

# 泥蚶(*Tegillarca granosa*)血细胞转录组序列中免疫相关基因的结构与 mRNA 表达量分析\*

钟爱华 储张杰 王伟洪 牟毅

(浙江海洋大学水产学院 舟山 316022)

**摘要** 贝类血细胞直接参与异物吞噬和包囊作用,还在伤口修复、炎症反应过程中发挥重要作用,是机体免疫防御的承担者。本研究采用 RNA-seq 测序技术研究了鳃弧菌感染前后泥蚶血细胞的转录组序列,通过序列组装,获得了  $\alpha$ -葡萄糖苷酶(GAA)、 $\alpha$ -甘露糖苷酶(LAMAN)、芳基硫酸酯酶 B(ASB)、天冬酰胺氨基葡萄糖酶(AGA)、整合素  $\beta$  3 亚基(ITB3)、热休克蛋白 9(HSPA9)和 Toll 相互作用蛋白(TOLLIP)序列,利用 Orffinder 获得了其编码蛋白序列,并对蛋白序列进行了理化性质、二级结构、保守结构域和三级结构分析。结果表明:7 种蛋白质二级结构均含有  $\alpha$  螺旋、 $\beta$  折叠及无规则卷曲,泥蚶和人类 ITB3 和 TOLLIP 具有相同的保守结构域,Swiss-model 同源建模获得的三级结构与模板同源性均超过 35%(TOLLIP 除外)。鳃弧菌刺激后,GAA、LAMAN、ASB、AGA 和 HSPA9 的表达量显著增加,ITB3 和 TOLLIP 表达量变化不明显。该研究补充和完善了对泥蚶免疫相关基因的认识,并为进一步开展泥蚶抗病机理研究提供了重要的理论依据。

**关键词** 泥蚶;血细胞;转录组测序;免疫基因

中图分类号 S917.4 doi: 10.11693/hyhz20191200251

泥蚶 (*Tegillarca granosa*) 属软体动物门 (Mollusca)、双壳纲(Bivalvia)、列齿目(Taxodonta)、蚶科(Arcidae),富含营养,且含有特有的维生素 B<sub>12</sub> 和血红蛋白,味道鲜美,具有补血养胃和除淤散积的功效,是我国传统养殖贝类(李太武等,2003)。血细胞是机体抵抗病原微生物的重要载体,贝类血细胞能吞噬病原微生物或形成包囊包裹异物,还在伤口修复、炎症反应过程中发挥重要作用(孙敬锋等,2006)。

转录组测序(RNA sequencing)是利用大规模测序技术直接对 cDNA 序列进行测序的技术,能获得细胞在某一状态下所能转录出来的所有 mRNA。RNA-seq 具有通量高、成本低、灵敏度高等优点,是一种有效的基因组研究和功能基因鉴定方法,利用转录组测序研究转录谱特征近年来呈爆发式增长(陈雪峰等,2019; Wang *et al.*, 2019)。目前有关泥蚶研究主要集中在养殖技术、病害防控以及单一基因克隆(李太武等,

2003; 董迎辉等, 2013; Liu *et al.*, 2017; 许家辉等, 2018)等方面。本研究采用 RNA-Seq 技术测序了鳃弧菌刺激前后泥蚶血细胞转录组,通过转录组序列组装获得了 7 个免疫相关基因蛋白序列,并对 7 个免疫相关基因蛋白序列理化性质、二级结构、三级结构、保守结构域、进化发生和 mRNA 表达水平进行了分析,探索其在细菌感染过程中所发挥的作用,以期对泥蚶抗病机理研究提供理论基础。

## 1 材料与方法

### 1.1 样本准备

实验用泥蚶来自浙江省舟山市沈家门国际水产城,选择贝壳无损伤、对外界刺激反应迅速的泥蚶,带回实验室暂养 5d,暂养盐度为 20±1、温度 (18±2)°C、溶解氧高于 5mg/L。暂养结束后,随机挑选 5 个大小相一致的泥蚶置于含有 5×10<sup>6</sup>cfu/mL 鳃弧

\* 浙江省“水产”一流学科开放课题,20160012 号;国家自然科学基金项目,31700322 号。钟爱华,硕士,讲师, E-mail: zhongpe@zjou.edu.cn

收稿日期: 2019-12-08, 收修改稿日期: 2020-01-06

菌悬液中, 5 个泥蚶置于无菌海水中。24h 后, 采用无菌注射器从两组泥蚶外套腔中取血, 获得的血淋巴立即与 Trizol (Invitrogen, CA, USA) 试剂按 1 : 5 混合, 将混合好的血淋巴冻存于液氮中备用。

### 1.2 RNA 提取、cDNA 文库构建与测序

按照 TRIzol<sup>®</sup> Reagent 试剂盒操作步骤提取总 RNA, 提取的总 RNA 采用 Bioanalyzer 2100 和 RNA 1000 Nano LabChip Kit (Agilent, CA, USA) 定性定量试剂盒进行质检, 获得的 RNA 满足总量 5 $\mu$ g、浓度 200ng/ $\mu$ L、OD<sub>260</sub>/OD<sub>280</sub> 为 1.8—2.2、RIN >8.0, 进行文库构建。先使用连接有 Oligo(dT) 的磁珠富集 mRNA, 富集的 mRNA 被片段化试剂(Fragmentation Buffer)随机打断成短片段, 以片段化的 mRNA 为模板, 用六碱基随机引物(Random hexamers)合成一链 cDNA, 随后加入缓冲液、dNTPs、RNaseH 和 DNA Polymerase I 进行二链 cDNA 合成。AMPure XP beads 纯化双链产物, 利用 T4 DNA 聚合酶和 Klenow DNA 聚合酶将 DNA 的粘性末端修复为平末端, 3'末端加碱基 A 并加接头, AMPureXP beads 进行片段选择, 之后用 USER 酶降解含有 U 的 cDNA 第二链, 然后进行 PCR 扩增获得最终测序文库。文库质检合格后采用 Illumina Hiseq4000 进行测序, 测序读长为双端 2\*150bp (PE150)。

### 1.3 转录组 *de novo* 组装

测序产生的原始数据(Raw Date)采用使用 Cutadapt 去除测序 reads 中接头后, 使用 Fqtrim 过滤掉不合格的序列后得到有效数据。具体为先对测序 reads 进行窗口法质量扫描, 扫描窗口默认为 6bp, 当窗口内平均质量值低于 20 时, 将 read 从窗口起始到 3'终止的部分截掉, 去除 poly-A/T、去除截断后长度小于 100bp 以及 N 的含量在 5%以上的序列, 余下的序列为有效数据。过滤后的数据(Clean reads), 使用 Trinity 软件进行从头组装, reads 首先被拼接成长片段 contigs, 去除冗余 contigs 后将 contigs 进行分组, 最后连接成两端不能再延长的转录片段(Unigene)。

