

基于中国明对虾不同组织 EST 数据的基因表达差异分析

隗健凯^{1,2}, 柳承璋¹, 张晓军¹, 王兵¹, 董波¹, 李富花¹, 相建海^{1*}

(1. 中国科学院海洋研究所实验海洋生物学重点实验室, 山东 青岛 266071;

2. 中国科学院大学, 北京 100049)

摘要: 对前期获得的中国明对虾头胸部、血液、眼柄、卵巢 4 种组织的 EST 测序数据进行了生物信息学挖掘和分析。4 种组织的原始 EST 序列分别为 10 446 条、2 690 条、1 067 条和 1 282 条, 通过聚类拼接得到 unigene 3 454 条、1 053 条、406 条和 544 条, 对其进行了 NR、GO、KEGG 等数据库的功能注释, 并在此基础上进行了组织差异分析。通过同源比对的方法找出了各组织特异的转录本, 并根据注释信息将其进行了 GO 功能分类。结果显示, 血液组在细胞杀伤、迁移和节律等功能分类上特异性富集; 眼柄组在色素生成、信号传递和生物学调控等功能分类上特异性富集; 卵巢组在营养贮存运输、繁殖和定位等功能分类上特异性富集。拼接时聚类到一起的 EST 数目在一定程度上可以代表基因的表达量, 据此对各组织的高表达基因进行了分析。结果显示, *peritrophin*、*elongation factor 1-alpha*、*thrombospondin* 和 *arginine kinase* 4 个基因在组织中分布较为广泛且表达量高, 提示其可能参与对虾多种重要的生物学过程。而在血液组中注释到的高表达基因还包括 *penaeidin*、*cytochrome c oxidase* 和 *14-3-3 like protein* 基因, 在眼柄组中注释到的高表达基因还包括 *arrestin* 和 *rhodopsin* 基因。此外, 为了解各组织共有基因的差异表达情况, 对 *peritrophin* 和 *peroxiredoxin* 进行了分析, 发现了 *peritrophin* 基因的多形式和高表达现象。

关键词: 中国明对虾; 表达序列标签(EST); 组织差异; 组织高表达基因

中图分类号: Q 785; S 917.4

文献标志码: A

表达序列标签(expressed sequence tag, EST)是将 mRNA 反转录成 cDNA 并克隆到载体上构建成 cDNA 文库后, 随机挑选 cDNA 克隆并对其 5'端或 3'端进行单向测序所获得的序列^[1]。近年来 EST 技术已经被广泛应用于基因识别、分子标记开发、基因图谱绘制、新基因寻找和基因芯片等领域, 是开展基因表达研究的重要参考数据。在海水养殖动物中, 对虾的 EST 研究开展较早, Leher 等^[2]在 1999 年首次报道了对虾 EST 测序, 并在斑节对虾(*Penaeus monodon*)中获得了 176 条 EST 序列。随后凡纳滨对虾(*Litopenaeus vannamei*)^[3]、白滨对虾(*L. setiferus*)^[3]、日本囊对虾(*P. japonicus*)^[4]和中国明对虾

(*Fenneropenaeus chinensis*)^[5]的 EST 测序和分析也相继开展。截止 2012 年 9 月 7 日, NCBI 数据库中公布的 EST 数目最多的对虾为凡纳滨对虾, 共计 161 241 条, 其次是斑节对虾, 共计 39 397 条, 第三为中国明对虾, 共计 10 446 条。在对虾中开展的 EST 研究主要集中在两方面: 一是从 EST 数据库中鉴定感兴趣的基因, 二是利用 EST 调查不同条件下基因表达差异情况^[6], 如通过分析 EST 数据库已有的大量与免疫、繁殖和性别相关的基因了解对虾在不同刺激下的基因差异表达情况。

中国明对虾是我国特有的海产品, 在海水养殖产业中占有重要的地位。近年来围绕其种质

收稿日期:2012-10-09 修回日期:2013-01-28

资助项目:国家“八六三”高技术研究发展计划(2012AA10A404, 2012AA092205); 国家自然科学基金项目(30972245, 41076101, 31172396)

通信作者:相建海, E-mail:jhxiang@qdio.ac.cn

资源、新品种培育、病害和环境等关键性问题开展了大量基础性研究^[7-10]。为了深入开展中国明对虾功能基因组学研究,本课题组构建了中国明对虾不同组织的 cDNA 文库^[11],并获得了 10 446 条头胸部 EST 测序结果,随后又开展了中国明对虾血液、眼柄和卵巢 3 种组织的 EST 测序并获得了 5 039 条测序结果。通过对中国明对虾头胸部文库的 10 446 条 EST 和血液文库的 2 371 条 EST 测序结果进行分析,已经获得了大量与对虾免疫相关的基因^[12-13],并利用其开展了 EST-SSR 的初步研究,得到了 229 个 SSR 分子标记^[14]。

对不同组织进行 EST 测序是在转录水平上研究对虾组织功能差异的重要手段,然而目前有关对虾多个特定组织基因表达情况的比较分析还很少有报道。通过对来自不同组织的 EST 进行生物信息学分析,有助于从整体角度了解对虾的免疫应答、激素调节、生长繁殖和性别控制等生理活动中的基因表达情况和比较不同组织间的差异,同时也为今后深入开展中国明对虾转录水平的相关研究提供了基础。

1 材料与方法

1.1 材料和数据来源

于 2011 年从构建的中国明对虾头胸部 cDNA 文库中随机挑取克隆,采用 MegaBACE1000 测序仪进行 5'端测序,得到 EST 测序原始数据 10 446 条,数据上传至 NCBI dbEST 数据库(登录号: BG891023-BG891738, BG901403-BG903971, BG936683-BG937304, BM293906-BM299611, BM302286-BM303124)。于 2004 年从构建的中国明对虾血液、眼柄和卵巢 3 个组织的 cDNA 文库中随机挑取克隆,采用 MegaBACE1000 测序仪进行 5'端测序。分别得到 EST 测序原始数据 2 690 条、1 067 条和 1 282 条。

