

马氏珠母贝(*Pinctada fucata*)血细胞转录组测序 数据中 SNP 标记的开发及其功能注释分析*

王忠良^{1,2} 丁 燊¹ 许尤厚³ 王 蓓¹ 张健东^{1,2①}

(1. 广东海洋大学水产学院 湛江 524088; 2. 南海水产经济动物增养殖广东省普通高校重点实验室 湛江 524088;
3. 广西北部湾海洋生物多样性养护重点实验室(钦州学院) 钦州 535000)

摘要 本文基于马氏珠母贝(*Pinctada fucata*)血细胞比较转录组学数据进行单核苷酸多态性标记(SNP)的开发和关联基因的功能注释分析,以期 SNP 标记的利用提供有效信息。在转录组测序获得的 67632 条 SNP-unigenes 中共获得 962995 个 SNP 位点,位点平均发生频率约为 1 个 SNP/100bp;转换 SNP 数目与颠换 SNP 数目比值约为 0.5,与理论比率相符;而 6 种单核苷酸变异类型中,以 A/T 颠换 SNP 最多。SNP-unigene 功能注释分析表明,30376 条 unigenes (44.91%)匹配到 GO 分类条目;18150 条 unigenes (26.84%)获得 COG 注释信息,并以“一般功能预测”类最多;10981 条 unigenes (16.24%)富集到 32 条 KEGG 子集,其中以“信号转导”子集中富集的 SNP-unigene 最多。进一步富集分析显示 629 个 SNP-unigenes 注释到“Platelet activation”等多条免疫防御相关信号通路中;同时,根据 SNP 位点分布情况筛选到 123 个溶藻弧菌感染前、后转录组中特异分布的免疫防御相关候选 SNP 位点。基于标准化的基因表达水平分析,在 17 个差异表达免疫防御相关基因中共检测到 1594 个 SNP 位点。

关键词 马氏珠母贝;单核苷酸多态性标记;转录组;分子标记

中图分类号 Q789 **doi:** 10.11693/hyhz201711000264

马氏珠母贝(*Pinctada fucata*)是人工培育海水珍珠的主要母贝,超过 90%的海水珍珠产量来自于马氏珠母贝,在我国海水养殖业中占有重要地位(Zhan *et al.*, 2016)。然而,由于长期的人工养殖和近亲繁殖,马氏珠母贝已呈现出种苗活力下降、生长速度减慢、死亡率增加、育珠贝变小、珍珠质量差等明显的种质退化现象,严重制约了我国海水珍珠产业的健康持续发展(Wang *et al.*, 2009)。因此,马氏珠母贝的遗传改良和优良品种培育是当前提高珍珠贝养殖效益和健康养殖水平的首要任务。

规模化开发马氏珠母贝分子标记是其种质资源发掘、经济性状关联分析、分子遗传育种技术建立和

优良品种培育的基础性工作之一。单核苷酸多态性标记(Single Nucleotide Polymorphism, SNP)是指单个核苷酸发生改变而引起的 DNA 序列多态性,最早由美国学者 Lander 于 1996 年提出,是继扩增片段长度多态性(AFLP)、微卫星(SSR)等之后的新一代分子标记(Lander, 1996)。由于具有数量多、密度大、遗传稳定性高、便于高通量自动化检测分析等优点,SNP 标记被广泛应用于高密度遗传连锁图谱构建、基因定位、品种鉴定、群体遗传学研究和分子辅助育种等领域(Liu *et al.*, 2004; Mir *et al.*, 2013; 张晓萌等, 2013)。

近年来,水产动物中 SNP 标记的开发和应用已有若干报道,如鱼类的大菱鲆(*Scophthalmus maximus*)

* 广东省省级科技计划项目, 2015A020209169 号; 广东海洋大学优秀青年骨干教师特别资助计划项目, GDOU2015050214 号; 国家自然科学基金项目, 31202023 号; 广西北部湾海洋生物多样性养护重点实验室(钦州学院)开放课题, 2017KB03 号。王忠良, 博士, 副教授, E-mail: leong2006@126.com

通讯作者: 张健东, 副教授, E-mail: yzxzjd@126.com

收稿日期: 2017-10-16, 收修改稿日期: 2017-11-22

(王婷等, 2014)、斑点叉尾鲷(*Ictalurus punctatus*) (Li *et al*, 2014)、大西洋鲑(*Salmo salar*) (Houston *et al*, 2014; Tsai *et al*, 2016)、甲壳类的凡纳滨对虾(*Litopenaeus vannamei*) (Yu *et al*, 2014)、斑节对虾(*Penaeus monodon*) (Sellars *et al*, 2014)、中华绒螯蟹(*Eriocheir sinensis*) (Cui *et al*, 2013)、三疣梭子蟹(*Portunus trituberculatus*) (张德宁等, 2015)、贝类的栉孔扇贝(*Chlamys farreri*) (李纪勤等, 2013)、长牡蛎(*Crassostrea gigas*) (Wang *et al*, 2017)。马氏珠母贝中也开发了少量经济性状相关 SNP 标记, 但数量还未达到高密度遗传连锁图谱构建和 QTL 精细定位的要求(Zhan *et al*, 2016)。

目前, SNP 标记的大规模开发主要依靠基因组测序和序列的拼接比对, 但这种方法费用高、适用范围有限, 限制了 SNP 标记在缺乏全基因组信息的非模式生物中的开发和利用。随着高通量测序技术的快速发展, 转录组测序成本大幅下降, 比较转录组学方法筛选性状相关基因已成为研究特定机理的普遍方法, 同时通过序列比对可识别大量的 SNP 位点(Pante *et al*, 2012; Núñez-Acuña *et al*, 2013; Cui *et al*, 2014; Yu *et al*, 2014; Mastrochirico-Filho *et al*, 2016; Li *et al*, 2017)。本文基于前期马氏珠母贝血细胞转录组测序数据进行 SNP 标记的开发和关联基因的功能注释分析, 以期马氏珠母贝遗传多样性分析、遗传图谱构建和分子辅助育种研究等提供候选标记资源, 促进马氏珠母贝种质资源保护、优良品种培育和珍珠养殖业的健康发展。