### 1.4 基因注释和结构分析

采用软件 DIAMOND, 设定期望值 E Value <1e-5, 在 NCBI-NR (<http://www.ncbi.nlm.nih.gov/>)、GO (<http://www.geneontology.org/>)、KEGG (<http://www.genome.jp/kegg/>)、Pfam (<http://pfam.xfam.org/>)和 Swiss-Prot (<https://www.uniprot.org/>)数据库中 进行 比 对, 获得 Unigenes 注释信息。基于注释信息, 利用 NCBI 数据库 SmartBlast 工具对 7 个候选基因进一步进行比对, 以保证注释信息正确, 7 个候选基因编码区由 NCBI

网站获得(<http://www.ncbi.nlm.nih.gov/>)。

利用 ExPaSy 中的 ProtParam (<http://web.expasy.org/protparam/>)对蛋白序列进行理化性质分析; 利用 SMART (<http://smart.embl-heidelberg.de/>)分析序列保守结构域(包括信号肽)特征; 利用 Prabi 程序([https://npsa-prabi.ibcp.fr/cgi-bin/secpred\\_hnn.pl](https://npsa-prabi.ibcp.fr/cgi-bin/secpred_hnn.pl))对蛋白质二级结构进行预测; 利用在线工具 Swiss-model (<http://swissmodel.expasy.org/>)进行三级结构分析, 选取同源性大于 30% (Seq Identity)、GMQE 和 QMEAN 值大的模板生成蛋白质的三级结构, 然后使用 PyMOL Viwer 软件进行编辑, 最终绘制出蛋白质的三级结构模型图。

### 1.5 差异表达基因分析

采用 Salmon 软件(Patro *et al*, 2017)将测序获得的 Clean reads 比对到 Trinity 组装的序列, 从而获得每个 unigene 上的 read 数目, 然后计算 TPM (Transcripts Per Kilobase Million), 从而获得 Unigene 的表达量。采用 R 软件包中的 edgeR 筛选差异表达基因, 筛选阈值为  $P$  值 <0.05 且  $|\log_2 \text{Fold Change}| > 1$ 。

### 1.6 系统发生分析

以人类 *Homo sapiens*、斑马鱼 *Danio rerio* 及其他水生生物作为参考物种, 利用 MEGA-X 对 7 种免疫相关基因进行 Clustal W 序列比对, 用最大似然法 (ML 法)构建系统发育树, 构建系统发育树前, 先计算最优模型, 采用最优模型进行建树, 同时进行步长循环 1000 检验, 所用模型及序列见表 1。

## 2 结果

### 2.1 测序数据统计、组装和注释

通过高通量测序泥蚶血细胞中共获得 9.01G raw base 和 60001612 raw reads, 过滤后获得 8.75G valid bases 和 59199480 valid reads, valid reads 的 GC (guanine-cytosine) 含量为 42.41%, 组装后, 共获得 21823 条 Unigenes, 原始数据已上传至国家基因组科学数据中心(<https://bigd.big.ac.cn/>, Project:PRJCA002006, 提交编号: subCRA002389)。通过与 Pfam、NR、GO、Swissprot 和 KEGG 数据库比对, 共有 10691 条 Unigenes 与已知基因同源, GO 数据库中注释到 8301 (38.04%) 条, KEGG 中注释到 5376 (24.63%) 条, Swissprot 中注释到 7686 (35.22%) 条, Pfam 中注释到 9868 (36.05%) 条, NR 中注释到 10636 条(48.74%) 条(图 1)。物种同源性分析表明, 与 NR 数据库中太平洋牡蛎 *Crassostrea gigas* 同源序列比例最高, 其次是栉孔扇贝 *Lottia gigantea*(图 1)。

表 1 构建系统进化树所用模型和序列  
Tab.1 Models and sequences for constructing phylogenetic trees

基因名	组装 ID	构建进化树模型	构建进化树所用序列 NCBI 登录号
溶酶体 $\alpha$ -葡萄糖苷酶	TRINITY_DN11864_c0_g5	WAG+G 模型	XP_022317334.1 ( <i>Crassostrea virginica</i> ), EKC19023.1 ( <i>Crassostrea gigas</i> ), XP_021379413.1 ( <i>Mizuhopecten yessoensis</i> ), NP_000143.2 ( <i>Homo sapiens</i> ), XP_009304541.2 ( <i>Danio rerio</i> ), XP_026063328.1 ( <i>Carassius auratus</i> ), TKS68573.1 ( <i>Collichthys lucidus</i> )
溶酶体 $\alpha$ -甘露糖苷酶	TRINITY_DN11881_c0_g1	LG+G 模型	XP_022316085.1 ( <i>Crassostrea virginica</i> ), EKC30993.1 ( <i>Crassostrea gigas</i> ), OWF45656.1 ( <i>Mizuhopecten yessoensis</i> ), NP_000519.2 ( <i>Homo sapiens</i> ), NP_001014354.1 ( <i>Danio rerio</i> ), XP_010737172.3 ( <i>Larimichthys crocea</i> ), XP_025088376.1 ( <i>Pomacea canaliculata</i> )
芳基硫酸酯酶 B	TRINITY_DN11890_c1_g2	LG+G 模型	XP_022329757.1 ( <i>Crassostrea virginica</i> ), XP_011454669.1 ( <i>Crassostrea gigas</i> ), XP_021349527.1 ( <i>Mizuhopecten yessoensis</i> ), NP_000037.2 ( <i>Homo sapiens</i> ), XP_021324609.1 ( <i>Danio rerio</i> ), XP_019120324.1 ( <i>Larimichthys crocea</i> ), XP_025088376.1 ( <i>Pomacea canaliculata</i> )
天冬酰胺氨基葡萄糖酶	TRINITY_DN12019_c0_g9	WAG+G 模型	XP_022321606.1 ( <i>Crassostrea virginica</i> ), EKC18130.1 ( <i>Crassostrea gigas</i> ), XP_021364372.1 ( <i>Mizuhopecten yessoensis</i> ), NP_000018.2 ( <i>Homo sapiens</i> ), NP_001103751.1 ( <i>Danio rerio</i> ), XP_027137564.1 ( <i>Larimichthys crocea</i> ), XP_025087972.1 ( <i>Pomacea canaliculata</i> )
整合素 $\beta$ 亚基	TRINITY_DN11770_c0_g3	WAG+G 模型	XP_022306553.1 ( <i>Crassostrea virginica</i> ), XP_011428304.1 ( <i>Crassostrea gigas</i> ), sp P05106.2  ( <i>Homo sapiens</i> ), NP_001076417.1 ( <i>Danio rerio</i> ), XP_010738332.3 ( <i>Larimichthys crocea</i> ), AVP12656.1 ( <i>Littorina littorea</i> )
热休克蛋白 9	TRINITY_DN11188_c0_g10	LG+G 模型	XP_022288463.1 ( <i>Crassostrea virginica</i> ), EKC18038.1 ( <i>Crassostrea gigas</i> ), XP_021341216.1 ( <i>Mizuhopecten yessoensis</i> ), NP_004125.3 ( <i>Homo sapiens</i> ), NP_958483.2 ( <i>Danio rerio</i> ), XP_010732361.2 ( <i>Larimichthys crocea</i> ), XP_025093768.1 ( <i>Pomacea canaliculata</i> ), XP_029655296.1 ( <i>Octopus vulgaris</i> ), XP_014775850.1 ( <i>Octopus bimaculoides</i> )
Toll 相互作用蛋白	TRINITY_DN9460_c0_g1	LG+G+F 模型	XP_022294590.1 ( <i>Crassostrea virginica</i> ), XP_011451178.1 ( <i>Crassostrea gigas</i> ), AKO62848.1 ( <i>Mizuhopecten yessoensis</i> ), NP_061882.2 ( <i>Homo sapiens</i> ), NP_996944.2 ( <i>Danio rerio</i> ), XP_027137215.1 ( <i>Larimichthys crocea</i> ), XP_025099118.1 ( <i>Pomacea canaliculata</i> ), AET79206.1 ( <i>Penaeus vannamei</i> )