1.2 EST 数据的预处理和聚类拼接

测序获得的 EST 数据需要先进行预处理才可以进行后续的聚类拼接。先通过对原始序列的过滤,去除载体序列、污染序列和 PolyA 部分,同时屏蔽重复序列和过短序列。根据本研究所获取的 EST 数据的特点,首先利用 seqclean 软件进行预处理,然后利用 TGICL 软件进行聚类和拼接

(<http://compbio.dfci.harvard.edu/tgi/software/>),最终得到 unigene。

1.3 unigene 的注释和功能分类

使用 BlastX 将 unigene 与 NR 数据库和 uniprot 数据库进行比对, E 值设置为 $1e-5$ 。使用 BlastN 将 unigene 与 NT 数据库进行比对, E 值设置为 $1e-5$ 。对在 NR 库中没有匹配的部分再使用 InterProscan 功能域搜索,完善其注释信息。使用自编脚本获取 GO 号,并设置为 WEGO 网站提供的 native format 格式后,提交到该网站(<http://wego.genomics.org.cn/cgi-bin/wego/index.pl>),进行 Gene ontology (GO) 功能分类。将 unigene 的序列信息提交到 KEGG 的 KAAS 在线工具(<http://www.genome.jp/tools/kaas/>),选择 BBH 方式与 KEGG 数据库比对,得到 unigene 在各通路图上的映射关系。

1.4 组织差异分析

利用大规模测序方法开展不同组织转录水平研究,能够有效地筛查组织差异表达的基因。本研究利用不同组织的 EST 序列单独聚类拼接产生的 unigene 作为参考进行组织特异性分析。将其中一组的 unigene 与其它三组合并在一起的 unigene 进行本地 BlastN 比对,没有产生匹配的,即作为组织特异的转录本,同时将每个组织特异的转录本进行注释与 GO 功能分类。根据其在 GO 分类中所占比例的组织差异得到各组织的特异性富集分类。

在使用 TGICL 进行聚类拼接时,有 overlap 区的 EST 会首先聚类到一起,然后利用重叠关系拼接成 contig。由于在测序过程中是随机挑取的克隆,因此这些聚类到一起的 EST 数目在一定程度上就代表了基因的表达量。在每个组织中按照 contig 包含的 EST 数目进行排列,选取 EST 数目最多的前十条 contig 进行分析,从而了解不同组织高表达基因的差异。同样,有些表达部位分布广泛的基因在多个组织中均被检测到,但基因表达的形式和数目会有所差别,可以通过比较不同组织中注释为同一基因的 contig 所包含的 EST 数目与总测序量的比值情况,来了解同一基因在不同组织的表达差异。对在 4 种组织中广泛存在的 *peritrophin* 和 *peroxiredoxin* 相关基因进行了分析,来说明 EST 数据在基因表达差异研究中的作用。

2 结果

2.1 原始数据的预处理和聚类拼接

原始数据如表 1 所示,从中可以看出第一批测序的头胸部 EST 数据量较大,但平均长度相对较短,只有 372 bp。第二批测序的血液、眼柄和卵巢 EST 数据平均长度可达 500 bp 左右。说明第二批 3 种组织的测序质量高于更早期进行的头胸部测序。经过预处理和聚类拼接之后得到的 unigene 的平均长度同样是血液、眼柄、卵巢这 3 种组织较长,可达到 600 bp 左右;而合并后得到的 unigene 平均长度为 478 bp,最长 unigene 为 2.9 kb(表 2)。

表 1 EST 测序原始数据的基本情况
Tab.1 The basic information of EST raw data

组织分布 tissue distribution	数量/条 number	平均长度/bp average length	GC 含量/% GC content
头胸部	10 446	372	42.73
血液	2 690	462	46.79
眼柄	1 067	539	49.91
卵巢	1 282	481	49.23
合并	15 485	408	44.82

表 2 EST 数据的预处理和聚类拼接结果
Tab.2 The result of preprocessing and assembling of EST data

	头胸部 cephalothorax	血液 blood	眼柄 eyestalk	卵巢 ovary	合并 total
原始 EST 数目	10 446	2 690	1 067	1 282	15 485
预处理后数目	10 075	2 367	1 023	1 169	14 632
contig 数目	1 025	277	90	148	1 369
singlet 数目	2 429	776	316	396	3 366
unigene 数目	3 454	1 053	406	544	4 735
unigene 平均长度	411	589	625	595	478
最长 unigene 长度	2 952	2 556	2 079	2 180	2 901

2.2 unigene 的注释和分类

各组织 unigene 的注释结果统计见表 3。从 GO 分类情况来看(图 1),注释到的蛋白从免疫调节到生长发育等多种生物学过程都有涉及,而每个组织均有其特异的功能分类特点。从柱形图的数量比例分布来看,其中卵巢组与其他组的 GO 分类柱形图差异最明显。显示卵巢在行使生物学功能上的独特性。从柱形图种类分布来看,GO 分类中 virion part、rhythmic process 和 locomotion 为血液组特有, nutrient reservoir

activity 为卵巢组特有, virion 仅在头胸部组和血液组中出现, transcription regulator activity 仅在头胸部组和眼柄组中出现, cell killing 仅在血液组和眼柄组中出现, viral reproduction 仅在血液组和卵巢组中出现。

从合并组的 KEGG 通路注释情况来看,注释到的通路涉及较广,包括 MAPK、Wnt、JAK-STAT、Apoptosis、Endocrine 等信号通路,但受到测序量和物种差异的限制,只有糖酵解等少数几个途径的注释较为完整,大部分通路的注释都不完整,仅仅是注释到通路上的个别基因,限制了在通路水平上开展中国明对虾组织差异的分析。比如免疫反应中的重要通路 JAK-STAT 只注释到 *PI3K* 和 *CBP* 两个基因。这表明要想得到较为理想的通路注释,依赖高通量测序获取大量数据将是一个重要途径。

表 3 各组织 unigene 的注释结果
Tab.3 Annotation result of unigenes in each tissue

	头胸部 cephalothorax	血液 blood	眼柄 eyestalk	卵巢 ovary	合并 total
unigene 数目	3 454	1 053	406	544	4 735
NR 注释数目	1 179	589	250	382	1 897
uniprot 注释数目	975	528	227	345	1 616
GO 注释数目	293	174	97	116	432
KEGG 注释数目	513	325	137	243	788
pathway 数目	191	197	147	156	238

