态结构和生活习性上差别巨大.幼虫通过变态过程转变为与成体形态结构和生活习性相同的稚贝.变态是贝类发育过程的关键环节,直接影响其种群分布、数量变动和物种进化^[1].笠贝(*Lottia gigantea*)隶属软体动物门(Mollusca)、腹足纲(Gastropoda)、前鳃亚纲(Prosobranchia)、笠形腹足目(Patellogastropoda)、莲花青螺科(Lottiidae)(国内分类为笠贝科(Acmaeidae)),又名加利福尼亚大戚(California Limpet, Owl Limpet),主要分布在美国加州北部到加利福尼亚半岛南部的太平洋沿岸.其壳长可达9 cm,寿命最长达20年,肉质鲜美,营养丰富,是一种重要的海洋经济软体动物.同其他贝类相比,其基因组比较小,成为研究进化和发育的重要模式物种^[2].

在我国与笠贝同科的物种有,鹌足青螺(鸡爪拟帽贝,*Patelloida saccharina*),主要分布在浙江以南各沿海海域;花青螺(史氏笠贝,*Nipponacmea schrenckii*),主要分布于我国南、北沿海以及日本、朝鲜等地;射线青螺(细线拟帽贝,*Acmaea striata*),主要分布于台湾海峡两岸、印度尼西亚等地;嫁戚(*Cellana toreuma*),主要分布于山东东海沿岸到浙江北部沿岸潮间带.其中,嫁戚分布纬度与笠贝相似,且其食物链短、营养价值高、活动范围小,自然增殖年限为3龄,是理想的岩相区域性增殖贝类^[3].研究笠贝的基因,构建其变态相关分子网络,对嫁戚的保护和养殖具有重要意义.

在NCBI的核酸数据库,该物种的cDNA序列众多,并且于2012年完成了基因组测序.如何利用这些海量的数据信息,成为生物科学工作者面临的一个重要问题.NCBI在线数字化差异显示工具DDD(Digital Differential Display, <http://www.ncbi.nlm.nih.gov/UniGene/ddd.cgi>)能够分析比较不同cDNA文库中转录子的表达差异,为充分利用大量的cDNA文库测序信息提供了一个解决途径.Blast2Go是一个综合型的序列注释分析软件,可以按照需要对DNA序列进行比对注释和分析,获得不同数据库中该序列的信息.Cytoscape是一款可图形化显示数据并具有分析和编辑功能的软件,是构建生物网络的常用软件.

本文采用DDD筛选出笠贝变态前后的差异表达的转录子,根据转录子的功能和表达量,通过Cytoscape软件构建了笠贝变态相关基因的调控网络,为合理利用丰富的网络核酸数据提供一个实例.

1 方法

1.1 文库构建

DDD基于UniGene数据库.UniGene将物种EST序列按最适标准的保守方法分配成群集(Clusters),每个群集代表唯一基因.DDD利用UniGene这个特点,比较不同cDNA文库间时间序列差异的表达谱序列,选出特定的UniGene Cluster^[4].在特定状态下,cDNA文库的UniGene中存在许多差异,其中部分差异显示序列具有重要研究价值.DDD利用Fisher精确检验(The Fisher Exact Test)方法,对比发现有重要研究价值的差异显示序列.这些序列条数越多,Fisher检验值越高,检索对比得到的序列价值就越高^[5].将有价值的UniGene的群集下载到本地系统,可以进行后续EST的拼接和组装工作^[6].

本研究利用DDD对笠贝(*L. gigantea*)EST数据库进行筛选,将不同的cDNA文库按照不同发育时间分成3个文库(pool),获得在统计学上表达量有显著差异的基因的EST序列.文库构建标准及每个文库包含的ESTs数量见表1.这些差异基因的表达量数据是本文进一步数据分析的基础.

1.2 序列拼接与结果筛选

将DDD筛选的ESTs以转录子为单位打包下载,使用UniGene自身的名称“Lgi. XXX”对ESTs命名.使用信息学软件DNASTAR中的Seqman功能对下载获得的所有EST序列分别进行拼接.分别将拼接获得的Contigs序列另存为fasta格式文件,并以对应的“Lgi. XXX”命名.