1 材料与方法

1.1 马氏珠母贝及高通量转录组测序数据

马氏珠母贝(*Pinctada fucata*) (平均壳长 70mm)购自广东省湛江市雷州后洪珍珠贝养殖场; 暂养于室内玻璃钢水槽中(80L 水体), 每槽饲养 20 只; 水温 $25\pm 1^{\circ}\text{C}$, 盐度 28, 饲养期间连续充气; 暂养过程中以螺旋藻粉为饵料投喂, 每天换水(100%)一次。室内暂养一周后将珠母贝分为溶藻弧菌感染组(AI)和对照组(BI); AI 组珠母贝闭壳肌注射 0.1mL、浓度为 5×10^7 CFU/mL 的溶藻弧菌悬液, BI 组注射同体积的 PBS; 细菌感染 4h 后闭壳肌采集血淋巴, $800\times g$ 、 4°C 离心 10min 收集血细胞; TRIzol 法提取总 RNA, 经浓度和纯度检测后用于 Illumina/Hiseq-2000 高通量转录组测序(Wang *et al*, 2016)。

转录组测序共产生 70407878 条 Raw Reads, 经去

除含有接头、重复及测序质量较低的原始读数后获得 56345139 条 Clean Reads; 使用 *de novo* 组装软件 Trinity (Grabherr *et al*, 2011)对 Clean Reads 进行组装, 并进行去冗余处理和进一步拼接, 共得到 74007 条 unigenes (Raw Reads 数据的 SRA 登录号为 SRP041567) (Wang *et al*, 2016)。

1.2 SNP 位点的检测

首先应用 SOAP 软件将各样本测序所得 Clean Reads 与 Trinity 拼接的转录本进行比对(错配 ≤ 2); 合并单端和双端比对结果, 过滤掉 Duplicated Reads 和 Multi-mapped Reads, 并将比对结果按转录本和坐标位置进行排序; 应用 SOAPsnp 软件(Li *et al*, 2009)对排序后的文件进行 SNP 位点检测, 将得到的 SNP 结果; 按质量值(≥ 20)、测序深度(≥ 10 和 ≤ 100)和 SNP 间距(≥ 5)等条件进行过滤并去杂合, 最终获得纯合 SNP 位点数据(Du *et al*, 2014; Lv *et al*, 2014)。

1.3 SNP 位点所在 unigene 的注释与功能分析

应用 BLAST 程序将含有 SNP 的 unigene (SNP-unigene)序列比对到 nr、nt、Swiss-Prot 和 COG(直系同源簇, Cluster of Orthologous Groups)等数据库, 从而获得该 unigene 的注释信息($E\text{-value} < 1.0\text{E}-5$); 采用 Blast2GO 软件获得 unigene 的 GO(基因本体论, Gene ontology)注释, 并使用 WEGO 软件对所有 unigene 做 GO 功能分类统计; 通过 BLASTx 比对 SNP-unigene 到 COG 数据库和 KEGG(京都基因与基因组百科, Kyoto Encyclopedia of Genes and Genomes)数据库, 分别获得 COG 分类注释和 KEGG 信号通路注释。

2 结果与分析

2.1 转录组数据及 SNP 位点数据分析

本研究的数据来源于对照组转录组(BI)和溶藻弧菌感染组转录组(AI)。经 SOAPsnp 软件检测, 67632 条 unigenes 中共获得 962995 个 SNP 位点(BI: 493243 个; AI: 469752 个)(表 1); BI 转录组和 AI 转录组的 SNP 发生频率分别为 0.0098 和 0.0093(约 1 个 SNP/100bp); SNP 密度频数分析显示, 每 1000 碱基含 0—5 个 SNP 位点的 unigene 最多(图 1)。所有 SNP 位点中, 纯合 SNP 位点 105232 个(BI: 50702 个; AI: 54530 个)(表 1); Unigene 中所含 SNP 位点数目统计表明, BI 转录组中含有 1 个 SNP 位点的 unigene 4597 条(13.64%), 含有 2—10 个 SNP 位点的 unigene 14354 条(42.60%), 含有 10 个以上 SNP 位点的 unigene

表 1 SNP 位点数量概况
Tab.1 Summary of the SNP identified in BI and AI transcriptomes

类别	BI 转录组	AI 转录组	总计
SNP-unigene	33697	33935	67632
SNP 位点总数	493243	469752	962995
纯合 SNP 位点	50702	54530	105232
颠换 SNP	33713	36327	70040
转换 SNP	16989	18203	35192

14746 条(43.76%); AI 转录组中含有 1 个 SNP 位点的 unigene 4632 条(13.65%), 含有 2—10 个 SNP 位点的 unigene 15140 条(44.61%), 含有 10 个以上 SNP 位点的 unigene 14163 条(41.74%)(图 2)。

SNP 类型分析表明, 来源于 BI 转录组的纯合 SNP 位点中, 颠换位点 33713 个(66.49%), 转换位点 16989

个(33.51%); 来源于 AI 转录组的纯合 SNP 位点中, 颠换位点 36327 个(66.62%), 转换位点 18203 个(33.38%)。BI 转录组和 AI 转录组转换 SNP 数目与颠换 SNP 数目比值分别为 0.504 和 0.501; 6 种单核苷酸变异类型中, 以 A/T 颠换 SNP 最多, 分别占纯合 SNP 总数的 25.45%(BI: 12906 个)和 25.72%(AI: 14024 个)(图 3)。

