

唐修阳,王传聪,项杰,等. 罗氏沼虾转录组免疫相关 SNP 的挖掘与分析[J]. 江苏农业科学,2019,47(4):145-148.
doi:10.15889/j.issn.1002-1302.2019.04.035

罗氏沼虾转录组免疫相关 SNP 的挖掘与分析

唐修阳,王传聪,项杰,欧江涛,王资生

(盐城工学院海洋与生物工程学院,江苏盐城 224051)

摘要:为了挖掘罗氏沼虾免疫相关单核苷酸多态性(SNP),采用第二代高通量测序技术,对罗氏沼虾的肝胰腺进行转录组深度测序分析。结果发现 SNP 位点 37 589 个,其中转换 25 188 个,颠换 12 401 个。再对 SNP 所在的基因进行 GO(gene ontology)分类和 KOG 注释,分别筛选到 2 451 个有 GO 注释和 2 458 个有 KOG 注释的 SNP,并重点对 4 个重要免疫通路的 145 个免疫相关基因,筛选出 129 个 SNP 位点。相关研究结果将为进一步系统挖掘免疫相关 SNP 的分子遗传标记奠定基础,为罗氏沼虾疾病防控和选育抗病新品种提供科学参考。

关键词:罗氏沼虾;转录组;SNP;高通量测序技术;免疫;抗病

中图分类号: S942.5 **文献标志码:** A **文章编号:** 1002-1302(2019)04-0145-04

罗氏沼虾(*Macrobrachium rosenbergii*)又名马来西亚大虾、淡水长臂虾等,广泛分布在江苏、广东和浙江等 17 个省(市、自治区),年产量超过 12 万 t,是重要的淡水养殖经济虾种之一^[1-2]。近年来,由于种质退化、病原滋生等原因,造成罗氏沼虾养殖生产中存在生长缓慢、抗病性减弱以及繁殖力下降等问题,给我国渔业带来严重的经济损失,同时也制约了水产甲壳类动物的可持续发展,相关问题的解决现已提上日程^[3-4]。

单核苷酸多态性(single nucleotide polymorphism,简称 SNP),通常是指基因组中单个碱基变异所引起的遗传多态性,其变异类型主要包括转换、颠换、缺失和插入。单碱基突变,可导致蛋白发生变异,从而影响生命体有关性状的生物学功能;已有大量研究表明,许多免疫通路中关键基因的 SNP 是很好的分子免疫遗传标记,在宿主的疾病防控和抗病育种中具有十分重要的意义^[5-7]。

通过转录组测序技术,结合同种基因组信息,可以更容易地找到与目标性状相关的 SNP。本研究通过对罗氏沼虾肝胰腺转录组的深度测序分析,筛选出大量 SNP 位点,并对这些 SNP 所在的基因进行功能注释,重点对重要免疫通路中的关键基因进行了系统的免疫相关 SNP 的挖掘与筛选。相关结果将为罗氏沼虾的抗病研究提供新的方法与思路,可为下一抗抗病分子设计育种提供科学的理论与试验依据。

1 材料与方法

1.1 试验动物与螺原体感染

动物的感染试验于 2013 年在南京师范大学的江苏省水

生甲壳动物病害重点实验室开展,试验动物为 100 尾罗氏沼虾,个体质量为 20~25 g,螺原体检测结果均为阴性,于 26~28 ℃水温养殖备用。螺原体在 R₂ 液体培养基中于 30 ℃孵育 48 h,待其生长到对数期备用^[8],将罗氏沼虾均分成 2 组。对照组注射 100 μL 无螺原体的 R₂ 培养基;试验组注射 100 μL 含螺原体 MR-1008 的 R₂ 培养基。