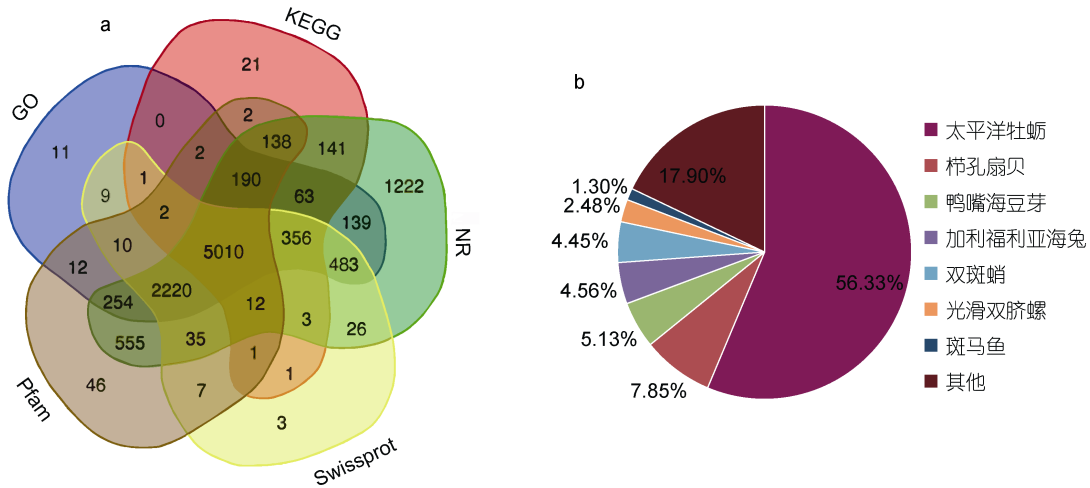


图 1 泥蚶血细胞转录组注释信息

Fig.1 Statistical parameters and functional annotation of *T. granosa* blood cell reference transcriptome  
注: a. 不同数据库 Unigenes 注释数; b. 同源性分析

2.2 溶酶体酶

泥蚶血细胞转录组序列中鉴定出 4 种溶酶体酶, 包括 2 种糖苷酶、硫酸酯酶 B 和天冬酰胺氨基葡萄糖

糖酶。溶酶体  $\alpha$ -葡萄糖苷酶(GAA)长度为 3168bp, 开放阅读框长度为 2347bp, 编码 779 个氨基酸, 预测分子量约为 87.9kDa、理论等电点 6.68; 丝氨酸和

苏氨酸含量最高, 占 7.7%; 不稳定系数为 34.03, 属于稳定蛋白。其 N 端(第 1—96 位氨基酸)是一个  $\beta$  桶状结构, 一般位于真核生物麦芽糖酶-葡糖淀粉酶催化域的 N 端, 属于糖基水解酶 31 的 N-末端保守结构域; 第 97—163 氨基酸属于糖苷水解酶家族 31 的保守结构域; 第 183—651 氨基酸为 O-糖基水解酶 31 结构域(图 2a)。二级结构分析其含有  $\alpha$  螺旋、 $\beta$  折叠及无规则卷曲。SWISS 同源性分析显示泥蚶溶酶体  $\alpha$ -葡萄糖苷酶与模板 5kzw.1.A、5nn3.1.A 同源性均超过 50%, 自动建模只有一种三级结构被建立(图 6a)。多重序列比对显示泥蚶  $\alpha$ -葡萄糖苷酶与人、小鼠和斑马鱼同源性为 50%左右, 与太平洋牡蛎 *Crassostrea gigas*、美洲巨牡蛎 *Crassostrea virginica* 和虾夷扇贝 *Mizuhopecten yessoensis* 同源性超过 60%; 系统进化树表明泥蚶  $\alpha$ -葡萄糖苷酶与虾夷扇贝优先聚为一支, 然后与太平洋牡蛎和美洲巨牡蛎相聚(图 2b)。

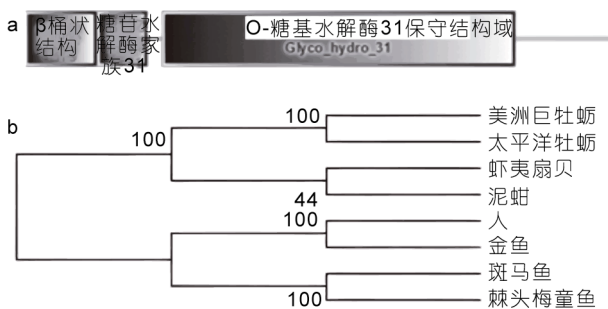


图 2  $\alpha$ -葡萄糖苷酶的结构域(Pfam 数据库)和系统进化树

Fig.2 The conserved domain of GAA and phylogenetic relationships

注: a. 结构域; b. 系统进化树

溶酶体  $\alpha$ -甘露糖苷酶(LAMAN)长度为 3473bp, 开放阅读框长度为 3180bp, 编码 1060 个氨基酸, 预测分子量约为 120.9kDa、理论等电点 8.77; 含量最高的氨基酸为亮氨酸, 占 9.4%; 不稳定系数为 32.94, 属于稳定蛋白; 二级结构由  $\alpha$  螺旋、 $\beta$  折叠及无规则卷曲组成。其 N 端(第 1—24 位氨基酸)具有信号肽, 第 43—360 氨基酸属于糖苷水解酶家族 38 的保守结构域, 第 365—442 为  $\alpha$ -甘露糖苷酶中间保守结构域, 第 491—1054 氨基酸为糖苷水解酶家族 38 的 C 端保守结构域(图 3a); 三级结构显示其与 Swiss-Model 数据库中  $\alpha$ -甘露糖苷酶模板 1o7d.1.A 和 1o7d.1.B 同源性超过 60%, 与模板 1o7d.1.D 同源性超过 55%, 自动建模获得地三级结构见图 6b。多重序列比对显示其与人类、斑马鱼、美洲巨牡蛎和虾夷扇贝同源性超过 50%。系统进化树分析显示泥蚶  $\alpha$ -甘露糖苷酶与美洲巨牡蛎亲缘关系最近(图 3b)。