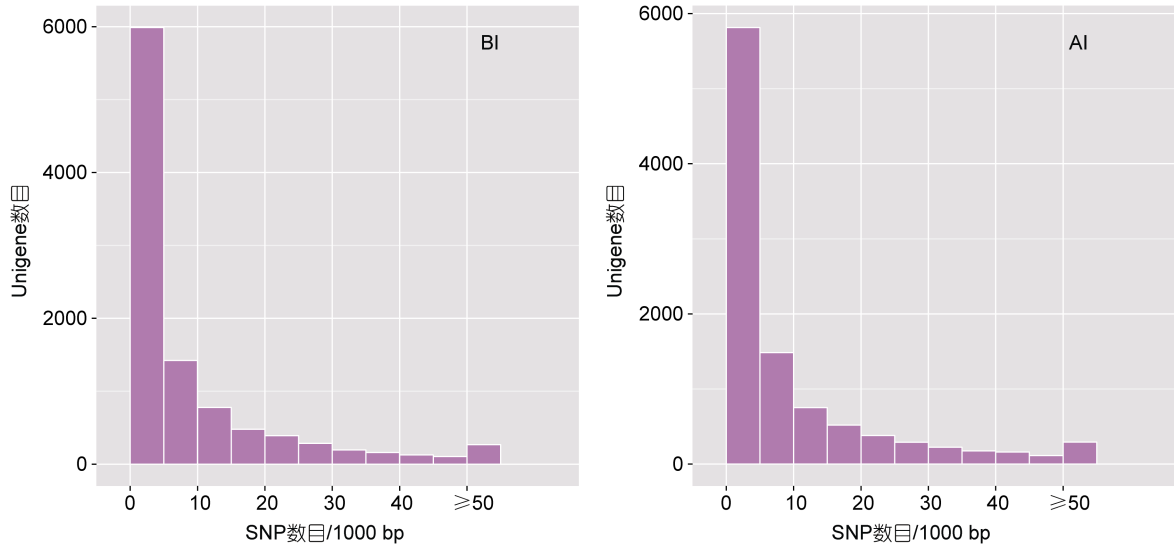


图 1 SNP 密度频数分布图
Fig.1 SNP frequency in BI and AI transcriptomes

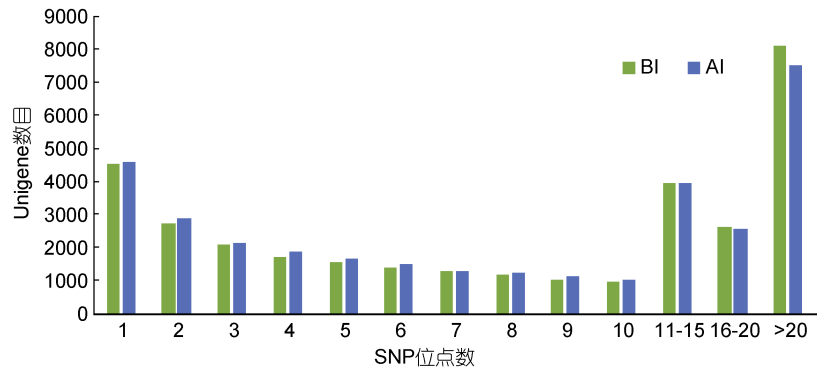


图 2 SNP 的分布统计
Fig.2 SNP distribution in BI and AI transcriptomes

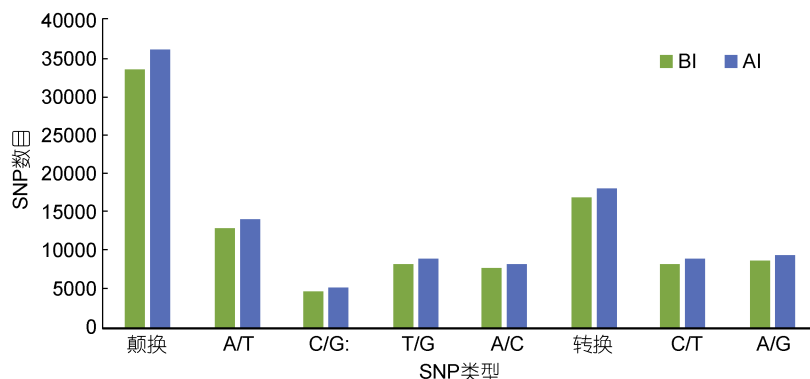


图3 SNP 类型分析结果

Fig.3 SNP numbers of different mutation types in BI and AI transcriptomes

2.2 SNP-unigene 的注释与功能分析

基于所有 67632 条 SNP-unigenes 序列与 Nr、Nt 及 Swiss-prot 数据库比对后获得的蛋白功能注释信息, 分别进行 GO、COG 基因功能分类及 KEGG 信号通路富集分析。GO 分析结果表明, 共有 30376 条 SNP-unigenes(44.91%)匹配到 GO 条目(GO term), 并被分配到生物过程、细胞组分及分子功能三大功能分类中的 42 个亚类(图 4); COG 结果显示, 共有 18150 条 SNP-unigenes(26.84%)能够匹配相应的 COG 注释信息; 根据功能信息可分为 26 类, 其中以“一般功能预测”类最多, 包含 3690 条 SNP-unigenes(图 5); KEGG 富集分析显示, 共有 10981 条 SNP-unigenes

(16.24%)富集到 32 条 KEGG 子集中, 其中以“信号转导”子集中富集的 SNP-Unigene 最多, 共计 1450 条(图 6)。

2.3 免疫防御相关 SNP-unigene 的富集及特异 SNP 位点分析

根据 KEGG 信号通路的富集分析, 共有 629 个 SNP-unigenes 注释到“Platelet activation”等 15 条与“病原感染”和“免疫功能”相关的一条或多条信号通路中, 其中以“Platelet activation”信号通路中注释的 unigene 最多(86 条), 其次分别为“Chemokine signaling pathway”、“Fc gamma R-mediated phagocytosis”、“Fc gamma R-mediated phagocytosis”和“Leukocyte transendothelial migration”等信号通路(表 2)。