1.2 罗氏沼虾肝胰腺转录组数据的获得和 SNP 检测

肝胰腺组织总 RNA 采用 Trizol(Invitrogen 公司)提取法提取,提取产物通过 Agilent 2100 Bioanalyzer(美国)($D_{260\text{ nm}}/D_{280\text{ nm}}$ 为 1.8~2.2)和 15 g/L 琼脂糖凝胶电泳进行质检(28S rRNA/18S rRNA > 1.0),样品质检合格。用干冰保护并送往杭州联川生物技术股份有限公司,然后按照 Illumina 公司的测序标准进行测序。由于测序存在一定错误,本研究以测序深度大于 100 为阈值避免假阳性 SNP 的产生,使用 SOAPsnpc 软件分析 SNP 位点信息。

1.3 GO(gene ontology)和 KOG 功能分析

把 SNP 所在的基因映射到罗氏沼虾 GO 注释文件中,并在 WEGO(<http://wego.genomics.org.cn/>)中对这些基因进行 GO 功能分类统计。以 KOG 数据库为参考,对这些基因的功能进行分析。

1.4 挖掘免疫相关 SNP

根据测序得到的整个罗氏沼虾肝胰腺转录组中的 SNP,结合挖掘到的 Toll、IMD、JAK/STAT 和 MAPK 4 个免疫通路中的免疫因子,进行免疫相关 SNP 的挖掘。

2 结果与分析

2.1 样品质量检测

采用 Trizol 方法提取肝胰腺组织的总 RNA,通过琼脂糖凝胶电泳检测,图 1 显示,5S、18S 和 28S RNA 条带清晰,28S RNA 条带亮度是 18S RNA 条带亮度的 2 倍关系,紫外分光光度 $D_{260\text{ nm}}/D_{280\text{ nm}}$ 测定结果为 1.9~2.0,说明 RNA 提取的质量较好。

2.2 SNP 分子标记筛选

通过 SOAPsnpc 软件分析,发现潜在 SNP 位点 37 589 个,

收稿日期:2017-10-16

基金项目:国家自然科学基金面上项目(编号:31570176、31872601);