芳基硫酸酯酶 B(ASB)长度为 1743bp, 开放阅读框为 1536bp, 编码 512 个氨基酸, 预测分子量约为 59.1kDa、理论等电点 6.32; 含量最高的氨基酸为甘氨酸, 占 8.2%; 不稳定系数为 47.08, 属于不稳定蛋白。二级结构分析其含有  $\alpha$  螺旋、 $\beta$  折叠及无规则卷曲。其 N 端具有信号肽和硫酸酯酶结构域(图 4a)。其与 Swiss-model 数据库中模板 B1fsu.1.A 同源性高达 47.01%, 自动建模只有一种三级结构被建立(图 6c)。多重序列比对显示其与太平洋牡蛎、美洲巨牡蛎和虾夷扇贝同源性高达 55%以上, 与人类芳基硫酸酯酶 B 同源性超过 40%; 系统进化分析显示泥蚶芳基硫酸酯酶 B 优先与虾夷扇贝聚为一支, 然后与太平洋牡蛎和美洲巨牡蛎相聚(图 4b)。

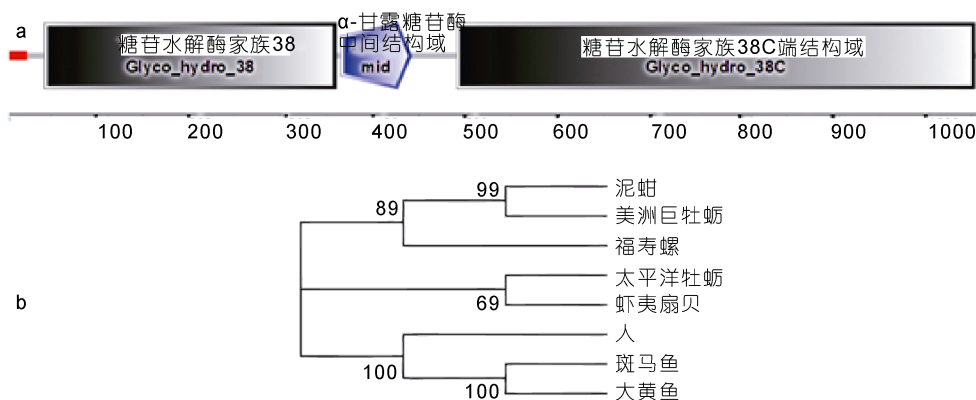


图 3 溶酶体  $\alpha$ -甘露糖苷酶的结构域和系统进化树

Fig.3 The conserved domain of LAMAN and phylogenetic relationships

注: a. 结构域; b. 系统进化树

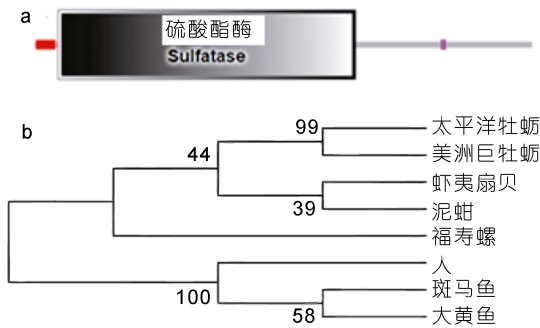


图 4 芳基硫酸酯酶 B 的结构域和系统进化树  
Fig.4 The conserved domain of ASB and phylogenetic relationships  
注: a. 结构域; b. 系统进化树

天冬酰胺氨基葡萄糖酶(AGA)长度为 1256bp, 开放阅读框为 1044bp, 编码 348 个氨基酸, 预测分子量约为 120.9kDa、理论等电点 8.77; 含量最高的氨基酸为亮氨酸, 占 9.4%; 不稳定系数为 32.94, 属于稳定蛋白; 具有  $\alpha$  螺旋、 $\beta$  折叠及无规则卷曲组成的二级结构。第 16—342 位氨基酸残基是天冬酰胺酶 2 结构域(图 5a)。三级结构分析显示其与模板 1p4k.1.A、

4r4y.1.A、同源性超过 40%, 与模板 1apz.1.A 同源性为超过 60%, 以模板 1apz.1.A 建立的三级结构见图 6d。多重序列比对显示其与太平洋牡蛎、美洲巨牡蛎和虾夷扇贝相似度高达 55%以上, 与人类、斑马鱼、大黄鱼等相似度超过 50%。系统进化分析显示泥蚶天冬酰胺氨基葡萄糖酶优先与虾夷扇贝聚为一支(图 5b)。

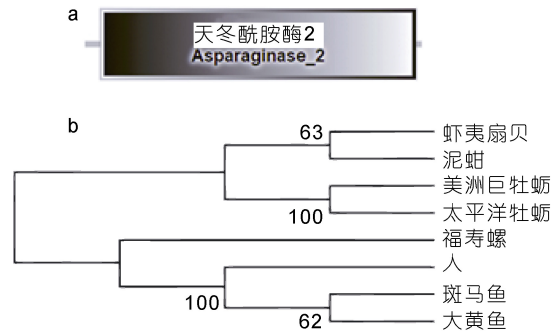


图 5 天冬酰胺氨基葡萄糖酶的结构域和系统进化树  
Fig.5 The conserved domain of AGA and phylogenetic relationships  
注: a. 结构域; b. 系统进化树