表 1 DDD 文库构建标准及 ESTs 数量
Tab.1 Standards and the number of ESTs for DDD library construction

文库名 Library Name	发育时间 Developmental Time	发育时期 Developmental Stage	EST 数目 EST Number	子文库名 Sub Library Name
poolA	3 ~ 18 h	面盘幼体 Veliger	49388	CBBG <i>L. gigantea</i> 12,15,18 h embryos CBGB <i>L. gigantea</i> 3,6,9,12 h embryos CBGC <i>L. gigantea</i> 15,18 h embryos
poolB	26 ~ 61 h	变态前 Before Metamorphosis	19896	CBBH <i>L. gigantea</i> 26,37,61 h Larvae (M) CBBI <i>L. gigantea</i> 26,37,61 h Larvae (L)
poolC	72 ~ 120 h	变态后 After Metamorphosis	16620	CBBN <i>L. gigantea</i> 3,4,5,6.5 d Larvae (M)

1.3 序列分析与本体论分类

Blast2Go 是一个综合型的序列注释分析软件. 优点是易使用、高通量, 且各部分相互协作, 可以自主进行调配, 按照需要对数据进行整理统计并做成想要的形式, 支持 GO、KEGG maps、Interpro 和 Enzyme Codes 等数据库^[7].

从 Blast2Go 官网(<http://www.Blast2Go.com/b2ghome>)下载并安装 Blast2Go 2.8 PC 端 (500 MB), 点击 File 中的 Load FASTA File (.fasta), 将所有拼接结果的 fasta 文件导入软件中. 运行 Blast 中的 Run BLAST, 并将 Blast Program 设置为 Blastx, Blast DB 选择为 nr, Blast Expectvalue 设置为1.0E-3. 运行 Blast 完毕后, 再选择运行 Mapping 中的 Run GO-Mapping, 接着运行 Annotation 中的 Run Annotation, 参数设置如下: E-Value-it-filter 为 1.0E-6, Annotation-Cutoff 为 55, Go-Weiht 为 5. 通过点击 Statistics 中的各项获得相关的统计学数据. 通过运行 Analysis 中的 Make Combined Graph 绘制本体论分类图, 并导出其 Excel 格式文件.

1.4 表达趋势归类

将 Blast2Go 的输出结果与表达量、序列信息一一对应整合入 Excel 表格中. 统计并制作各基因在不同时期的表达量变化趋势图. 按变化趋势对基因进行分类. 本文将稚贝 3 个变化阶段 (受精到面盘幼体、面盘幼体到变态前、变态前到变态后) 不同表达趋势的基因分为以下几类:

- I) 表达量升高, 即受精至稚贝表达量一直上升;
- II) 表达量降低, 即受精至稚贝表达量一直下降;
- III) 表达量先升高后降低, 即变态阶段的表达量最高, 高于变态前和变态后;
- IV) 表达量先降低后升高, 即变态阶段的表达量最低, 低于变态前和变态后.

对各分类间基因变化关系进行分析, 并依据 William^[8]的分类标准, 将筛选获得的 124 条变态相关基因分为 9 大类: 钙离子结合、分子相互作用、代谢、免疫应激、膜蛋白、基质蛋白、细胞骨架、核糖体组分和信号传递. 结合表达趋势, 获得基因功能的表达趋势分类结果.

1.5 表达量数据标准化与调控相关系数计算

对表达趋势归类中单独列出的基因的表达量数据进行对数转换后做标准化处理^[9]. 公式为:

$$Z_t = (X_t - \bar{X}_t) / \sqrt{\sum_{i=1}^N (X_i - \bar{X})^2 / 3}, (N = 1, 2, 3),$$

其中: X_t 为该基因 t 时序时对数处理后的表达量向量, t 为表达时序, N 为 t 的具体取值范围, Z_t 为标准化的表达量向量. 利用 Pearson 相关系数评估不同基因之间表达量相似性, 以进一步确定它们之间是否有调控关系, 筛选出起调控作用的基因^[10].

使用矩阵中任意基因 E 和 F 的标准化表达量向量计算 E 、 F 的相关系数 r_{ef} . 公式为:

$$r_{ef} = \frac{\sum_{t=1}^N (Z_{et} - \bar{Z}_e)(Z_{ft} - \bar{Z}_f)}{\sqrt{\sum_{t=1}^N (Z_{et} - \bar{Z}_e)^2 \sum_{t=1}^N (Z_{ft} - \bar{Z}_f)^2}},$$

$(N = 1, 2, 3; Z_e = Z_{e1}, Z_{e2}, \cdots, Z_{et}; Z_f = Z_{f1}, Z_{f2}, \cdots, Z_{ft}).$

当 r 值为 1 时，说明两条基因是正调控；当 r 值为 -1 时，说明两条基因是负调控；当 r 值为 0 时，说明两条基因没有调控关系。 r 值越接近 1 或 -1 ，两条基因相关性越显著^[9]。使用 SPASS 软件计算 r 值，并检验其显著性 P ，取 $P < 0.05$ 的 r 值（显著差异）作为网络构建的依据^[9,11-12]。