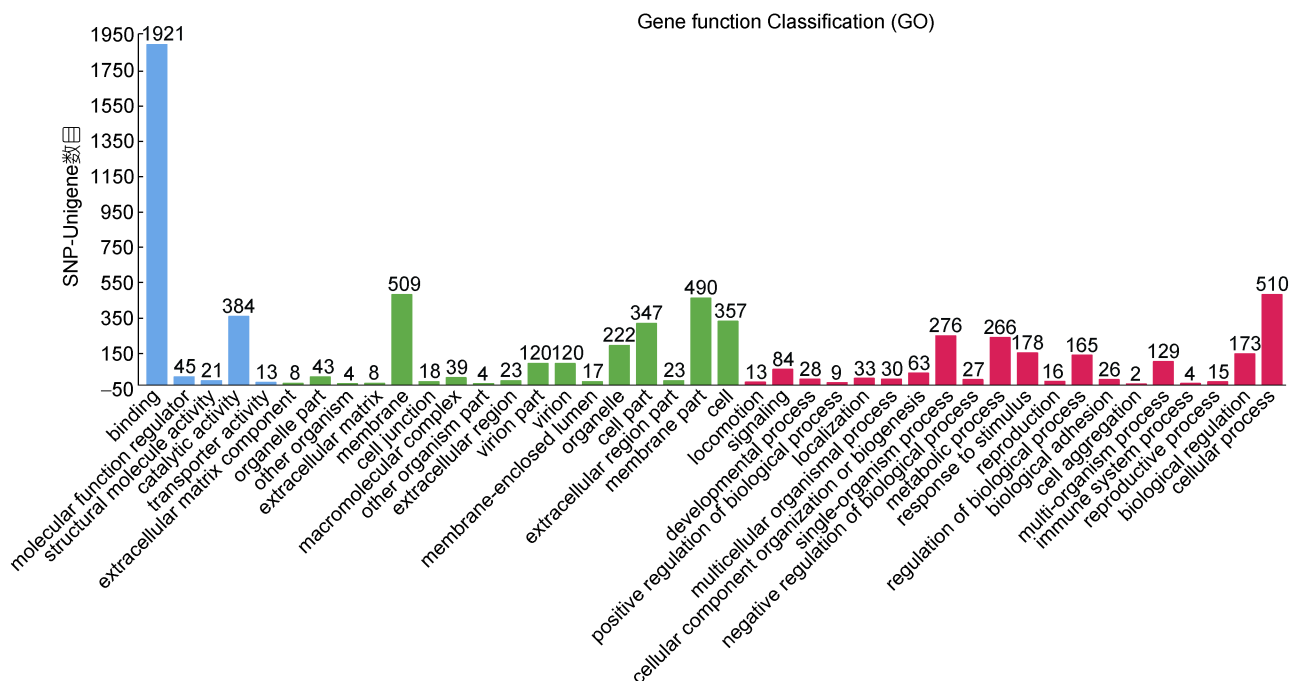


图4 SNP-unigenes GO 功能注释分析

Fig.4 Gene ontology (GO) classification of SNP-unigenes

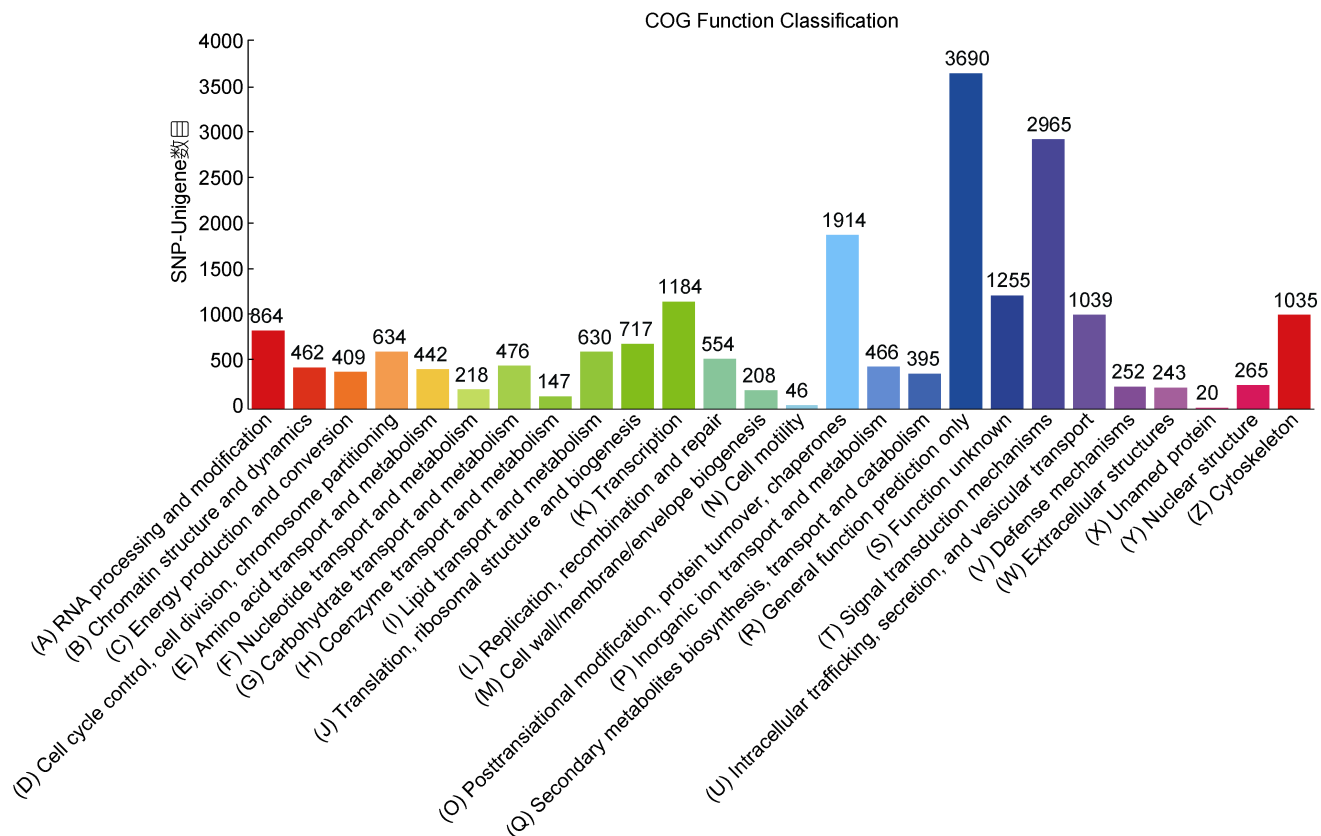


图 5 SNP-unigenes GOG 功能注释分析

Fig.5 The cluster of orthologous groups (COG) classification of SNP-unigenes

表 2 免疫防御相关 SNP-unigene 的 KEGG 富集分析

Tab.2 The Kyoto encyclopedia of genes and genomes (KEGG) enrichment analysis of immune-related SNP-unigenes

免疫防御相关 KEGG 信号通路	信号通路 ID	SNP-unigene 数目
Platelet activation	ko04611	86
Chemokine signaling pathway	ko04062	78
Fc gamma R-mediated phagocytosis	ko04666	69
Leukocyte transendothelial migration	ko04670	63
T cell receptor signaling pathway	ko04660	50
Toll-like receptor signaling pathway	ko04620	46
Fc epsilon RI signaling pathway	ko04664	39
B cell receptor signaling pathway	ko04662	35
NOD-like receptor signaling pathway	ko04621	35
Natural killer cell mediated cytotoxicity	ko04650	34
RIG-I-like receptor signaling pathway	ko04622	30
Cytosolic DNA-sensing pathway	ko04623	27
Antigen processing and presentation	ko04612	19
Complement and coagulation cascades	ko04610	9
Hematopoietic cell lineage	ko04640	9
合计		629