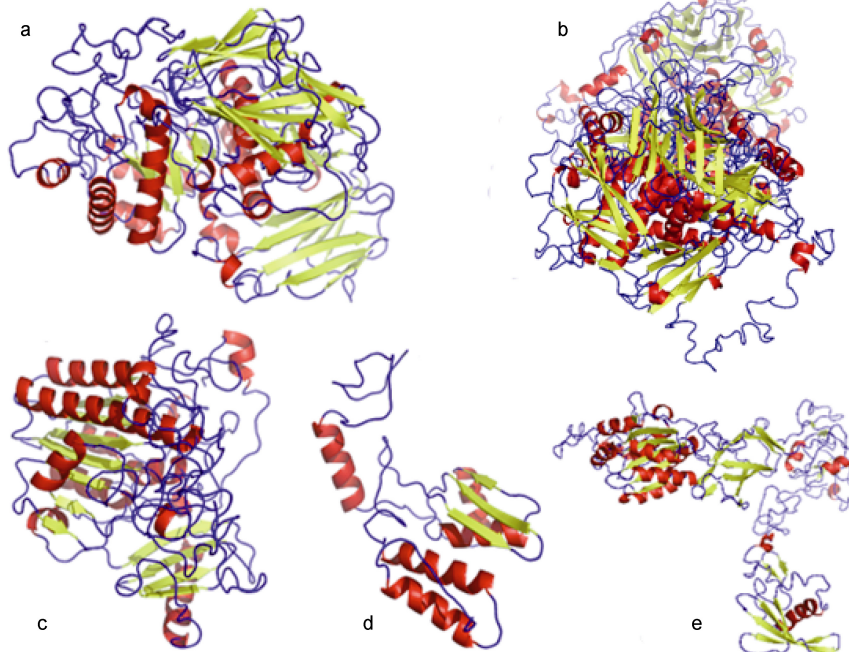


图 6 5 种蛋白质三级结构图  
Fig.6 The tertiary structure of 5 proteins  
注: a. 溶酶体  $\alpha$ -葡萄糖苷酶; b. 溶酶体  $\alpha$ -甘露糖苷酶; c. 芳基硫酸酯酶 B; d. 天冬酰胺氨基葡萄糖酶; e. 整合素  $\beta$ 3 亚基

### 2.3 整合素

泥蚶血细胞转录组序列中, 鉴定出整合素  $\beta$ 3 亚基

(ITB3),  $\beta$ 3 亚基长度为 3455bp, 编码 793 个氨基酸, 预测分子量约为 87.9kDa、理论等电点 4.94; 苏氨酸含量



最高, 占 7.9%; 不稳定系数为 44.51, 属不稳定蛋白。SMART 保守结构域分析其具有人整合素  $\beta 3$  相似结构域, 胞外序列 N 端有信号肽和整合素亚基保守结构域(integrin beta subunits domain, INB)(第 27—452 位氨基酸), 第 631—714 位氨基酸是整合素  $\beta$  尾部保守区域, 第 714—736 位氨基酸为跨膜区域, 胞内有整合素  $\beta$ -cyt 结构域, 泥蚶整合素第 461—622 位氨基酸位置上还有 3 个 EGF 样(epidermal growth factor)重复(图 7b)。具有  $\alpha$  螺旋、 $\beta$  折叠及无规则卷曲组成的二级结

构。通过 Swiss-model 进行蛋白质三级结构同源建模, 结果显示泥蚶整合素  $\beta 3$  与模板 3ije.1.B、6avr.1.B、6bxj.1.A、4mmy.1.B、4mmx.1.B 等同源性均超过 35%, 模板 6avr.1.B 建模效果最优, 其 GMQE(0.63)和 QMEAN(-3.31)值均最大, 泥蚶整合素  $\beta 3$  三级结构见图 6e。NCBI 结果显示泥蚶整合素  $\beta 3$  序列与美洲牡蛎整合素  $\beta 3$  同源性超过 45%, 与人和斑马鱼同源性超过 35%。系统进化树表明: 泥蚶整合素  $\beta 3$  亚基与太平洋牡蛎和美洲巨牡蛎亲缘关系最近(图 7c)。

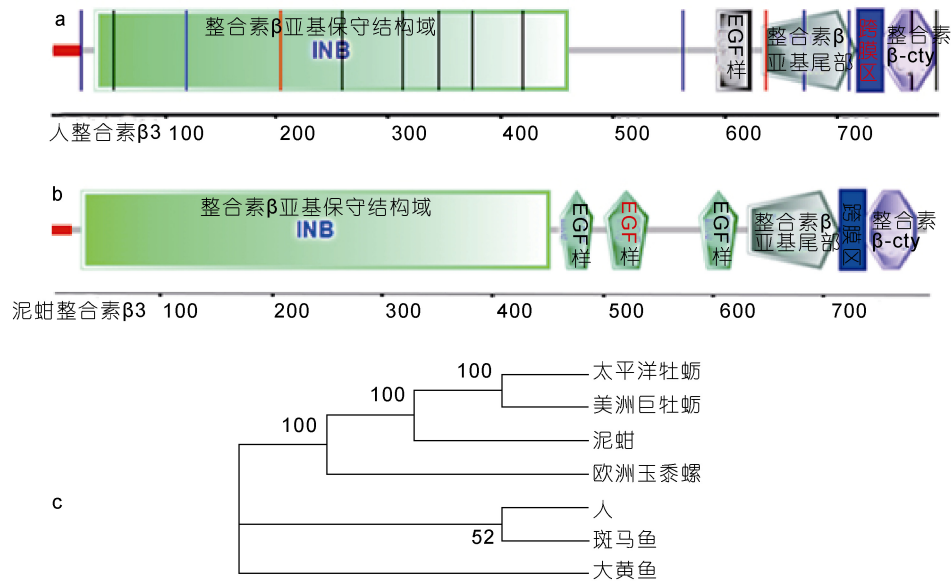


图 7 整合素  $\beta 3$  亚基的结构域和系统进化树

Fig.7 The conserved domain of ITB3 and phylogenetic relationships

注: a. 人整合素  $\beta 3$  亚基保守结构域; b. 泥蚶整合素  $\beta 3$  亚基保守结构域; c. 系统进化树

#### 2.4 热休克蛋白

泥蚶血细胞转录组中鉴定出 1 种新型热休克蛋白, 为热休克蛋白 9 (strss-70 protein, mitochondrial, HSPA9)。热休克蛋白 9 长度为 2189bp, 开放阅读框长度为 2067bp, 编码 688 个氨基酸, 编码蛋白分子量大小约为 75.6kDa、理论等电点 6.32; 谷氨酸含量最高, 占 8.4%; 不稳定系数为 41.96, 属于稳定不蛋白; 具有  $\alpha$  螺旋、 $\beta$  折叠及无规则卷曲组成的二级结构; 三级结构分析显示其与模板 2kho.1.A、1yuw.1.A 和 4f19.1.A 同源性超过 55%, 以同源性高的模板 2kho.1.A 生成三级结构见图 8a。NCBI 多重序列比对结果显示泥蚶热休克蛋白 9 氨基酸序列与哺乳类(人 *Homo sapiens*, 小鼠 *Mus musculus*)、斑马鱼、太平洋牡蛎和虾夷扇贝等序列相似度都在 70%以上。系统进化树表明: 泥蚶 HSPA9 与虾夷扇贝优先聚为一支, 然后与太平洋牡蛎和美洲巨牡蛎相聚, 表明它们之