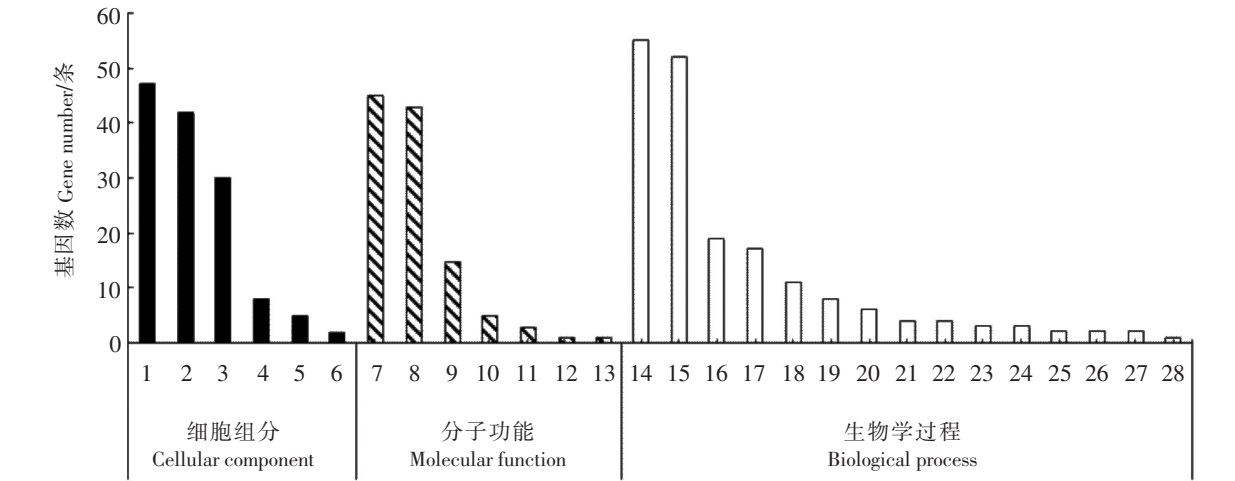
1.6 核心调控网络生成

将获得的基因、Blast2Go 结果、表达量数据、调控关系和相关系数运用 CSV 文件编辑器软件，按照 Cytoscape 格式（基因 1、基因间的关系、基因 2）编辑生成 genelist 格式的文件，将文件导入 Cytoscape 中生成初步调控网络图^[13]。从初步网络中筛选出与神经发育、细胞分化、次生壳形成、能量代谢相关的基因，构建核心调控网络。

2 结果

2.1 序列分析与分类

用 Blast2Go 软件对序列拼接和筛选得到的 203 条序列进行 Blast、Mapping 和 Annotation 分析，剔除无结果序列（24 条）、结果重复及错误序列（13 条）和只有假定蛋白序列（42 条），最终选出具有有效功能注释的序列 124 条。通过基因本体论（GO）分类，可以将这些基因归为细胞组分（cellular component）、分子功能（molecular function）和生物学过程（biological process）3 类，分别占到总基因（203 条）的 66.00%、55.66% 和 88.18%（如图 1 所示）。



说明：1—细胞；2—细胞器；3—高分子配合物；4—胞外区域；5—膜；6—膜封闭腔；7—结合；8—催化活性；9—结构分子活性；10—运输活性；11—酶调节活性；12—分子传递活性；13—核酸结合转录因子活性；14—细胞过程；15—代谢过程；16—单一生物过程；17—组织起源细胞组件；18—局部过程；19—刺激反应；20—发育过程；21—繁殖；22—生物调节；23—信号；24—多细胞生物过程；25—运动；26—免疫系统过程；27—生长；28—生物黏附。

Notes:1—cell;2—organelle;3—macromolecular complex;4—extracellular region;5—membrane;6—membrane—enclosed lumen;7—binding;8—catalytic activity;9—structural molecule activity;10—transporter activity;11—enzyme regulator activity;12—molecular transducer activity;13—nucleic acid binding transcription factor activity;14—cellular process;15—metabolic process;16—single—organism process;17—cellular component organization or biogenesis;18—localization;19—response to stimulus;20—developmental process;21—reproduction;22—biological regulation;23—signaling;24—multicellular organismal process;25—locomotion;26—immune system process;27—growth;28—biological adhesion.