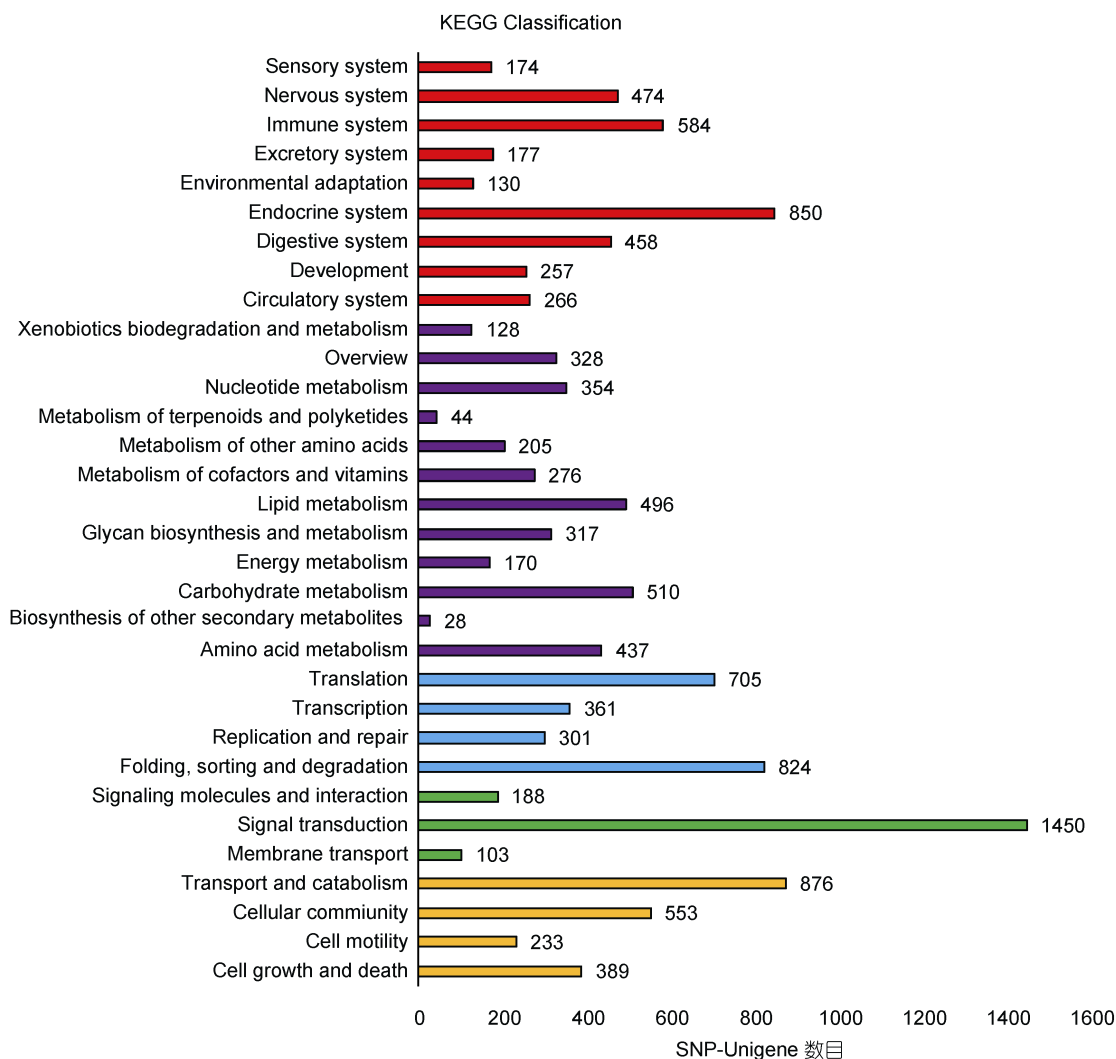


图 6 SNP-unigenes 的 KEGG 信号通路富集分析

Fig.6 The Kyoto encyclopedia of genes and genomes (KEGG) classification of SNP-unigenes

同时, 根据转录组中 unigene 的 SNP 位点分布情况及 KEGG 功能注释信息筛选 BI 和 AI 转录组中特异分布的免疫防御相关基因 SNP 位点, 共获得 123 个特异分布的 SNP 位点, 其中 BI 转录组中 36 个 SNP 位点, 存在于 12 个免疫防御相关基因中; AI 转录组中 87 个 SNP 位点, 存在于 22 个免疫防御相关基因中 (表 3)。

2.4 差异表达免疫防御相关基因 SNP 位点分析

基于标准化的基因表达水平分析, 在溶藻弧菌感染前、后的血细胞转录组中共检测到 636 个差异表达的 unigenes (Differentially Expressed Genes, DEGs) (Wang *et al.*, 2016)。根据 KEGG 信号通路分析结果及 unigene 的 SNP 位点检测信息, 获得了部分重要差异表达免疫防御相关基因 SNP 位点的分布情况。结果显示, 在 BI 和 AI 转录组中分别检测到 700 和 894 个

SNP 位点(共 1594 个 SNP 位点), 并分布于 17 个差异表达免疫防御相关基因中(表 4)。

3 讨论

在基因组中, SNP 标记相比于 SSR 等分子标记具有更高的稳定性, 且符合孟德尔遗传规律, 方便用于遗传研究分析中; 同时, SNP 标记的准确性和重现性也优于其它分子标记。因此, 作为第三代分子标记的 SNP 具有广阔的应用前景。然而, SNP 开发、检测的技术难点限制了 SNP 标记在非模式生物中的开发和应用。目前, SNP 标记的大规模开发主要是利用基因组测序及序列的拼接比对产生, 但这种途径成本较高(庄伟, 2010); 而对于没有基因组参考序列的非模式生物, 多利用公共数据库的 EST 序列进行筛选, 但 EST 序列与实际序列间可能存在差异, 所以这种方法

表 3 BI 及 AI 转录组特异分布免疫防御相关基因 SNP 位点分析
Tab.3 Specific immune-related SNPs identified in BI and AI transcriptomes