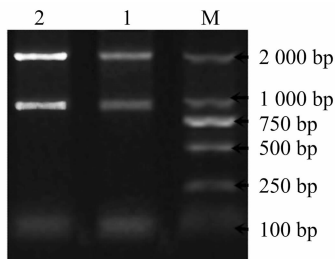
国家自然科学基金青年科学基金(编号:31000072)。

作者简介:唐修阳(1995—),男,江苏盐城人,主要从事海洋生物技术研究。E-mail:654743945@qq.com。

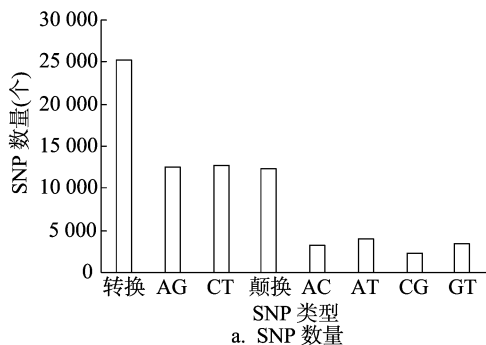
通信作者:欧江涛,博士,副教授,主要从事海洋生物技术研究。

E-mail:ojt110@126.com;王资生,博士,教授,主要从事海洋生态

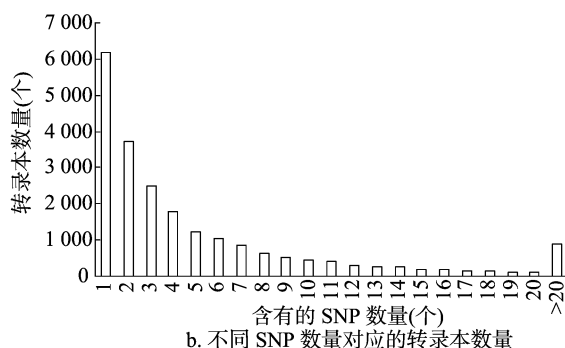
学研究。E-mail:wzs399@126.com。



M—2 000 bp marker; 1—健康罗氏沼虾肝胰腺总 RNA; 2—螺旋体感染罗氏沼虾肝胰腺总 RNA
图1 总 RNA 完整性检测

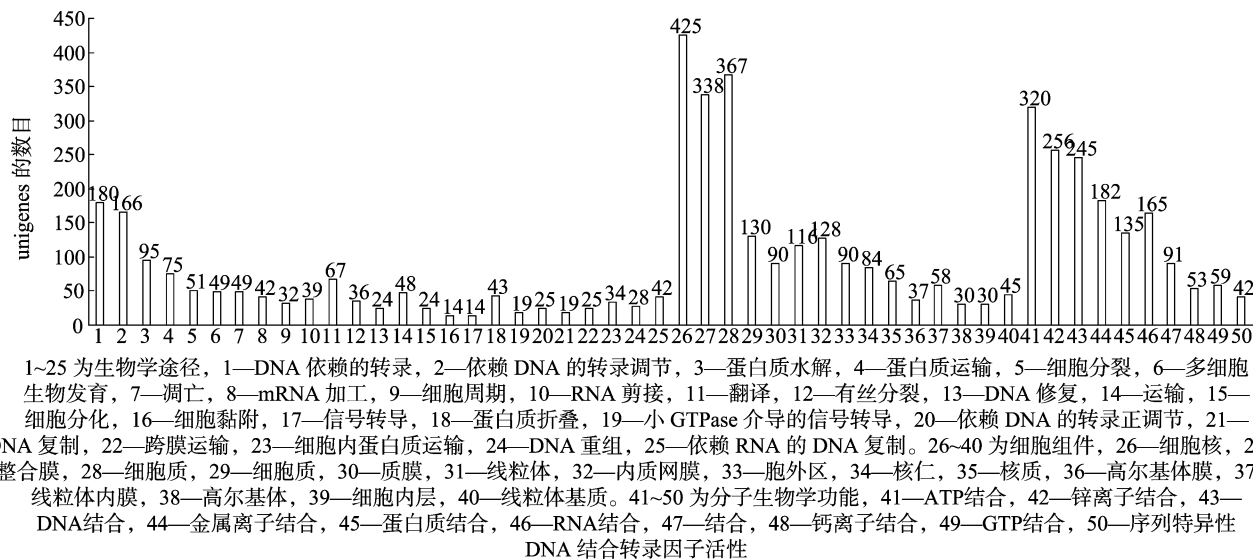


a. SNP 数量



b. 不同 SNP 数量对应的转录本数量

图2 SNP 的类型统计



1~25 为生物学途径, 1—DNA 依赖的转录, 2—依赖 DNA 的转录调节, 3—蛋白质水解, 4—蛋白质运输, 5—细胞分裂, 6—多细胞生物发育, 7—凋亡, 8—mRNA 加工, 9—细胞周期, 10—RNA 剪接, 11—翻译, 12—有丝分裂, 13—DNA 修复, 14—运输, 15—细胞分化, 16—细胞黏附, 17—信号转导, 18—蛋白质折叠, 19—小 GTPase 介导的信号转导, 20—依赖 DNA 的转录正调节, 21—DNA 复制, 22—跨膜运输, 23—细胞内蛋白质运输, 24—DNA 重组, 25—依赖 RNA 的 DNA 复制。26~40 为细胞组件, 26—细胞核, 27—整合膜, 28—细胞质, 29—细胞质, 30—质膜, 31—线粒体, 32—内质网膜, 33—胞外区, 34—核仁, 35—核质, 36—高尔基体膜, 37—线粒体内膜, 38—高尔基体, 39—细胞内层, 40—线粒体基质。41~50 为分子生物学功能, 41—ATP 结合, 42—锌离子结合, 43—DNA 结合, 44—金属离子结合, 45—蛋白质结合, 46—RNA 结合, 47—结合, 48—钙离子结合, 49—GTP 结合, 50—序列特异性 DNA 结合转录因子活性