图 1 基因 GO 注释一级和二级分布

Fig.1 Distribution of GO annotations of identified genes that belong to both the primary and secondary functions

在 134 条具有细胞组分注释的基因中，细胞（cell）、细胞器（organelle）和高分子配合物（marcomolecular complex）条数最多，分别为 47 条、42 条和 30 条，占差异基因的 23.15%、20.69% 和 14.78%。有分子功能注释的基因有 113 条，其中具有结合（binding）、催化活性（catalytic activity）和结构分子活性（structural molecule activity）功能的基因条数最多，分别为 45 条、43 条

和 15 条, 占差异基因的 22. 17%、21. 18% 和 7. 39%。具有生物学过程注释的基因有 179 条, 其中细胞过程 (cellular process)、代谢过程 (metabolic process) 和单—的生物过程 (single-organism process) 条数最多, 分别为 55 条、52 条和 19 条, 占差异基因的 30. 73%、29. 05% 和 10. 61%。

2.2 表达趋势归类及网络构建

根据表达趋势划分的 4 类基因 (见表 2) 中, I 类 (表达量升高) 基因有 42 条, 占总数的 33. 87%, 其中代谢最多 (12 条), 信号传递次之 (8 条), 基质蛋白和免疫应激并列第三 (5 条), 核糖体组分最少 (1 条)。II 类 (表达量降低) 基因有 50 条, 占总数的 40. 32%, 其中, 代谢最多 (17 条), 分子相互作用基因次之 (10 条), 细胞骨架基因第三 (8 条), 核糖体组分基因最少 (1 条)。III 类 (变态表达量最高) 基因有 28 条, 其中核糖体组分基因最多, 有 7 条; 信号传递基因次之, 有 5 条; 没有发现钙离子结合功能的基因。IV 类 (变态表达量最少) 基因仅有 4 条, 其中 3 条属于代谢基因, 1 条属于信号传递基因。

表 2 基因表达量及功能分类
Tab.2 Expression and functional classification of identified genes

条										
类别 Category	钙离子 结合 Calcium binding	分子相互 作用 Molecular interaction	代谢 Metabolism	免疫应激 Immune stress	膜蛋白 Membrane protein	基质蛋白 Matrix protein	细胞骨架 Cytoskeleton	核糖体组分 Ribosome	信号传递 Signal transmission	总计 Total
I	1	4	12	5	3	5	3	1	8	42
II	1	10	17	4	2	2	8	1	5	50
III	0	3	4	2	1	3	3	7	5	28
IV	0	0	3	0	0	0	0	0	1	4
总计 Total	2	17	36	11	6	10	14	9	19	124

2.3 分子相互作用网络构建

用统计学软件 SPSS 对所有基因两两之间的表达量相关系数 r 进行计算。结果共获得 7099 个配对, 其中显著性结果有 560 个, 占总结果数的 7. 9%; 极显著结果有 216 个, 占总结果数的 3. 0%。运用 Excel 提取 $P < 0. 05$ 的 r 值, 并使对应的基因——配对。将前期获得的注释结果、变化趋势和相关系数经 CSV 文件编辑器处理后, 提交到 Cytoscape 软件中生成初步相互作用网络。从初步网络中手工筛选神经发育、组织分化、次生壳形成、能量代谢相关基因 38 个, 再以这 38 个基因构建核心调控网络 (见表 3、图 2)。

由表 3、图 2 可见, 网络有 7 个核心基因, 分别是辅酶合成加氧酶 7 同系物 (*coq7*)、刺猬蛋白 (*hh*)、下连合器脊椎蛋白 (*sspo*)、DNA 结合蛋白抑制剂 (*id2*)、G 蛋白亚基 1 (*gng1*)、神经增殖及分化和控制蛋白 1 (*npdc1*)、未知软体动物壳蛋白 13 (*lusp-13*)。以它们为中心, 分别形成了 2 个大集团 (集团 1 和集团 2) 和 1 个小集团 (集团 3)。集团 1 主要由 III 类和 IV 类基因组成。核心分子 *coq7* 为 III 类基因, 与其他相互作用分子有正相关关系, 并且这些相互作用分子几乎都为 III 类基因。另一个核心分子 *hh* 与集团 2 的核心分子 *npdc1* 具有正相关关系, 将两个集团联系起来。集团 2 主要由 I 类和 II 类基因组成, 形成了 2 条具有正相关关系的相互作用链 *dpys-flot1-try-3-csad-npdc1-sspo* 和 *cyt-c-dynll2-nxnl-id2-h3f3a-npdc1*, 1 条具有负相关关系的作用链 *dpys-gabarap-nlp-dynlt1-if-h3f3a-npdc1*。此外, *gng1* 等 3 个变态期表达量最高的基因与钙调蛋白 (*cam*) 形成一个具有正相关作用的小集团。