2.3 不同组织特异 unigene 的分析

从各组特异 unigene 所占比例来看(表 4),头胸部的 unigene 中不同于其他 3 种组织的特异 unigene 比例最高,其次是血液组,而眼柄组和卵巢组的特异 unigene 所占比例较低数量也较少。从各组织特异 unigene 对 NR 蛋白数据库的注释情况来看,头胸部因包含对虾大量功能器官,其相对其他 3 组的特异 unigene 注释的蛋白种类也较为多样。血液组的特异 unigene 注释到多种与细胞凋亡相关的蛋白,如 caspase、autophagy protein 5、inhibitor of apoptosis protein 等,提示细胞凋亡相关信号通路在该时期血细胞中相当活跃。眼柄组特异 unigene 注释到的蛋白主要与视觉相关(如 arrestin、opsin、rhodopsin)和神经调控相关(如 ceramide kinase、survival motor neuron protein)。卵巢组特异 unigene 注释到的蛋白主要包括与生长繁殖相关(Ras-related protein Rab-1A、juvenile

hormone epoxide hydrolase 和 vitellogenin)、性别相关(anti-estrogen resistance protein)和线粒体相关(mitochondrial ubiquinol-cytochrome c reductase)的多种功能蛋白。

根据特异 unigene 的 GO 分类(图 2)可知,血液组的特异 unigene 在 virion part、locomotion、rhythmic process、cell killing 和 multi-organism process 特异性富集,主要与细胞杀伤、迁移和节律等功能相关。眼柄组的特异 unigene 在

molecular transducer activity、response to stimulus、multicellular organismal process、pigmentation 和 biological regulation 特异性富集,主要与信号传递和生物学调控等功能相关。而卵巢组的特异 unigene 在 nutrient reservoir activity、reproduction、anatomical structure formation、transporter activity 和 establishment of localization 特异性富集,主要与营养贮存运输、繁殖和定位等功能相关。

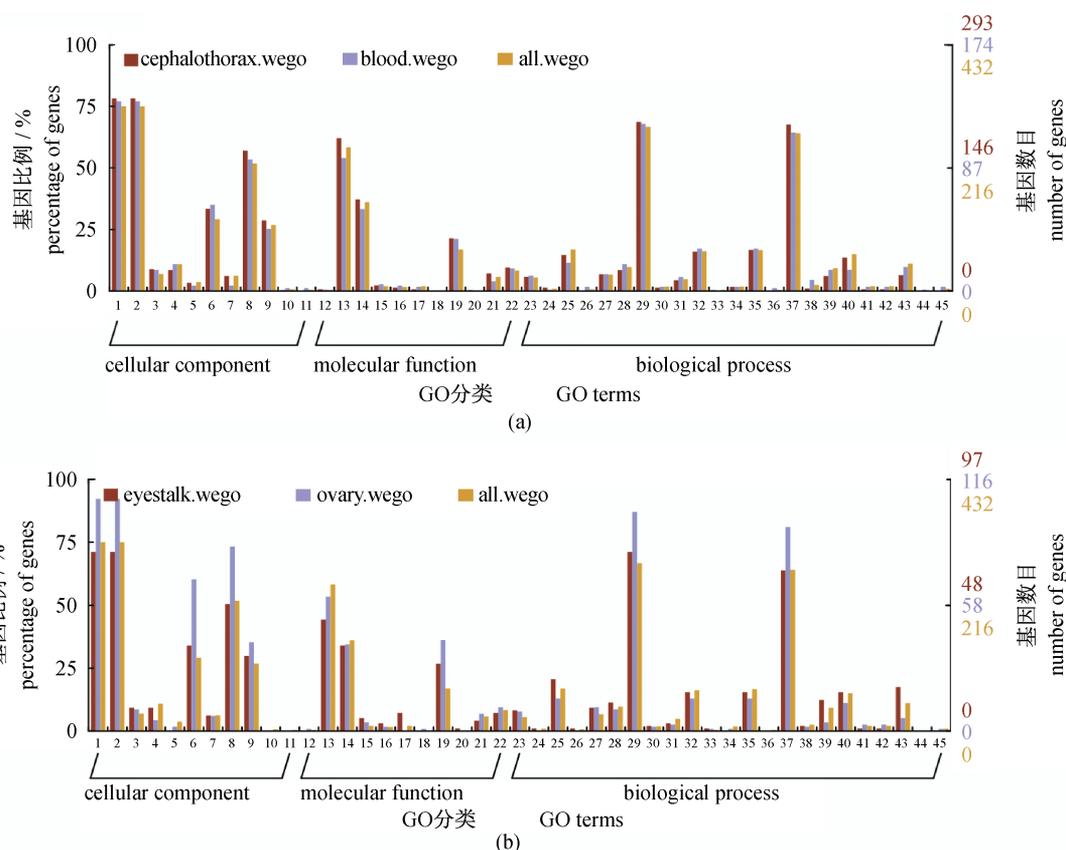


图 1 各组织 unigene 的 GO 分类柱形图

(a) 头胸部、血液、合并组的 GO 分类结果; (b) 眼柄、卵巢、合并组的 GO 分类结果。

Fig. 1 GO categories of unigenes in each tissue

(a) GO categories in cephalothorax, blood and total group; (b) GO categories in eyestalk, ovary and total group.

1. Cell; 2. Cell part; 3. Envelope; 4. Extracellular region; 5. Extracellular region part; 6. Macromolecular complex; 7. Membrane-enclosed lumen; 8. Organelle; 9. Organelle part; 10. Virion; 11. Virion part; 12. Antioxidant; 13. Binding; 14. Catalytic; 15. Electron carrier; 16. Enzyme regulator; 17. Molecular transducer; 18. Nutrient reservoir; 19. Structural molecule; 20. Transcription regulator; 21. Translation regulator; 22. Transporter; 23. Anatomical structure formation; 24. Biological adhesion; 25. Biological regulation; 26. Cell killing; 27. Cellular component biogenesis; 28. Cellular component organization; 29. Cellular process; 30. Death; 31. Developmental process; 32. Establishment of localization; 33. Growth; 34. Immune system process; 35. Localization; 36. Locomotion; 37. Metabolic process; 38. Multi-organism process; 39. Multicellular organismal process; 40. Pigmentation; 41. Reproduction; 42. Reproductive process; 43. Response to stimulus; 44. Rhythmic process; 45. Viral reproduction. The same as below.