转录组	SNP-unigene ID	所属 KEGG 信号通路	SNP 数目	合计
BI	comp39780_c0	B cell receptor signaling pathway/ Toll-like receptor signaling pathway/ T cell receptor signaling pathway/ Fc epsilon RI signaling pathway/ Fc gamma R-mediated phagocytosis/ NOD-like receptor signaling pathway/ Platelet activation	7	36
	comp671474_c0	Complement and coagulation cascades/Platelet activation	3	
	comp290115_c0	Cytosolic DNA-sensing pathway	5	
	comp732503_c0	Cytosolic DNA-sensing pathway	4	
	comp51341_c0	Fc gamma R-mediated phagocytosis	1	
	comp57855_c0	Hematopoietic cell lineage	1	
	comp44879_c0	Leukocyte transendothelial migration/ Platelet activation	3	
	comp59941_c0	Leukocyte transendothelial migration/Platelet activation / T cell receptor signaling pathway	3	
	comp1007_c0	NOD-like receptor signaling pathway	3	
	comp653385_c0	NOD-like receptor signaling pathway	2	
	comp229504_c0	Platelet activation	2	
	comp738091_c0	RIG-I-like receptor signaling pathway	2	
AI	comp1306306_c0	Leukocyte transendothelial migration/Platelet activation	1	87
	comp14833_c0	Hematopoietic cell lineage	1	
	comp1620486_c0	Chemokine signaling pathway	4	
	comp19568_c0	Leukocyte transendothelial migration/Platelet activation	1	
	comp22997_c0	B cell receptor signaling pathway/Chemokine signaling pathway/Fc epsilon RI signaling pathway/Fc gamma R-mediated phagocytosis/Leukocyte transendothelial migration/Natural killer cell mediated cytotoxicity/Platelet activation/T cell receptor signaling pathway/Toll-like receptor signaling pathway	4	
	comp2349_c0	Complement and coagulation cascades/Platelet activation	2	
	comp340170_c0	Chemokine signaling pathway/Platelet activation	5	
	comp35583_c0	Cytosolic DNA-sensing pathway	3	
	comp364321_c0	Fc epsilon RI signaling pathway/Fc gamma R-mediated phagocytosis/Platelet activation	5	
	comp44949_c0	Fc gamma R-mediated phagocytosis	9	
	comp47012_c0	Fc gamma R-mediated phagocytosis	5	
	comp509876_c0	Hematopoietic cell lineage	3	
	comp51181_c0	Complement and coagulation cascades	8	
	comp524944_c0	Chemokine signaling pathway/Platelet activation	1	
	comp54383_c0	Leukocyte transendothelial migration/Platelet activation	1	
	comp55367_c0	Chemokine signaling pathway/Fc epsilon RI signaling pathway/Fc gamma R-mediated phagocytosis/Leukocyte transendothelial migration/Natural killer cell mediated cytotoxicity/Platelet activation/T cell receptor signaling pathway/Toll-like receptor signaling pathway/Platelet activation	2	
	comp57390_c0	Hematopoietic cell lineage/Platelet activation	7	
	comp579306_c0	Fc gamma R-mediated phagocytosis	6	
	comp60708_c0	Fc epsilon RI signaling pathway/Fc gamma R-mediated phagocytosis/Leukocyte transendothelial migration/T cell receptor signaling pathway	6	
	comp631969_c0	Hematopoietic cell lineage/Platelet activation	1	
	comp63931_c0	Cytosolic DNA-sensing pathway	5	
	comp748376_c0	Platelet activation	7	
	合计			123

表 4 差异表达免疫防御相关基因 SNP 位点分析
Tab.4 SNPs identified in differentially expressed immune-related unigenes

基因名称	免疫后表达水平	BI-SNP 数目	AI-SNP 数目
Alpha-L-iduronidase	上调	14	49
Apoptosis regulator R1	上调	51	95
Arrestin	上调	79	45
cAMP-responsive element modulator	上调	64	69
Cytosolic phospholipase A2	上调	41	37
Dual specificity protein phosphatase 10	上调	29	43
Dual specificity protein phosphatase 7	上调	49	46
Heat shock protein 90	上调	56	47
Interleukin-1 receptor	上调	53	60
MAP kinase-activated protein kinase 2	上调	75	77
myeloid differentiation primary response protein 88	上调	0	41
Nuclear transcription factor Y subunit alpha	上调	34	62
Putative inhibitor of apoptosis	上调	6	15
Serine/threonine protein kinase NLK	上调	20	68
Toll-like receptor 2	下调	39	37
Toll-like receptor 3	上调	38	45
Tumor necrosis factor	上调	52	58
合计		700	894

的工作量大、开发效率低，且覆盖不广(刘庆明, 2012)。随着高通量测序技术的发展，转录组测序很好地弥补了以上方法的不足和缺陷，其不仅成为研究生物免疫防御、生长发育、环境互作等复杂分子机制的重要手段，也是多态性 SNP 分子标记鉴定的重要数据来源(周华等, 2012)。

本研究基于 Hiseq-2000 测序技术从 67632 条 unigenes 中共获得 962995 个候选 SNP 位点，SNP 发生频率为 0.0098 和 0.0093(平均约 100bp 含有 1 个 SNP)，研究结果与女王扇贝(*Aequipecten opercularis*)和黑扇贝(*Mimachlamys varia*)SNP 分布频率相近(1 个 SNP/100bp) (Arias *et al*, 2009)，稍低于长牡蛎的 SNP 分布频率(1 个 SNP/40—60bp) (Sauvage *et al*, 2007)，但明显高于栉孔扇贝的 SNP 分布频率(1 个 SNP/426.5bp) (李纪勤等, 2013)。各物种 SNP 分布频率的差异可能与其生长环境的复杂程度及遗传背景的差异有关(张德宁等, 2015)，另一个可能的原因是在 EST 序列中 SNP 的筛选条件不同(李纪勤等, 2013)。SNP 碱基替换类型分为转换(由嘌呤置换嘌呤或嘧啶置换嘧啶)和颠换(由嘌呤置换嘧啶或由嘧啶置换嘌呤)两类，理论上发生转换的概率与发生颠换的概率比值应该为 0.5。本研究中，BI 转录组和 AI 转录组转换 SNP 数目与颠换 SNP 数目比值分别为 0.504 和 0.501，