表 3 核心网络中基因注释结果
Tab.3 Annotation of genes of core network

序号 Number	注释 Annotation	基因 Gene	功能分类 Functional Classification	类别 Category
1	神经外营养蛋白 Nucleoredoxin-like	<i>nxnl</i>	代谢 Metabolism	II
2	细胞色素 C Cytochrome C	<i>cyt-c</i>	代谢 Metabolism	II
3	视网膜色素上皮脊椎蛋白 Rpe-spondin	<i>rpesp</i>	代谢 Metabolism	IV
4	死亡相关蛋白 Death-associated protein	<i>dap</i>	代谢 Metabolism	IV
5	二氢嘧啶酶 Dihydropyrimidinase	<i>dpys</i>	代谢 Metabolism	I
6	假定酪氨酸酶蛋白 3 Putative tyrosinase-like protein tyr-3	<i>try-3</i>	代谢 Metabolism	I
7	半胱亚磺酸脱羧酶 Cysteine sulfinic acid decarboxylase,partial	<i>csad</i>	代谢 Metabolism	I
8	NADH 脱氢酶亚基 6 Nadh dehydrogenase subunit 6	<i>nd6</i>	代谢 Metabolism	III
9	肌动蛋白抑制蛋白 Profilin	<i>pfn</i>	分子相互作用 Molecular Interaction	II
10	微管蛋白阿尔法 3 链 Tubulin alpha-3 chain	<i>tub3</i>	分子相互作用 Molecular Interaction	II
11	下连合器脊椎蛋白 Sco-spondin	<i>sspo</i>	分子相互作用 Molecular Interaction	I
12	神经增殖、分化和控制蛋白 1 Neural proliferation differentiation and control protein 1	<i>npdc1</i>	分子相互作用 Molecular Interaction	I
13	钙调蛋白 Calmodulin	<i>cam</i>	钙离子结合 CalciumBinding	II
14	40S 核糖体蛋白质 S6 40 S ribosomal protein S6	<i>rps6</i>	核糖体组分 Ribosome	III
15	珍珠质样蛋白 2 Nacrein-like 2 protein	<i>nlp</i>	基质蛋白 Matrix Protein	I
16	未知软体动物壳蛋白 13 Uncharacterized shell protein 13(usp13)	<i>lusp13</i>	基质蛋白 Matrix Protein	III
17	筑丝蛋白 3 Tektin-3	<i>tekt3</i>	基质蛋白 Matrix Protein	III
18	筑丝蛋白 3 样蛋白 Tektin-3-like	<i>tekt3-like</i>	基质蛋白 Matrix Protein	III
19	组蛋白 H3.3 Histone h3.3	<i>h3f3a</i>	免疫应激 Immune Stress	II
20	细胞质膜微囊蛋白-1(窖蛋白) Caveolin-1	<i>cav1</i>	膜蛋白 Membrane Protein	II
21	浮舰蛋白 Flotillin-1	<i>flot1</i>	膜蛋白 Membrane Protein	I
22	鞘脂激活蛋白原 Proactivator polypeptide	<i>psap</i>	膜蛋白 Membrane Protein	III
23	外周致密纤维蛋白 3 Outer dense fiber protein 3	<i>odf3</i>	细胞骨架 Cytoskeleton	II
24	胞质动力蛋白轻链 2 Dynein light chain 2, cytoplasmic-like	<i>dynll2</i>	细胞骨架 Cytoskeleton	II
25	胞质动力蛋白轻链 1 Dynein light chain tetex-type 1	<i>dynlt1</i>	细胞骨架 Cytoskeleton	II
26	软骨基质蛋白 Cartilage matrix protein	<i>cmp</i>	细胞骨架 Cytoskeleton	I
27	细胞质中间丝蛋白 Cytoplasmic intermediate filament protein a	<i>if</i>	细胞骨架 Cytoskeleton	I
28	肌球蛋白必需轻链 Myosin essential light chain	<i>melc</i>	细胞骨架 Cytoskeleton	III
29	轴丝相关蛋白 mst101 Axoneme-associated protein mst101-like	<i>cep101</i>	细胞骨架 Cytoskeleton	III
30	DNA 结合蛋白抑制剂 DNA-binding protein inhibitor id-2	<i>id2</i>	信号传递 Signal Transmission	II
31	γ-氨基丁酸受体相关蛋白 Gamma-aminobutyric acid receptor-associated protein	<i>gabarp</i>	信号传递 Signal Transmission	II
32	刺猬蛋白 Protein hedgehog	<i>hh</i>	信号传递 Signal Transmission	IV
33	前列腺癌抗原 Protein dd3-3-like	<i>dd3</i>	信号传递 Signal Transmission	I
34	G 蛋白亚基 1 Guanine nucleotide-binding protein subunit gamma-1	<i>gng1</i>	信号传递 Signal Transmission	III
35	细胞周期 S 期激酶相关蛋白 1 S-phase kinase-associated protein 1	<i>skp1</i>	信号传递 Signal Transmission	III
36	B 细胞迁移基因 B-cell translocation gene 1	<i>btg1</i>	信号传递 Signal Transmission	III
37	超气门蛋白 Usp-like protein isoform 1	<i>usp</i>	信号传递 Signal Transmission	III
38	辅酶合成加氧酶 7 同系物 Ubiquinone biosynthesis protein coq7 homolog	<i>coq7</i>	信号传递 Signal Transmission	III