表 4 各组织特异 unigene 数目和注释情况
Tab.4 The number and annotation of tissue specific unigenes

	头胸部 cephalothorax	血液 blood	眼柄 eyestalk	卵巢 ovary
unigene 数目	3 454	1 053	406	544
特异 unigene 数目(相对其他 3 种组织)	2 699	438	94	75
特异 unigene 占总 unigene 的比例	78%	42%	23%	14%
特异 unigene 中 singlet 数目	2 029	378	89	70
特异 unigene 中 contig 数目	670	60	5	5
特异 unigene 中 singlets 所占比例	75%	86%	95%	93%
有注释的特异 unigene 数目	745	189	41	39
有 GO 分类注释的特异 unigene 数目	149	44	13	9

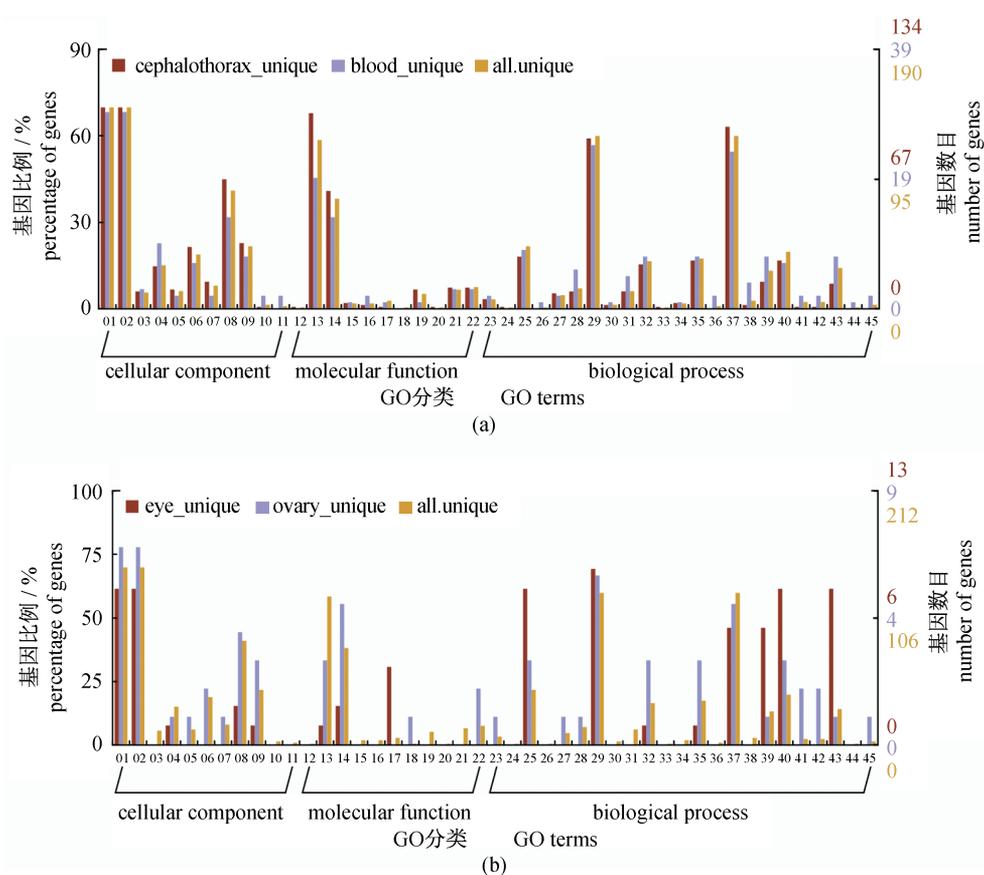


图 2 各组织特异 unigene 的 GO 分类柱形图

(a) 头胸部、血液、合并组特异 unigene 的 GO 分类结果; (b) 眼柄、卵巢、合并组特异 unigene 的 GO 分类结果。

Fig.2 GO categories of tissue specific unigenes

(a) GO categories in cephalothorax, blood and total group; (b) GO categories in eyestalk, ovary and total group.

2.4 组织高表达 unigene 的分析

不同组织中,表达量前十位的 unigene 如表 5 所示。在头胸部组中,包含 EST 数目最多的 contig 为 CL1Contig11,其最佳比对结果是 hypothetical protein [*Thermobia domestica*],但是只有 158 bp 具有同源比对结果,e 值也只有 $4e-09$ 。经过 NT 库比对发现,此序列全长与中国明对虾线粒体基因组

IrRNA 序列高度同源,而 IrRNA 是线粒体基因组中编码线粒体核糖体大亚基 (rrnL) 的区域。推测这是源于建库过程中混入了对虾的线粒体 DNA。同样在血液组中的 CL1Contig1、眼柄组中的 CL6Contig1 和卵巢组中的 CL1Contig54 也是如此。这也说明了转录组数据 NT 库注释对于 NR 库注释具有补充作用。

表 5 表达量前十位的 unigene 统计
Tab.5 The top ten highly expressed unigenes

(1) 头胸部组包含的 EST 数目在前十位的 unigene 列表

编号 name	包含的 EST 数目 EST number	NR 库的注释信息 NR annotation
CL1Contig11	712	hypothetical protein[<i>Thermobia domestica</i>]
CL1Contig17	380	peritrophin[<i>Fenneropenaeus chinensis</i>]
CL1Contig14	327	peritrophin[<i>Fenneropenaeus chinensis</i>]
CL1Contig15	278	peritrophin[<i>Fenneropenaeus chinensis</i>]
CL1Contig18	141	peritrophin[<i>Fenneropenaeus chinensis</i>]
CL1Contig7	140	peritrophin[<i>Fenneropenaeus chinensis</i>]
CL1Contig19	139	elongation factor 1-alpha[<i>Marsupenaeus japonicus</i>]
CL2Contig1	113	thrombospondin[<i>Penaeus monodon</i>]
CL5Contig1	97	putative thrombospondin[<i>Fenneropenaeus chinensis</i>]
CL4Contig2	84	unknown