间亲缘关系比较近(图 8b)。

#### 2.5 Toll 相互作用蛋白

Toll 相互作用蛋白(TOLLIP)长度为 995bp, 开放阅读框为 867bp, 编码 289 个氨基酸, 分子量约为 32.4kDa、理论等电点 4.77; 含量最高的氨基酸为谷氨酰胺, 占 10.03%; 不稳定系数为 57.98, 属于稳定不蛋白; 二级结构由  $\alpha$  螺旋、 $\beta$  折叠及无规则卷曲组成,  $\alpha$  螺旋占比最低。第 68—165 位置上具有 C2(蛋白激酶 C 保守结构域 2)结构域, C 端第 245—287 具有 toll 相互作用蛋白 CUE 保守结构域(图 9b)。Swiss-model 数据库中暂时没有 toll 相互作用蛋白模板。多重序列比对显示其与太平洋牡蛎和虾夷扇贝相似度高达 60%以上, 与人类和斑马鱼等相似度超过 50%。系统进化分析显示泥蚶 toll 相互作用蛋白与太平洋牡蛎、美洲巨牡蛎和虾夷扇贝亲缘关系比较近(图 9c)。

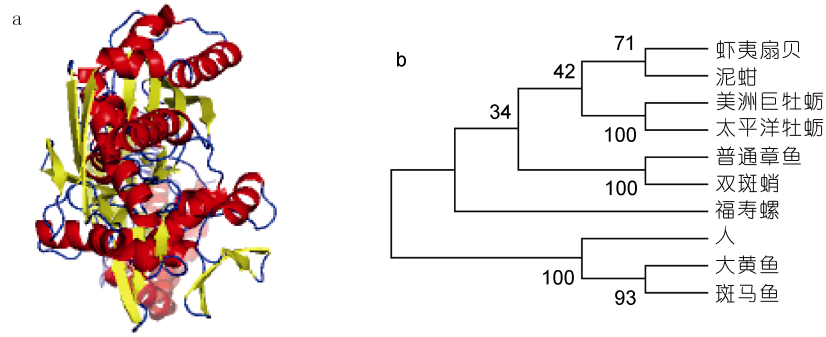


图 8 热休克蛋白 9 三级结构和系统进化树

Fig.8 The tertiary structure and phylogenetic relationships of HSPA9  
注: a. 三级结构; b. 系统进化树

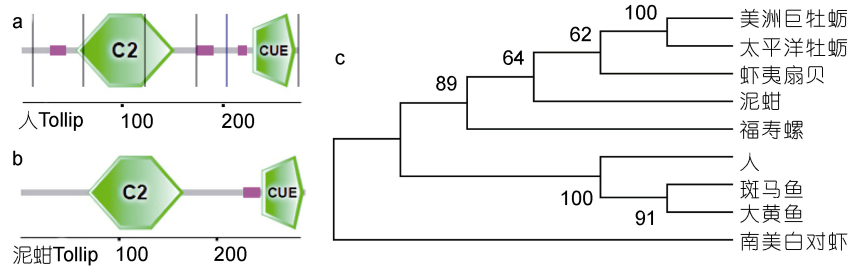


图 9 Toll 相互作用蛋白的结构域和系统进化树

Fig.9 The conserved domain of TOLLIP and phylogenetic relationships  
注: a. 人 Toll 相互作用保守结构域; b. 泥蚶 Toll 相互作用蛋白保守结构域; c. 系统进化树

2.6 感染前后 7 种免疫相关基因表达差异

基因丰度反映了基因表达水平，鳗弧菌刺激后，7 种免疫相关基因丰度 TPM 见图 10，整合素  $\beta 3$  以及 Toll 相互作用蛋白刺激前后丰度变化不大，四种溶酶体酶和热休克蛋白 9 刺激后表达丰度显著增加，以  $P < 0.05$  且  $|\log_2 \text{Fold Change}| > 1$  进行差异基因表达分析，四种溶酶体酶和热休克蛋白 9 表达上调。

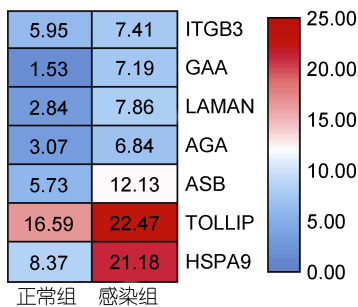


图 10 感染前后 7 种免疫相关基因表达量

Fig.10 The differential expression of seven immune related genes

3 讨论

3.1 免疫相关基因的进化特点

本文采用转录组测序获得了泥蚶 7 个免疫相关基因序列，并分析了其一级、二级和三级结构，采用 Swiss-model 进行同源建模，一般要求序列同源性大于 30% (Arnold *et al*, 2006; Marco *et al*, 2014)，本研究中 6 个蛋白序列与 Swiss-model 数据库中的模板同源性均超过 35%，表明三级结构可靠。利用最大似然法 (ML 法) 并采用步长循环 1000 进行进化树检验，构建了 7 种免疫相关基因的系统进化树，系统进化树显示泥蚶与虾夷扇贝 *Mizuhopecten yessoensis*、太平洋牡蛎 *Crassostrea gigas* 和美洲巨牡蛎 *Crassostrea virginica* 亲缘关系比较近，四者常聚为一个紧密的簇；福寿螺 *Pomacea canaliculata* 与泥蚶亲缘关系次于三种双壳贝类；斑马鱼 *Danio rerio*、大黄鱼 *Larimichthys crocea* 和人 *Homo sapiens* 与泥蚶亲缘关系较远。与多数双壳贝类不同，泥蚶血液呈红色，以血红蛋白作为呼吸色素，系统进化关系显示免疫相关基因与无脊

椎动物同源性更高些,这与血细胞形态学研究结果相一致,朱泽闻等(2011)通过研究发现泥蚶血细胞由红色颗粒细胞、嗜碱性颗粒细胞和透明细胞为主,血细胞组成上不同于脊椎动物。

### 3.2 溶酶体酶

泥蚶属低等无脊椎动物,血细胞在免疫防御中起重要作用,血细胞直接参与异物的免疫黏附、包裹、吞噬等过程,细胞内溶酶体酶的水解作用是杀伤和清除吞噬病原微生物主要途径,部分溶酶体酶还能释放到血淋巴中,通过改变病原菌表面的分子结构,有利于吞噬细胞识别病原微生物(Cheng, 1983)。Carballal 等用半定量光密度法(API ZYM 试剂盒)测定了贻贝血细胞中与免疫防御功能相关的酶类,贻贝血细胞具有  $\alpha$ -葡萄糖苷酶活性(Carballal *et al.*, 1997); Xue 等用 API ZYM 试剂盒在健康欧洲扁牡蛎血细胞中发现 14 种酶,没有发现  $\alpha$ -葡萄糖苷酶活性,原生动物的 *Bonamia ostreae* 感染后血细胞具有  $\alpha$ -葡萄糖苷酶活性(Xue *et al.*, 2000)。通过转录组测序,健康泥蚶血细胞中有  $\alpha$ -葡萄糖苷酶表达,细菌感染后泥蚶血细胞中  $\alpha$ -葡萄糖苷酶表达量显著增加。生理状态下,血细胞中  $\alpha$ -葡萄糖苷酶是否具有活性,不同贝类研究结果不一样,表明其存在形式具有种属特异性,欧洲扁牡蛎血细胞中  $\alpha$ -葡萄糖苷酶可能以无活性前体形式存在,而贻贝是以活性形式存在,泥蚶  $\alpha$ -葡萄糖苷酶分子特征显示其没有信号肽,长牡蛎和斑马鱼溶酶体  $\alpha$ -葡萄糖苷酶二级结构分析显示其没有信号肽,而人类  $\alpha$ -葡萄糖苷酶前体 N 端含有信号肽,暗示泥蚶血细胞中溶酶体  $\alpha$ -葡萄糖苷酶以活性形式存在。