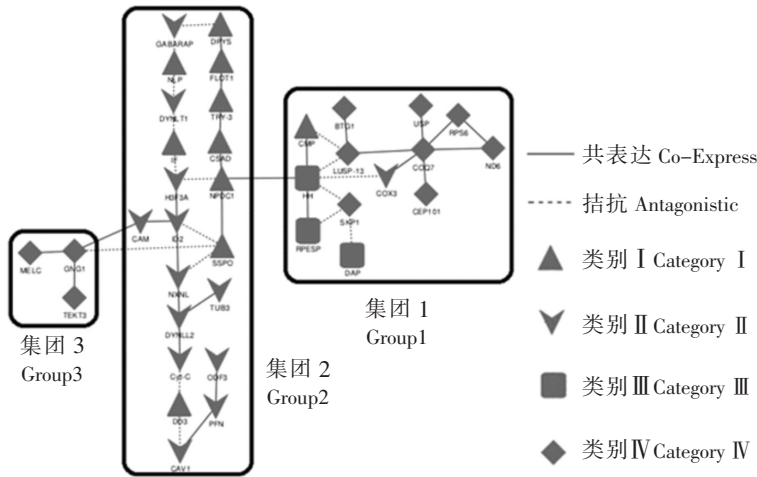


图 2 笠贝幼虫变态相关基因相互作用核心网络
Fig.2 Core network of *L. gigantea* larval of metamorphosis-related genes

3 讨论

利用表达量数据构建分子调控网络是分子生物学系统研究分子相互作用的新方法，它能有效利用几何级数增长的生物信息数据，从更宏观、更全面的角度去研究各类生物学问题^[14]。NCBI 的 DDD 基因差异表达分析，是通过比较不同 cDNA 文库中同一基因 EST 数量，来获得这一基因的表达差异的，其本质上与比较二代测序转录组差异基因分析是相同的。而且 DDD 利用的 cDNA 文库是 NCBI 整理认可的，文库测序数据皆由实验获得，其可信度有保障^[15]。目前有很多利用二代测序转录组的方式进行基因变态差异的研究，定量 PCR 的结果与测序分析具有很好的一致性^[16-18]。这表明基于测序的基因差异表达分析具有较高的可信度。本文的分析方法是对爆炸性增长的测序数据发掘利用的一次尝试。

笠贝与鲍同属于腹足类，变态过程类似。笠贝已经完成基因组测序，转录组数据非常丰富。利用笠贝丰富的核酸序列资源，能够方便地获得笠贝幼虫变态的基因表达过程。这对分析相近物种鲍的变态分子机制，具有重要参考意义。笠贝是目前在 DDD 中与杂色鲍唯一近缘的物种，其原产地在北美，材料获取困难，因此本研究结果并未进行定量 PCR 的验证。但是笔者根据本文预测结果，在杂色鲍中进行相关基因的定量 PCR 检测，其结果与构建的网络一致性很好。这表明 DDD 的分析结果具有较高的可靠性。利用 DDD 中的近缘物种进行分析，能够提高研究的目的性。

在所获得差异表达基因中，参与细胞学过程的基因最多，细胞、细胞器相关基因占总体比例最高，表明幼虫变态过程中细胞分化增殖剧烈，这与变态过程中新组织的形成和幼虫特有组织的崩解凋亡相关。参与代谢过程的基因数量次之，表明变态过程中能量分配对变态成功率有重要影响，这与 Moran 和 Shilling 的研究结果相符^[19-20]。贝类幼虫变态前能量的储备是变态成功的关键，一般个体较大的幼虫变态率较高^[21]。处于变态期的幼虫同初期面盘幼虫相比，自身体重增加显著，其为保持浮游状态需要消耗更多能量。变态期幼虫运动活跃，面盘纤毛摆动力度和频率都明显增加。这些因素可能是代谢相关基因占总体比重较高的部分原因。

根据表达趋势划分的 4 类基因中，I 类基因可能与幼虫的发育成熟关系较大，它们为变态后幼虫的存活进行物质和功能上的储备。如，ODC 是多胺产生的关键酶。多胺参与昆虫卵黄发生、胚胎发育、变态、行为等生理调节过程^[22]；CSAD 是牛磺酸和亚牛磺酸合成的关键酶，二者能促进脂肪及脂溶性维生素的消化吸收。CYP356A1 可能与外源性生物转化、类固醇代谢有关^[23]。这些基因表达量的上升表明幼虫在消化系统方面一直在完善，与鲍幼虫具有“‘Anticipatory’ Developmental Program”