(2) 血液组包含的 EST 数目在前十位的 unigene 列表

编号 name	包含的 EST 数目 EST number	NR 库的注释信息 NR annotation
CL1Contig1	173	hypothetical protein[<i>Thermobia domestica</i>]
CL3Contig1	49	elongation factor 1-alpha[<i>Marsupenaeus japonicus</i>]
CL5Contig1	36	cytochrome c oxidase subunit I[<i>Fenneropenaeus chinensis</i>]
CL2Contig2	35	beta-actin[<i>Litopenaeus vannamei</i>]
CL6Contig1	32	The Human Nucleosome Structure Containing Histone Variant H3.3
CL7Contig1	30	unknown
CL9Contig1	25	arginine kinase[<i>Fenneropenaeus chinensis</i>]
CL4Contig2	21	penaeidin 3-1[<i>Fenneropenaeus chinensis</i>]
CL2Contig1	18	actin E[<i>Litopenaeus vannamei</i>]
CL10Contig1	18	14-3-3-like protein[<i>Penaeus monodon</i>]

(3) 眼柄组包含的 EST 数目在前十位的 unigene 列表

编号 name	包含的 EST 数目 EST number	NR 库的注释信息 NR annotation
CL3Contig1	74	arrestin2 precursor[<i>Dianemobius nigrofasciatus</i>]
CL1Contig2	62	peritrophin[<i>Fenneropenaeus chinensis</i>]
CL5Contig1	55	arginine kinase[<i>Fenneropenaeus chinensis</i>]
CL6Contig1	44	hypothetical protein[<i>Thermobia domestica</i>]
CL2Contig4	34	arrestin-like protein[<i>Camponotus floridanus</i>]
CL4Contig2	30	rhodopsin[<i>Litopenaeus vannamei</i>]
CL2Contig1	24	arrestin homolog[<i>Locusta migratoria</i>]
CL8Contig1	21	arrestin-like protein[<i>Acromyrmex echinaior</i>]
CL1Contig1	16	peritrophin[<i>Fenneropenaeus chinensis</i>]
CL4Contig4	14	rhodopsin[<i>Litopenaeus vannamei</i>]

(4) 卵巢组包含的 EST 数目在前十位的 unigene 列表

编号 name	包含的 EST 数目 EST number	NR 库的注释信息 NR annotation
CL1Contig47	44	arginine kinase[<i>Fenneropenaeus chinensis</i>]
CL1Contig18	39	peritrophin[<i>Fenneropenaeus chinensis</i>]
CL1Contig7	31	peritrophin[<i>Fenneropenaeus chinensis</i>]
CL1Contig21	30	thrombospondin[<i>Penaeus monodon</i>]
CL2Contig1	24	elongation factor 1-alpha[<i>Marsupenaeus japonicus</i>]
CL1Contig38	20	peritrophin[<i>Fenneropenaeus chinensis</i>]
CL1Contig54	20	putative puroindoline b protein[<i>Triticum aestivum subsp. macha</i>]
CL1Contig19	19	unknown
CL1Contig24	16	peritrophin[<i>Fenneropenaeus chinensis</i>]
CL3Contig1	16	unknown

在头胸部组中,排在 2~6 位的 5 个 contig 全部注释为中国明对虾 *peritrophin*。另外排在前十位的还包括注释为 elongation factor 1-alpha、thrombospondin 的 unigene 和一个没有注释的 unigene。在血液组中,排在前十位的包括注释到 elongation factor 1-alpha、cytochrome c oxidase、beta-actin、Histone Variant H3.3、arginine kinase、penaeidin 3-1、actin E 和 14-3-3-like protein 的 unigene 以及一个没有注释的 unigene。对于眼柄组而言,排在前十位的 unigene 的注释包括 arrestin、peritrophin、rhodopsin 和 arginine kinase 4 类蛋白。对于卵巢组而言,包括注释到 arginine

kinase、peritrophin、thrombospondin 和 elongation factor 1-alpha 4 类和两个没有注释信息的 unigene。

2.5 共有基因的差异表达情况

编码 *peritrophin* 的基因在不同组织中分布广泛且表达量高,为了深入分析 *peritrophin* 基因的功能,对各组织中 *peritrophin* 基因的表达差异进行了分析。从表 6 和图 3 可知,注释为 *peritrophin* [*Fenneropenaeus chinensis*] 的 unigene 在头胸部、眼柄、卵巢 3 组中包含的 EST 数目比例相当大,而血液中只有 1 条包含 4 个 EST 的 contig 注释为 *peritrophin*。同时在不同组织中还注释到了其他

表 6 各组织注释为 *peritrophin* 及其相关基因的 unigene 所对应的 EST 总数统计

Tab. 6 The EST number of unigenes annotated as *peritrophin*

Nr 库注释结果 Nr annotation	头胸部 cephalothorax	血液 blood	眼柄 eyestalk	卵巢 ovary
<i>peritrophin</i> [<i>Fenneropenaeus chinensis</i>]	1 291	4	90	115
ovarian <i>peritrophin</i> [<i>Fenneropenaeus merguensis</i>]	42	-	-	-
<i>peritrophin</i> -like protein 1 [<i>Penaeus semisulcatus</i>]	2	-	-	-
<i>peritrophin</i> -like protein 2 [<i>Penaeus semisulcatus</i>]	23	-	-	-
ovarian <i>peritrophin</i> 1 precursor [<i>Penaeus monodon</i>]	1	-	-	-
ovarian <i>peritrophin</i> 2 precursor [<i>Penaeus monodon</i>]	-	-	10	-