溶酶体  $\alpha$ -甘露糖苷酶也属于糖苷酶,参与甘露糖苷键水解,贻贝、欧洲扁牡蛎、太平洋牡蛎血细胞中均发现  $\alpha$ -甘露糖苷酶,鳎弧菌感染后泥蚶血细胞中溶酶体  $\alpha$ -甘露糖苷酶表达量增加,与太平洋牡蛎研究结果相似(Xue *et al.*, 2000)。泥蚶溶酶体  $\alpha$ -甘露糖苷酶其序列 N 端具有信号肽,人类和斑马鱼溶酶体  $\alpha$ -甘露糖苷酶前体二级结构显示其 N 端具有信号肽(Paciotti *et al.*, 2017),表明泥蚶血细胞中溶酶体  $\alpha$ -甘露糖苷酶以前体的形式存在。

芳基硫酸酯酶的作用是水解体内的硫酸盐,芳基硫酸酯酶 B 主要水解体内的糖胺聚糖(gags),特别是水解硫酸皮肤素和硫酸软骨素,二级结构分析显示其与人芳基硫酸酯酶前体一样, N 端具有信号肽,以前体存在。天冬酰胺氨基葡萄糖酶能水解与天冬酰

胺残基相连的 N-糖苷键(GlcNAc-Asn),二级结构显示其没有信号肽,表明其以活性形式存在溶酶体中。

鳎弧菌感染后,泥蚶血细胞中四种酶表达量均增加,表明其在水解细菌过程中发挥作用,溶酶体酶水解作用是血细胞杀伤和清除细菌的重要途径,与泥蚶血细胞超微结构研究相一致(Zhou *et al.*, 2019)。

### 3.3 整合素

整合素是一种跨膜蛋白,由  $\alpha$  和  $\beta$  两个亚基通过非共价结合形成二聚体,介导细胞与细胞以及细胞与细胞外基质相互识别和粘附(马进, 2004)。在人类研究中发现,整合素  $\beta 3$  有两种存在形式:与  $\alpha I I b$  形成二聚体以及与  $\alpha V$  形成二聚体,  $\alpha I I b \beta 3$  主要表达于巨核细胞和血小板上,  $\alpha V \beta 3$  存在于多种细胞上(姜泓等, 2007)。泥蚶没有巨核细胞和血小板,  $\beta 3$  应以  $\alpha V \beta 3$  形式存在于血细胞上。整合素  $\beta 3$  的天然配体都含有一段精氨酸-甘氨酸-天冬氨酸(RGD)序列,包括纤连蛋白、玻连蛋白和层粘连蛋白等配体,整合素  $\beta 3$  与配体识别后会引起细胞内信号转导过程,使下游激酶发生磷酸化修饰,如丝裂原活化蛋白激酶(Mitogen-activated protein kinases, MAPK)、磷酸肌醇激酶(Phosphoinositide 3-kinase, PI3K)和细胞外信号调节激酶(Extracellular signal regulated kinase, ERK)磷酸化,从而发挥生理学功能,如细胞增殖、细胞迁移等(Sun *et al.*, 2016; Zoppi *et al.*, 2018)。转录组测序获得的泥蚶整合素  $\beta 3$  序列具有典型的整合素  $\beta$  结构域,包括胞外区域、跨膜区域和胞内区域(Xiong *et al.*, 2001),泥蚶整合素  $\beta 3$  具有 3 个 EGF 样结构域,与人类整合素 3 保守结构域稍有差异。鳎弧菌刺激后,血细胞中整合素  $\beta 3$  表达没有明显变化,暗示泥蚶整合素  $\beta 3$  不直接参与鳎弧菌识别与吞噬,而 Lv 等(2019)研究发现长牡蛎血细胞中整合素  $\beta 5$  参与结合和吞噬大肠埃希杆菌。

### 3.4 热休克蛋白和 Toll 相互作用蛋白

HSPA9 是热休克蛋白 70 家族成员之一, HSPA9 存在于线粒体,是线粒体主要蛋白质之一,与细胞多种生理活动密切相关,如负责精确调控线粒体蛋白的输入和输出、线粒体产能、分子伴侣以及应激反应等(张国彬等, 2009)。鳎弧菌感染后,泥蚶血细胞中 HSPA9 表达显著增高,表明其参与机体应激反应,与斑点叉尾鲷研究结果相一致(Song *et al.*, 2016)。

Toll 相互作用蛋白是一种接头蛋白,在天然免疫中发挥重要作用,其是 Toll 样受体信号通路的组成部分。未受刺激的泥蚶血细胞中, Toll 相互作用蛋白表



大量较高,表明其也是一种负调控因子, Tollip 能下调 MyD88 依赖性的 TLRs 信号通路转导从而抑制细胞活化和炎症反应,使 NF- $\kappa$ B 的转录活性降低,以避免免疫反应对机体本身的损伤(陈维英等, 2011)。受鳃弧菌刺激 24h 后, 泥蚶血细胞中 Tollip 表达量变化不显著, 鳃弧菌刺激后虾夷扇贝血细胞中 Tollip 表达量 24h 时显著上升(李若佼, 2015), 而日本鳃弧菌感染嗜水气单胞菌 24h 后肝脏、脾脏和肾中 Tollip 表达量没有明显变化(Feng *et al*, 2019), 表明 Tollip 响应细菌感染的表达模式有种属差异。

#### 4 结论

本研究通过对泥蚶血细胞进行 RNA-Seq 转录组测序, 对测序序列进行基因信息注释和生物信息学分析, 获得了 7 个免疫相关基因的蛋白序列, 包括溶酶体酶、整合素、热休克蛋白和 Toll 相互作用蛋白, 分析了 7 个基因蛋白序列理化性质、保守结构域、二级结构和三级结构, 鳃弧菌感染后 4 种溶酶体酶和热休克蛋白 9 表达量显著升高, 获得地这些基因序列有利于更深入研究泥蚶的抗病机理。