的现象一致^[24]. II 类基因与胚胎发育关系密切, 所以随着发育进程结束表达量逐渐降低. II 类基因中代谢类的基因最多, 其中变化最剧烈的 5 个基因 *cox2*、*cox3*、*cyt - c*、*nd2*、*nd3* 都参与氧化磷酸化过程, 17 条代谢类基因中有 12 条与氧化磷酸化有直接关系. 这表明笠贝胚胎发育至面盘幼虫阶段能量消耗巨大, 且与 Vavra 等^[25]在红鲍中的研究结果一致. III 类基因在变态过程表达量最高, 可能是变态过程的重要参与者. III 类基因中核糖体组分最多, Dong 等^[26]在研究棉铃虫 (*Helicoverpa armigera*) 的变态过程时也发现核糖体组分基因差异的表达, 可能如文献 [21] 所说是与作为细胞增殖和细胞凋亡的调节蛋白直接参与了变态过程有关; 信号传递基因数量次之, 表明这一时期幼虫的信号传递系统进一步完善, 能够响应外界变态诱导因子启动变态过程, 这是贝类变态的特点之一^[21,28]. IV 类基因只有 4 个, 分别是 *dap*、*hh*、*abat*、*rpesp*, 它们可能是变态过程的抑制因子. 例如 *dap* 与 TGFB 相关, 共同调节细胞凋亡^[28]; *hh* 能影响 MAPK 通路, 抑制脂肪细胞发育^[29]. 在核心分子网络中, 集团 1 主要由 III 类和 IV 类基因组成, *coq7* 和 *hh* 是其核心分子. 与 *coq7* 相关的基因主要以能量代谢为主, 如 *usp* 是蜕皮激素受体靶标基因, 参与脂肪代谢, 而 *nd6* 和 *cox3* 参与氧化磷酸化过程; 与 *hh* 相关的基因有 *cmp*、*lusp - 13*、*rpesp*、*skp1*、*dap* 等, 它们与细胞周期有密切联系. 集团 2 以 *npdc1* 为核心分子, 形成了 2 条具有正相关关系的相互作用链和 1 条具有负相关关系的作用链, 多与神经发育有关.

目前关于软体动物的各类研究成果, 相对于人、鼠等高等生物来说比较少, 数据的可获得性和数据质量在一定程度上限制了网络的预测精度. 另外, 本文中选取的神经发育、组织分化、次生壳形成、能量代谢相关基因的限制范围较小, 对最终的结果有一定影响, 这也是本研究的局限之一. 随着生物信息学技术的不断进步和可利用的生物信息学研究方法的完善, 将不断有更简便、高效、准确的方法诞生, 同时数据库的样本丰度也会不断提高. 下一步, 笔者将会使用更先进的数据筛选和关联模型, 通过增加差异化表达数据来源, 参考同源比对、亚细胞定位、Gene Ontology 信息来进一步提高网络模型的预测精度.

[参 考 文 献]

- [1] UNDERWOOD A, KEOUGH M J. Supply-side ecology: the nature and consequences of variations in recruitment of intertidal organisms [J]. Marine Community Ecology, 2001, 87(1): 183-200.
- [2] VEENSTRA J A. Neurohormones and neuropeptides encoded by the genome of *Lottia gigantea*, with reference to other mollusks and insects [J]. General and Comparative Endocrinology, 2010, 167(1): 86-103.
- [3] 王志铮, 吴常文. 浙北沿海嫁戚 *Cellanator* 年龄与生长的研究 [J]. 浙江海洋学院学报: 自然科学版, 2000(4): 316-323.
- [4] SCHEURLE D, DEYOUNG M P, BINNINGER D M, et al. Cancer gene discovery using digital differential display [J]. Cancer Research, 2000, 60(15): 4037-4043.
- [5] OLESEN C, HANSEN C, BENDSEN E, et al. Identification of human candidate genes for male infertility by digital differential display [J]. Molecular Human Reproduction, 2001, 7(1): 11-20.
- [6] 顾朝辉. 生物信息挖掘新基因预测平台的建立与 TSEG-3 的克隆和功能初步研究 [D]. 武汉: 华中科技大学, 2010.
- [7] CONESA A, GÖTZ S, GARCÍA-GÓMEZ J M, et al. Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research [J]. Bioinformatics, 2005, 21(18): 3674-3676.
- [8] WILLIAMS E A, DEGNAN B M, GUNTER H, et al. Widespread transcriptional changes pre-empt the critical pelagic-benthic transition in the vetigastropod *Haliotis asinina* [J]. Molecular Ecology, 2009, 18(5): 1006-1025.
- [9] 栾德琴. 鸡肌肉生长相关基因的表达与肌苷酸关键酶基因网络调控的构建 [D]. 扬州: 扬州大学, 2012.
- [10] 刘天飞, 唐国庆, 李学伟. 不同实验类型的基因表达数据聚类分析方法研究 [J]. 畜牧兽医学报, 2009, 40(2): 180-184.
- [11] 崔焕先. 肉鸡肌肉脂肪形成的分子调控网络及相关基因研究 [D]. 北京: 中国农业科学院, 2011.