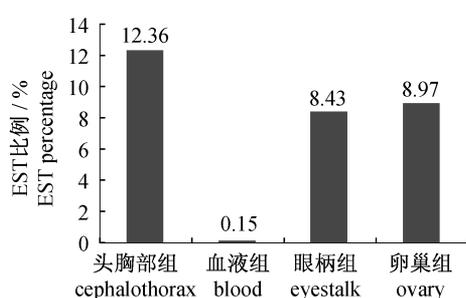


图 3 各组织注释为 *peritrophin* [*Fenneropenaeus chinensis*] 的 unigene 对应的 EST 占总 EST 的比例统计

Fig. 3 The EST percentage of unigenes annotated as *peritrophin* [*Fenneropenaeus chinensis*]

的一些 *peritrophin* 相关蛋白,比如在头胸部当中注释到了不同形式的 ovarian *peritrophin*,眼柄中也注释到一个特异表达的 *peritrophin* 类型 ovarian *peritrophin* 2 precursor。这些都表明 *peritrophin* 是一类表达部位分布较广泛且表达量很高的基因,可能行使多种重要的生物学功能。

从表 7 可知,peroxiredoxin 也在 4 个组织当中均被注释到,但对应的 unigene 包含的 EST 数目均较少。从注释情况来看,peroxiredoxin-6 分布广泛,peroxiredoxin-5 则只在血液组织的 EST 中发现。这种组织分布上的差异提示这两种不同的 peroxiredoxin 在行使功能上的区别。

表 7 注释为 peroxiredoxin 相关基因的 unigene 所包含的 EST 数目

Tab. 7 The EST number of unigenes annotated as peroxiredoxin

	头胸部 cephalothorax	血液 blood	眼柄 eyestalk	卵巢 ovary
peroxiredoxin [<i>Fenneropenaeus indicus</i>]	7	2	2	1
peroxiredoxin [<i>Litopenaeus vannamei</i>]	1	-	-	-
peroxiredoxin-5 [<i>Aedes aegypti</i>]	-	1	-	-
peroxiredoxin-6 [<i>Caligus rogercresseyi</i>]	5	-	-	-
peroxiredoxin 6 [<i>Scylla paramamosain</i>]	-	1	-	1

3 讨论

对虾头胸部包含了鳃、心脏、肝胰腺、胃、肠道和淋巴等众多器官,因此头胸部 EST 在一定程度上可以代表整个对虾的转录本情况,其基因表达的多样性相应更高,对其进行的测序量也最高,主要用于获取更全面的转录本信息。随后选择血液、眼柄和卵巢 3 个文库进行 EST 序列的测定,目的在于深入了解对虾的免疫应答、激素调节、生长繁殖和性别控制等生理活动。课题组前期对这些 EST 序列的应用侧重于寻找对虾重要的免疫相关功能基因^[12-13];而本研究则侧重于利用生物信息学手段进行对虾基因表达的组织差异分析,对揭示对虾组织功能的特异性和开展相关功能基因的研究提供重要的参考。

各组织特异 unigene 的比例不同,出现这种情况一方面与组织的功能特异性相关,头胸部包含众多功能器官,其基因表达的多样性也相应较高。另一方面也受到测序量的影响,本研究所测数据还不足以检测到各组织表达的所有基因。测序深度增大才会检测到更多低丰度的基因,而特异性表达的调控基因很多倾向于低的表达丰度^[15],在各组特异 unigene 中 singlet 所占比例都很高也说明了这一点。另外,很多组织特异表达的基因具有时空特异性,往往只有在特定的发育时期或者特定的刺激下才会表达^[16]。受限于 Sanger 测序通量的影响,还有很多发挥重要调控功能的低丰度表达基因未能检测到,通路的注释也不够完善,需要通过高通量测序来弥补。因此开展不同组织、不同胁迫条件和不同发育时期的转录组高通量测序,对于进一步开展中国明对虾转录水平的研究具有重要意义,也是对虾重要功能基因筛选和开发的必然趋势。

受到测序深度的影响,本研究通过相互同源比对得到的特异 unigene,其是否为组织特异表达还需要进一步实验验证。然而这些特异 unigene 在功能上的分布却可以直接体现各组织的功能差异。血液组主要在细胞杀伤、迁移和节律等功能上特异性富集,与血液中血细胞负责完成物质运输和免疫防御两大功能是相关的;眼柄组主要在信号传递和生物学调控等功能上特异性富集,体现了眼柄是发挥内分泌调节功能的重要部位,同时提示眼柄可能是大量神经细胞的富集部位^[17];

卵巢组主要在营养贮存运输、繁殖和定位等功能上特异性富集,暗示卵巢在发育过程中物质贮存的重要性^[18]。

在高表达基因的注释结果中,编码 peritrophin、elongation factor 1-alpha、thrombospondin 和 arginine kinase 的 4 个基因在组织中分布较为广泛且表达量高。其中 peritrophin 除了有助于对摄入食物的消化外,还具有保护动物抵御病原体入侵的功能。已发现的对虾类围食膜蛋白大都含有 3 个或 4 个几丁质结合结构域^[19]。elongation factor 1-alpha 不仅是翻译必须的蛋白,而且是一个重要的多功能蛋白。它参与信号传导、翻译控制、凋亡、细胞骨架组成和病毒复制等过程^[20]。thrombospondin 作为细胞外的糖蛋白,参与了细胞与细胞之间以及细胞与基质之间的信息交流,并在细胞和组织的生长与分化中起重要作用^[21]。arginine kinase 在无脊椎动物体内能量代谢过程中能够提供和维持能量的平衡,在生命活动的调控中占据重要地位^[22]。

在血液组中,注释到的高表达基因还包括编码 penaeidin、cytochrome c oxidase 和 14-3-3-like protein 的基因。其中 penaeidin 是在中国明对虾中发现的最大的抗菌肽家族,主要在血细胞中表达和储存,并在受到刺激后释放^[23]。而 cytochrome c oxidase 是位于线粒体呼吸链终末端的关键酶,不但在耗氧产能的过程中发挥重要的调节作用,而且还可以影响应激条件介导的细胞凋亡^[24]。14-3-3-like protein 是一种广泛分布于各种真核生物中的多功能蛋白,参与细胞信号转导、代谢酶合成和细胞凋亡等多种重要生命活动的调控^[25]。同时也有研究表明,它与生物体渗透压的调节密切相关,在一些海洋生物中,盐度的变化可以引起 14-3-3 基因的显著变化^[26],这为研究生物体盐度耐受的机理提供了很好的参考。在眼柄组中,注释到的高表达基因还包括编码 arrestin 和 rhodopsin 的基因。其中 arrestin 蛋白主要分布在视网膜上,对介导光信号传导的视紫红质脱敏起重要作用^[27],而 rhodopsin 是一类重要的 G 蛋白偶联受体,在受到光子激活时可以发生变构^[28]。对 rhodopsin 的研究有助于进一步了解对虾对光线强弱的感应和生物节律现象的分子机理。