#### 参 考 文 献

- 马 进, 2004. 整合素的研究进展. 上海第二医科大学学报, 24(1): 11—13
- 朱泽闻, 徐立蒲, 吴信忠等, 2011. 泥蚶血细胞的形态结构特征及部分免疫功能. 水产学报, 35(10): 1494—1504
- 许家辉, 陈世雄, 俞兴发等, 2018. 投喂海水小球藻和三角褐指藻对泥蚶成贝铜排出的影响. 海洋与湖沼, 49(3): 624—629
- 孙敬锋, 吴信忠, 2006. 贝类血细胞及其免疫功能研究进展. 水生生物学报, 30(5): 601—607
- 李太武, 李成华, 宋林生等, 2003. 5 个泥蚶群体遗传多样性的 RAPD 分析. 生物多样性, 11(2): 118—124
- 李若佼, 2015. 虾夷扇贝 NF- $\kappa$ B 家族基因和 Tollip 基因鉴定及表达分析. 青岛: 中国海洋大学硕士学位论文, 59—62
- 张国彬, 张 宇, 孙秀红, 2009. 热休克蛋白 70 的研究进展. 中国综合临床, 25(8): 892—894
- 陈雪峰, 王春琳, 顾志敏等, 2019. 罗氏沼虾(*Macrobrachium rosenbergii*)卵巢发育不同时期转录组分析. 海洋与湖沼, 50(2): 398—408
- 陈维英, 李龙江, 汤为学, 2011. Tollip 的研究进展. 中华医学教育探索杂志, 10(8): 1020—1022
- 姜 泓, 白雪帆, 2007. 整合素  $\alpha$ v $\beta$ 3 与病毒感染. 病毒学报, 23(6): 494—497
- 董迎辉, 项 翔, 姚韩韩等, 2013. 泥蚶(*Tegillarca granosa*)生长因子受体结合蛋白 2(GRB2)基因的克隆与表达分析. 海洋与湖沼, 44(4): 937—943
- Arnold K, Bordoli L, Kopp J *et al*, 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 22(2): 195—201
- Carballal M J, López C, Azevedo C *et al*, 1977. Enzymes involved in defense functions of hemocytes of mussel *Mytilus galloprovincialis*. *Journal of Invertebrate Pathology*, 70(2): 96—105
- Cheng T C, 1983. The role of lysosomes in molluscan inflammation. *American Zoologist*, 23(1): 129—144
- Feng J J, Lin P, Wang Y L *et al*, 2019. Molecular characterization, expression patterns, and functional analysis of toll-interacting protein (Tollip) in Japanese eel *Anguilla japonica*. *Fish & Shellfish Immunology*, 90: 52—64
- Liu G S, Chen M L, Yu C *et al*, 2017. Molecular cloning, characterization and functional analysis of a putative mitogen-activated protein kinase kinase kinase 4 (MEKK4) from blood clam *Tegillarca granosa*. *Fish & Shellfish Immunology*, 66: 372—381
- Lv Z, Qiu L M, Jia Z H *et al*, 2019. The activated  $\beta$ -integrin (Cg $\beta$ V) enhances RGD-binding and phagocytic capabilities of hemocytes in *Crassostrea gigas*. *Fish & Shellfish Immunology*, 87: 638—649
- Marco B, Stefan B, Waterhouse A *et al*, 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, 42(W1): W252—W258
- Paciotti S, Codini M, Tasegian A *et al*, 2017. Lysosomal alpha-mannosidase and alpha-mannosidosis. *Frontiers in Bioscience*, 22(1): 157—167
- Patro R, Duggal G, Love M I *et al*, 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14(4): 417—419
- Song L, Li C, Xie Y J *et al*, 2016. Genome-wide identification of Hsp70 genes in channel catfish and their regulated expression after bacterial infection. *Fish & Shellfish Immunology*, 49: 154—162
- Sun C C, Qu X J, Gao Z H, 2016. Arginine-glycine-aspartate-binding integrins as therapeutic and diagnostic targets. *American Journal of Therapeutics*, 23(1), e198—e207
- Wang Q, Li J, Guo H M, 2019. Transcriptome analysis and discovery of genes involved in immune pathways in *Solen strictus* (Gould, 1861) under *Vibrio anguillarum*. *Fish & Shellfish Immunology*, 88: 237—243
- Xiong J P, Stehle T, Diefenbach B *et al*, 2001. Crystal structure of the extracellular segment of integrin  $\alpha$ V $\beta$ 3. *Science*, 294(5541): 339—345
- Xue Q G, Renault T, 2000. Enzymatic activities in European flat oyster, *Ostrea edulis*, and pacific oyster, *Crassostrea gigas*, hemolymph. *Journal of Invertebrate Pathology*, 76(3): 155—163
- Zhou L Q, Zhao D, Wu B *et al*, 2019. Ark shell *Scapharca broughtonii* hemocyte response against *Vibrio anguillarum* challenge. *Fish & Shellfish Immunology*, 84: 304—311
- Zoppi N, Chiarelli N, Ritelli M *et al*, 2018. Multifaced roles of the  $\alpha$ v $\beta$ 3 integrin in Ehlers-Danlos and arterial tortuosity syndromes' dermal fibroblasts. *International Journal of Molecular Sciences*, 19(4): 982

## ANALYSIS OF THE STRUCTURE AND EXPRESSION OF IMMUNE-RELATED GENES IN BLOODY CLAM *TEGILLARCA GRANOSA* BASED ON HAEMOCYTE TRANSCRIPTOME SEQUENCING

ZHONG Ai-Hua, CHU Zhang-Jie, WANG Wei-Hong, MOU Yi  
(School of Fishery, Zhejiang Ocean University, Zhoushan 316022, China)

**Abstract** The haemocytes of shellfish are directly involved in the phagocytosis and encapsulation of foreign materials. They also play an important role in the process of wound repair and inflammatory response. They are the undertaker of immune defense. RNA sequencing technology was used to study the transcriptome of haemocytes from bloody clam *Tegillarca granosa*. By sequence assembly, the sequences of  $\alpha$ -glucosidase (GAA),  $\alpha$ -mannosidase (LAMAN), arylsulfatase B (ASB), asparagine glucosaminase (AGA), integrin  $\beta$ 3 subunit (ITB3), heat shock protein 9 (HSPA9) and toll interaction protein (TOLLIP) were obtained. The encoded proteins were obtained by Orfinder. The physical and chemical properties, secondary structure, conserved domain, and tertiary structure were analyzed. The results showed that the secondary structures of the seven proteins contained  $\alpha$ -helix,  $\beta$ -fold, and irregular curl. The ITB3 and TOLLIP conserved domains of clam and human are similar. The homology of the tertiary structure and template from Swiss-Model was more than 35% (except TOLLIP). After stimulation by *Vibrio anguillarum*, the expression of GAA, LAMAN, ASB, AGA, and HSPA9 increased significantly, while the expression of ITB3 and Tollip did not change significantly. This study complemented and improved the understanding of immune related genes, and provided important information for further research on the disease resistance mechanism of *T. granosa*.

**Key words** bloody clam *Tegillarca granosa*; haemocyte; RNA sequencing; immune-related gene