- [12] 张永新, 李子健, 闫洁, 等. 利用苯肾上腺素诱导新生大鼠心肌细胞肥大基因表达谱构建其调控网络 [J]. 中国病理生理杂志, 2009(11): 2081-2087.
- [13] NEEDHAM C J, MANFIELD I W, BULPITT A J, et al. From gene expression to gene regulatory networks in *Arabidopsis thaliana* [J]. Bmc Systems Biology, 2009, 3(1): 85.
- [14] 张哲, 荣起国. 数据整合方法构建大鼠分子调控网络 [J]. 系统仿真学报, 2009(5): 1479-1483.
- [15] HE T, CHEN J, ZHANG J, et al. Sarp19 and Vdg3 gene families are functionally related during abalone metamorphosis [J]. Dev Genes Evol, 2014, 224(4): 197-207.
- [16] ZHANG X, MAO Y, HUANG Z, et al. Transcriptome analysis of the octopus vulgaris central nervous system [J]. Plos One, 2012, 7(6): 825-833.
- [17] QIN J, HUANG Z, CHEN J, et al. Sequencing and denovo analysis of *Crassostrea angulata* (Fujian oyster) from 8 different developing phases using 454 gsflx [J]. Plos One, 2012, 7(8): 1322-1337.
- [18] HUANG Z, CHEN Z, KE C, et al. Pyrosequencing of *Haliotis diversicolor* transcriptomes: insights into early developmental molluscan gene expression [J]. Plos One, 2012, 7(12): 1599-1614.
- [19] SHILLING F M, HOEGH-GULDBERG O, MANAHAN D T. Sources of energy for increased metabolic demand during metamorphosis of the abalone *Haliotis rufescens* (mollusca) [J]. The Biological Bulletin, 1996, 191(3): 402-412.
- [20] MORAN A L, MANAHAN D T. Energy metabolism during larval development of green and white abalone, *Haliotis fulgens* and *H. sorenseni* [J]. The Biological Bulletin, 2003, 204(3): 270-277.
- [21] 王国栋, 张丽莉, 王艺磊. 鲍幼虫变态分子机制的研究进展 [J]. 集美大学学报: 自然科学版, 2012, 17(2): 101-108.
- [22] SPARKS R B, VASKE D, LILLEBERG S, et al. Temporal expression of ornithine decarboxylase in developing embryos of *Musca domestica* [J]. Archives of Insect Biochemistry and Physiology, 1991, 16(3): 177-187.
- [23] TOLEDO-SILVA G D, SIEBERT M N, MEDEIROS I D, et al. Cloning a new cytochrome P450 isoform (*cyp356a1*) from oyster *Crassostrea gigas* [J]. Marine Environmental Research, 2008, 66(1): 15-18.
- [24] DEGNAN B M, MORSE D E. Developmental and morphogenetic gene regulation in *Haliotis rufescens* larvae at metamorphosis [J]. American Zoologist, 1995, 35(4): 391-398.
- [25] VAVRA J, MANAHAN D T. Protein metabolism in lecithotrophic larvae (gastropoda: *Haliotis rufescens*) [J]. The Biological Bulletin, 1999, 196(2): 177-186.
- [26] DONG D, HE H, CHAI L, et al. Identification of genes differentially expressed during larval molting and metamorphosis of *Helicoverpa armigera* [J]. Bmc Developmental Biology, 2007, 7(1): 73.
- [27] 柯才焕, 冯丹青. 海洋底栖动物浮游幼体附着和变态的研究 [J]. 厦门大学学报: 自然科学版, 2006, 45(A02): 77-82.
- [28] 李仲诚, 阮礼波, 杨帆, 等. 死亡相关凋亡诱导蛋白激酶2 (DRAK2) 研究进展 [J]. 生命科学, 2014, 26(8): 823-828.
- [29] CHEN J, TANG Q. Hedgehog signaling pathway and adipocyte development [J]. Chinese Journal of Biochemistry and Molecular Biology, 2011, 27(1): 6-10.

(责任编辑 朱雪莲 英文审校 张子平)