在 peritrophin 高表达的 3 组中,均是多条不

同的 contig 同时注释为 peritrophin [*Fenneropenaeus chinensis*], 例如在头胸部中包含 EST 数目排在 2 ~ 6 位的 contig 均注释为 peritrophin [*Fenneropenaeus chinensis*]. 推测中国明对虾存在多种形式的 peritrophin 序列, 该基因在基因组上也可能为多基因家族类型。在昆虫等其他节肢动物中也有类似的情况和报道, Jasarapuria 等^[29] 曾在赤拟谷盗 (*Tribolium castaneum*) 基因组中对包含此类几丁质结合结构域的蛋白进行搜索, 找到了 29 个相关蛋白并根据系统发生、表达及功能将其分为 PMP、CPAP3 和 CPAP1 3 类。在对虾 EST 中发现 peritrophin 的高表达和多形式现象暗示该类蛋白在对虾适应环境和形成自我防护屏障过程中也可能具有重要作用。

peroxiredoxin 是广泛存在于原核生物和真核生物体内的一类抗氧化蛋白酶, 它的主要功能是清除机体内的过氧化氢, 调节由过氧化氢介导的信号转导, 是机体实现先天性免疫反应的重要组成部分^[30]。peroxiredoxin-6 与过氧化物还原酶家族的其它成员不同, 它不使用硫氧还蛋白作为还原剂。越来越多的研究证明 peroxiredoxin-6 在体内具有多方面的作用, 包括可以降解脂质氢过氧化物, 而其他的 peroxiredoxin 类型则还没有发现具有这样的功能^[31], 这也体现了 peroxiredoxin 在对虾发挥免疫功能上的多样性。

参考文献:

- [1] Nagaraj S H, Gasser R B, Ranganathan S. A hitchhiker's guide to expressed sequence tag (EST) analysis [J]. Briefings in Bioinformatics, 2007, 8 (1):6 - 21.
- [2] Lehnert S A, Wilson K J, Byrne K, et al. Tissue-specific expressed sequence tags from the black tiger shrimp *Penaeus monodon* [J]. Marine Biotechnology, 1999, 1(5):465 - 476.
- [3] Gross P S, Bartlett T C, Browdy C L, et al. Immune gene discovery by expressed sequence tag analysis of hemocytes and hepatopancreas in the Pacific white shrimp, *Litopenaeus vannamei*, and the atlantic white shrimp, *L. setiferus* [J]. Developmental and Comparative Immunology, 2001, 25(7):565 - 577.
- [4] Rojtiinnakorn J, Hirono I, Itami T, et al. Gene expression in haemocytes of kuruma prawn, *Penaeus japonicus*, in response to infection with WSSV by EST approach [J]. Fish & Shellfish Immunology, 2002, 13(1):69 - 83.
- [5] Xiang J, Bing W, Li F, et al. Generation and Analysis of 10, 443 ESTs from cephalothorax of *Fenneropenaeus chinensis* [C]. Shanghai: The 2nd International Conference on Bioinformatics and Biomedical Engineering (iCBBE 2008), 2008.
- [6] Leu J H, Chen S H, Wang Y B, et al. A Review of the major Penaeid shrimp EST studies and the construction of a shrimp transcriptome database based on the ESTs from four Penaeid shrimp [J]. Marine Biotechnology, 2011, 13(4):608 - 621.
- [7] Xiang J H, Li F H, Zhang C S, et al. Evaluation of induced triploid shrimp *Penaeus (Fenneropenaeus chinensis)* cultured under laboratory conditions [J]. Aquaculture, 2006, 259(1 - 4):108 - 115.
- [8] 李健, 刘萍, 何玉英, 等. 中国对虾快速生长新品种“黄海 1 号”的人工选育 [J]. 水产学报, 2005, 29(1):1 - 5.
- [9] Wang B, Li F, Dong B, et al. Discovery of the genes in response to white spot syndrome virus (WSSV) infection in *Fenneropenaeus chinensis* through cDNA microarray [J]. Marine Biotechnology, 2006, 8(5):491 - 500.
- [10] Gao H A, Kong J, Li Z J, et al. Quantitative analysis of temperature, salinity and pH on WSSV proliferation in Chinese shrimp *Fenneropenaeus chinensis* by real-time PCR [J]. Aquaculture, 2011, 312(1 - 4):26 - 31.
- [11] 张晓军, 王兵, 张绍萍, 等. 中国对虾 6 种组织 cDNA 文库的构建 [J]. 海洋学报: 中文版, 2005, 27(5):92 - 95.
- [12] Shen Y Q, Xiang J H, Wang B, et al. Discovery of immune related factors in *Fenneropenaeus chinensis* by annotation of ESTs [J]. 自然科学进展: 英文版, 2004, 14(1):8.
- [13] Dong B, Xiang J H. Discovery of genes involved in defense/immunity functions in a haemocytes cDNA library from *Fenneropenaeus chinensis* by ESTs annotation [J]. Aquaculture, 2007, 272(1 - 4):208 - 215.
- [14] Wang H X, Li F H, Xiang H H. Polymorphic EST-SSR markers and their mode of inheritance in *Fenneropenaeus chinensis* [J]. Aquaculture, 2005, 249(1 - 4):107 - 114.
- [15] Morozova O, Hirst M, Marra M A. Applications of new sequencing technologies for transcriptome analysis [J]. Annual Review of Genomics and