与理论比率相符。然而，长牡蛎、三疣梭子蟹、中国对虾(*Fenneropenaeus chinensis*)等无脊椎动物 SNP 标记中转换的比例远大于颠换，转换/颠换比值大于 0.5(王家丰, 2013；张德宁等, 2015；逢锦菲, 2013)。这种差异被称为“转换偏差”(Gojobori *et al*, 1982; Li *et al*, 1984; Collins *et al*, 1994)。这种不同物种间转换/颠换比值的较大差异可能与不同生物在进化中承受的压力不同有一定关系(Zhao *et al*, 2002)。

利用转录组测序数据开发 SNP 标记具有一个极为显著的优点，即标记位点来自与功能基因直接相关的编码序列，便于开展进一步的基因功能及遗传性状解析研究(王婷等, 2014)。本研究获得的 67632 条 SNP-unigenes 中，分别有 30376 条、18150 条和 10981 条 SNP-unigenes 在 GO、COG 及 KEGG 数据库中获得功能注释，大部分 unigene 参与细胞代谢、生物合成、生物(非生物)胁迫(刺激)反应、信号转导、机体组成及激素分泌等过程，表明这些 SNP-unigene 广泛参与马氏珠母贝的各种生命活动，并影响其各种遗传性状的建立；同时，这些 SNP 标记的开发为将来的基因功能、基因表达调控模式等研究提供了跟踪标志。此外，来自于编码区的 SNP 因其对氨基酸编码的影响可能导致疾病发生或对环境毒物敏感度的改变，因此 SNP 也被用于疾病遗传学和抗病性状关联

性研究中。如刘敬文(2014)开展了凡纳滨对虾免疫相关基因 SNP 标记与 WSSV 抗性的关联分析, 最终确定了 681 个 SNP 组成的免疫基因候选标记库以指导抗性品种的选育; 逢锦菲(2013)利用开发的 SNP 候选位点, 以中国对虾的抗病群体和敏感群体为研究材料进行 SNP 基因型与抗病性状的相关性分析, 并鉴别出抗病相关 SNP 位点。为了进一步研究马氏珠母贝免疫防御机制, 筛选抗病相关 SNP 位点, 本文根据转录组 unigene 的注释信息, 共获得了 629 个含有 SNP 位点的免疫防御相关 unigenes, 并被注释到“Platelet activation”等多条重要的免疫防御信号通路中; 同时, 根据 KEGG 功能注释信息及 SNP 位点分布情况筛选到 123 个细菌感染前、后转录组中特异分布的免疫防御相关基因 SNP 位点。以上免疫防御相关 SNP-unigene 及特异分布 SNP 位点的获得为以分子标记应用为基础的马氏珠母贝抗病家系选育提供了有力辅助。

4 结论

本研究首次应用高通量转录组测序数据实现了马氏珠母贝 SNP 标记的高效率大规模开发, 共获得 962995 个 SNP 位点, 位点分布频率约为 1 个 SNP/100bp; 共有 629 个 SNP-unigenes 被注释到多种免疫防御相关信号通路中; 根据 SNP 位点分布情况, 筛选到 123 个特异分布的免疫防御相关候选 SNP 位点。以上研究结果对于丰富马氏珠母贝分子标记类型、阐明免疫防御分子机制、探究种群遗传多样性和保护种质资源均具有重要意义。

参 考 文 献

- 王 婷, 黄智慧, 马爱军等, 2014. 基于转录组数据的大菱鲂 (*Scophthalmus maximus*) SNP 标记开发及多态性分析. 海洋与湖沼, 45(6): 1300—1307
- 王家丰, 2013. 长牡蛎基因组 SNP 标记规模开发及其在遗传育种研究中的应用. 青岛: 中国科学院研究生院(海洋研究所)博士学位论文
- 刘庆明, 2012. 应用高分辨率溶解曲线开发大菱鲂 SNP 标记研究. 上海: 上海海洋大学硕士学位论文
- 刘敬文, 2014. 凡纳滨对虾免疫基因 SNPs 开发及其与 WSSV 抗性的关联分析. 青岛: 中国科学院研究生院(海洋研究所)硕士学位论文
- 庄 伟, 2010. 半滑舌鲷(*Cynoglossus semilaevis*)基因组 SNP 标记的开发与检测. 青岛: 中国海洋大学硕士学位论文
- 张晓萌, 马 普, 王洪迪等, 2013. SNPs 在水产动物中的研究进展. 生物技术通报, (8): 7—11
- 张德宁, 吕建建, 刘 萍等, 2015. 三疣梭子蟹生长相关 SNP 位点的鉴定. 中国水产科学, 22(3): 393—401
- 李纪勤, 包振民, 李 玲等, 2013. 栉孔扇贝 EST-SNP 标记开发及多态性分析. 中国海洋大学学报: 自然科学版, (1): 56—63
- 周 华, 张 新, 刘腾云等, 2012. 高通量转录组测序的数据分析与基因发掘. 江西科学, 30(5): 607—611
- 逢锦菲, 2013. “黄海 2 号”中国对虾高通量 SNP 筛选及其与抗 WSSV 性状的关联分析. 上海: 上海海洋大学硕士学位论文
- Arias A, Freire R, Boudry P *et al*, 2009. Single nucleotide polymorphism for population studies in the scallops *Aequipecten opercularis* and *Mimachlamys varia*. Conservation Genetics, 10(5): 1491—1495
- Collins D W, Jukes T H, 1994. Rates of transition and transversion in coding sequences since the human-rodent divergence. Genomics, 20(3): 386—396
- Cui J, Wang H D, Liu S K *et al*, 2014. Transcriptome analysis of the gill of *Takifugu rubripes* using Illumina sequencing for discovery of SNPs. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 10: 44—51
- Cui Z, Li X, Liu Y *et al*, 2013. Transcriptome profiling analysis on whole bodies of microbial challenged *Eriocheir sinensis* larvae for immune gene identification and SNP development. PLoS One, 8(12): e82156
- Du X, Li L, Zhang S *et al*, 2014. SNP identification by transcriptome sequencing and candidate gene-based association analysis for heat tolerance in the bay scallop *Argopecten irradians*. PLoS One, 9(8): e104960
- Gojobori T, Li W H, Graur D, 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. Journal of Molecular Evolution, 18(5): 360—369
- Grabherr M G, Haas B J, Yassour M *et al*, 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature Biotechnology, 29(7): 644—652
- Houston R D, Taggart J B, Cézard T *et al*, 2014. Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). BMC Genomics, 15(1): 90
- Lander E S, 1996. The new genomics: global views of biology. Science, 274(5287): 536—539
- Li R Q, Li Y R, Fang X D *et al*, 2009. SNP detection for massively parallel whole-genome resequencing. Genome Research, 19(6): 1124—1132
- Li S J, Liu H, Bai J J *et al*, 2017. Transcriptome assembly and identification of genes and SNPs associated with growth traits in largemouth bass (*Micropterus salmoides*). Genetica, 145(2): 175—187
- Li W H, Wu C I, Luo C C, 1984. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. Journal of Molecular Evolution, 21(1): 58—71
- Li Y, Liu S K, Qin Z K *et al*, 2014. Construction of a high-density, high-resolution genetic map and its integration with BAC-based physical map in channel catfish. DNA Research, 22(1): 39—52
- Liu Z J, Cordes J F, 2004. DNA marker technologies and their applications in aquaculture genetics. Aquaculture, 238(1-4):

- 1—37
- Lv J J, Liu P, Gao B Q *et al*, 2014. Transcriptome analysis of the *Portunus trituberculatus*: de novo assembly, growth-related gene identification and marker discovery. *PLoS One*, 9(4): e94055
- Mastrochirico-Filho V A, Hata M E, Sato L S *et al*, 2016. SNP discovery from liver transcriptome in the fish *Piaractus mesopotamicus*. *Conservation Genetics Resources*, 8(2): 109—114
- Mir R R, Hiremath P J, Riera-Lizarazu O *et al*, 2013. Evolving molecular marker technologies in plants: from RFLPs to GBS. In: Lübberstedt T, Varshney R eds. *Diagnostics in Plant Breeding*. Dordrecht: Springer, 229—247
- Núñez-Acuña G, Gallardo-Escárate C, 2013. Identification of immune-related SNPs in the transcriptome of *Mytilus chilensis* through high-throughput sequencing. *Fish & Shellfish Immunology*, 35(6): 1899—1905
- Pante E, Rohfritsch A, Becquet V *et al*, 2012. SNP detection from de novo transcriptome sequencing in the bivalve *Macoma balthica*: marker development for evolutionary studies. *PLoS One*, 7(12): e52302
- Sauvage C, Bierne N, Lapègue S *et al*, 2007. Single nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. *Gene*, 406(1-2): 13—22
- Sellers M J, Dierens L, McWilliam S *et al*, 2014. Comparison of microsatellite and SNP DNA markers for pedigree assignment in Black Tiger shrimp, *Penaeus monodon*. *Aquaculture Research*, 45(3): 417—426
- Tsai H Y, Robledo D, Lowe N R *et al*, 2016. Construction and annotation of a high density SNP linkage map of the Atlantic salmon (*Salmo salar*) genome. *G3: Genes, Genomes, Genetics*, 6(7): 2173—2179
- Wang J L, Li Q, 2017. Characterization of novel EST-SNP markers and their association analysis with growth-related traits in the Pacific oyster *Crassostrea gigas*. *Aquaculture International*, 25(5): 1707—1719
- Wang Z L, Wu Z H, Jian J C *et al*, 2009. Cloning and expression of heat shock protein 70 gene in the haemocytes of pearl oyster (*Pinctada fucata*, Gould 1850) responding to bacterial challenge. *Fish & Shellfish Immunology*, 26(4): 639—645
- Wang Z L, Wang B, Chen G *et al*, 2016. Transcriptome analysis of the pearl oyster (*Pinctada fucata*) hemocytes in response to *Vibrio alginolyticus* infection. *Gene*, 575(2): 421—428
- Yu Y, Wei J K, Zhang X J *et al*, 2014. SNP discovery in the transcriptome of white Pacific shrimp *Litopenaeus vannamei* by next generation sequencing. *PLoS One*, 9(1): e87218
- Zhan X, Wen H Y, Shi Y H *et al*, 2016. Association between novel EST-SNPs and commercial traits in *Pinctada fucata martensii*. *Aquaculture Reports*, 3: 209—213
- Zhao Z, Boerwinkle E, 2002. Neighboring-nucleotide effects on single nucleotide polymorphisms: a study of 2.6 million polymorphisms across the human genome. *Genome Research*, 12(11): 1679—1686

SNP DISCOVERY AND FUNCTIONAL ANNOTATION IN TRANSCRIPTOME DATASETS FROM HEMOCYTES OF *PINCTADA FUCATA*

WANG Zhong-Liang^{1,2}, DING Yu¹, XU You-Hou³, WANG Bei¹, ZHANG Jian-Dong^{1,2}

(1. Fisheries College, Guangdong Ocean University, Zhanjiang 524088, China; 2. Key Laboratory of Aquaculture in South China Sea for Aquatic Economic Animal of Guangdong Higher Education Institutes, Zhanjiang 524088, China; 3. Guangxi Key Laboratory of Beibu Gulf Marine Biodiversity Conservation, Qinzhou University, Qinzhou 535000, China)

Abstract To enrich more SNP markers for prospective genetic studies in *Pinctada fucata*, SNPs were predicted and characterized by analyzing reads from hemocytes transcriptome. A total of 962995 SNPs of high quality were predicted from 67632 SNP-unigenes. The average SNP frequency was one SNP per 100 bp and the estimated ratio of transition to transversion was 0.5, which is in coincidence with the theoretical ratio. A/T was the most abundant SNP type among the six SNP mutation types. By matching GO, COG and KEGG databases, 30376 (44.91%), 18150 (26.84%, “general function prediction only”) was the most dominant group) and 10981 (16.24%, “signal transduction” was the largest group) SNP-unigenes were annotated, respectively. Further enrichment and SNP distribution analysis indicated that 629 SNP-unigenes were enriched in some well-known immune-related signal pathways, and 36 and 87 specific immune-related SNPs (123 SNPs in total) were detected from pre- and post-challenged transcriptomes, respectively. Additionally, 1594 SNPs were predicted in 17 differentially expressed immune-related unigenes based on the analysis on normalized gene expressed level.

Key words *Pinctada fucata*; single nucleotide polymorphism; transcriptome; molecular marker