- Human Genetics, 2009, 10: 135 – 151.
- [16] Wang Z, Gerstein M, Snyder M. RNA-Seq: A revolutionary tool for transcriptomics [J]. Nature Reviews Genetics, 2009, 10(1): 57 – 63.
- [17] Liu J S, Li S Y, Ye H H, *et al.* Sodium current of neurosecretory cells in the eyestalk from the shrimp *Penaeus japonicus* [J]. Progress in Natural Science, 2009, 19(11): 1493 – 1499.
- [18] Wu P, Qi D, Chen L Q, *et al.* Gene discovery from an ovary cDNA library of oriental river prawn *Macrobrachium nipponense* by ESTs annotation [J]. Comparative Biochemistry and Physiology-Part D: Genomics & Proteomics, 2009, 4(2): 111 – 120.
- [19] Du X J, Wang J X, Liu N, *et al.* Identification and molecular characterization of a peritrophin-like protein from fleshy prawn (*Fenneropenaeus chinensis*) [J]. Molecular Immunology, 2006, 43(10): 1633 – 1644.
- [20] 周冰, 曹诚, 刘传暄. 翻译延伸因子 1A 的研究进展 [J]. 生物技术通讯, 2007, 18(2): 281 – 284.
- [21] 陈春球, 范跃祖. 血小板反应素与消化道肿瘤 [J]. 同济大学学报: 医学版, 2003, 24(4): 292 – 294.
- [22] 苏晓峰, 陆国清, 程红梅. 精氨酸激酶蛋白及分子生物学的研究进展 [J]. 生物技术通报, 2011(4): 26 – 30.
- [23] Kang C J, Xue J F, Liu N, *et al.* Characterization and expression of a new subfamily member of penaeidin antimicrobial peptides (penaeidin 5) from *Fenneropenaeus chinensis* [J]. Molecular Immunology, 2007, 44(7): 1535 – 1543.
- [24] Kadenbach B, Arnold S, Lee I, *et al.* The possible role of cytochrome c oxidase in stress-induced apoptosis and degenerative diseases [J]. Biochimica et Biophysica Acta, 2004, 1655(1–3): 400 – 408.
- [25] Aitken A. 14-3-3 proteins: A historic overview [J]. Seminars in Cancer Biology, 2006, 16(3): 162 – 172.
- [26] Kaeodee M, Pongsomboon S, Tassanakajon A. Expression analysis and response of *Penaeus monodon* 14-3-3 genes to salinity stress [J]. Comparative Biochemistry and Physiology-Part B: Biochemistry & Molecular Biology, 2011, 159(4): 244 – 251.
- [27] 程智洁, 裴钢. Arrestins 家族: 一类在 G 蛋白偶联受体的信号传导中起关键作用的蛋白质 [J]. 生命科学, 1999, 11(4): 169 – 171.
- [28] Hofmann K P, Scheerer P, Hildebrand P W, *et al.* A G protein-coupled receptor at work: The rhodopsin model [J]. Trends in Biochemical Sciences, 2009, 34(11): 540 – 552.
- [29] Jasrapuria S, Arakane Y, Osman G, *et al.* Genes encoding proteins with peritrophin A-type chitin-binding domains in *Tribolium castaneum* are grouped into three distinct families based on phylogeny, expression and function [J]. Insect Biochemistry and Molecular Biology, 2010, 40(3): 214 – 227.
- [30] 张庆利, 李富花, 张继泉, 等. 中国明对虾过氧化物还原酶基因在大肠杆菌中的重组表达、产物纯化及活性测定 [J]. 中国生物工程杂志, 2008, 28(2): 47 – 52.
- [31] 张成林, 李建远. 抗氧化蛋白 Peroxiredoxin 6 研究进展 [J]. 医学研究杂志, 2010, 39(6): 121 – 123.

Differential gene expression analysis based on expressed sequence tags(EST) from different tissues of *Fenneropenaeus chinensis*

WEI Jiankai^{1,2}, LIU Chengzhang¹, ZHANG Xiaojun¹, WANG Bing¹,
DONG Bo¹, LI Fuhua¹, XIANG Jianhai^{1*}

(1. The Key Laboratory of Experimental Marine Biology, Institute of Oceanology,
Chinese Academy of Sciences, Qingdao 266071, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Bioinformatics analysis was conducted based on the acquired expressed sequence tags(ESTs) data from different tissues of *Fenneropenaeus chinensis*. A total of 10 446, 2 690, 1 067 and 1 282 original ESTs were obtained from cephalothorax, blood, eyestalk, and ovary of adult *F. chinensis* respectively. By clustering and assembling, 3 454, 1 053, 406 and 544 unigenes were generated, and then annotated by searching in NR, GO, KEGG databases. The tissue specific transcripts were identified through sequence homology analysis and classified by GO annotation. The result indicated that specific transcripts in blood group were specifically enriched in the GO term of virion part, locomotion, rhythmic process, cell killing and multi-organism process. Specific transcripts in eyestalk group were specifically enriched in the GO term of molecular transducer activity, response to stimulus, multicellular organismal process, pigmentation and biological regulation. Specific transcripts in ovary group were specifically enriched in the GO term of nutrient reservoir activity, reproduction, anatomical structure formation, transporter activity and establishment of localization. We also analyzed the highly expressed genes in each tissue according to the number of ESTs of each unigene. The result indicated that genes encoding *peritrophin*, *elongation factor 1-alpha*, *thrombospondin* and *arginine kinase* were widespread and highly expressed in tissues, suggesting that they were involved in a variety of important biological processes in shrimp. In addition, genes encoding *penaeidin*, *cytochrome c oxidase* and *14-3-3-like protein* were highly expressed in blood group and genes encoding *arrestin* and *rhodopsin* were highly expressed in eyestalk group. For understanding the expressions of common genes, we analyzed genes related to *peritrophin* and *peroxiredoxin*, revealing that genes of *peritrophin* in *F. chinensis* were multiform and highly expressed. This result may provide reference for further functional study of *peritrophin*.

Key words: *Fenneropenaeus chinensis*; expressed sequence tags(EST); tissue variability; tissue highly expressed genes

Corresponding author: XIANG Jianhai. E-mail: jhxiang@ qdio. ac. cn