图3 SNP-Unigenes 的 GO 分类

(molecular function)、生物学途径 (biological process) 和细胞组件 (cellular component) 3 方面进行聚类分析。

KOG 是真核生物直系同源蛋白簇 (clusters of orthologous groups for eukaryotic complete genomes), 共分为 24 类。结果显示, 在一般功能预测 (general function prediction only, 479 个), 翻译后的修饰、蛋白周转、分子伴侣 (posttranslational modification, protein turnover, chaperones, 303 个) 和信号转导机制 (signal transduction mechanisms, 193 个) 3 个方面聚集的最多, 细胞运动 (cell motility, 3 个) 最少 (表 1 和图 4)。

2.4 免疫相关 SNP 筛选

对于 Toll、IMD、JAK/STST 和 MAPK 4 个主要免疫通路, 本研究通过筛选得到 145 个免疫关键基因, 并挖掘到 129 个 SNP 位点 (表 2)。其中 4 个通路中这 145 个基因相关免疫 SNP 数

其中转换 25 188 个, 颠换 12 401 个, 转换的 SNP 位点数约占 2/3, 明显高于颠换。对不同转录本中 SNP 数量进行统计发现, 大部分转录本中的 SNP 数量少于 6 个 (图 2)。

2.3 含 SNP 位点 Unigene (SNP-Unigenes) 的功能注释

将得到注释的 7 062 个 Unigenes 与 GO、KOG 数据库中的蛋白序列进行 BLAST 比对, 取阈值 $e \leq 10^{-10}$, 分别筛选到 2 451 个有 GO 注释和 2 458 个有 KOG 注释的 SNP 标记。

GO 是适用于各物种, 且不断更新描述和限定基因与蛋白功能的语言词汇标准, SNP-Unigenes 的 GO 分类结果见图 3。通过软件对获得 GO 注释的基因从分子生物学功能

分别为 1、10、27、91 个。这些 SNP 位于 5' 区、CDS 区 (蛋白质编码区) 和 3' 区的 SNP 所占比例分别为 6.98%、53.49%、39.53%, 位于 CDS 区的 SNP 数量超过总数量的 1/2。

3 讨论

本研究发现潜在 SNP 位点 37 589 个, SNP 发生频率是 1 152 bp 中就有 1 个 SNP, 与 Jung 等对罗氏沼虾的表达序列标签 (EST) 分析得出的 945 bp 1 个 SNP^[9] 和人基因组中 1 kb 1 个 SNP^[10] 大致同等水平, 低于中华绒螯蟹的 189 bp 1 个 SNP^[11] 和大西洋鲑的 641 bp 1 个 SNP^[12]。这些 SNP 的发现丰富了水产甲壳动物的分子标记的数据, 为促进分子标记挖掘和育种提供了分子材料。尤其是针对 4 个免疫通路 145 个免疫关键基因中挖掘到 129 个 SNP 位点, 经过后续的试验验证

表 1 SNP - unigenes 的 KOG 分类

KOG 功能代号	KOG 功能注释信息	Unigene 数量(条)
A	RNA 加工和修饰	122
B	染色质结构和动力学	39
C	能源生产与转化	127
D	细胞周期控制,细胞分裂,染色体分配	73
E	氨基酸转运和代谢	96
F	核苷酸转运和代谢	28
G	碳水化合物运输和代谢	137
H	辅酶运输和新陈代谢	30
I	脂质运输和新陈代谢	122
J	翻译,核糖体结构和生物合成	177
K	转录	140
L	复制,重组和修复	56
M	细胞壁/膜/包膜生物合成	22
N	细胞运动	3
O	翻译后修饰,蛋白质周转,分子伴侣	303
P	无机离子运输和代谢	43
Q	次生代谢物生物合成、运输和分解代谢	73
R	一般功能预测	479
S	功能未知	177
T	信号转导机制	193
U	胞内运输、分泌和囊泡运输	137
V	防御机制	20
W	细胞外结构	16
Y	核结构	10
Z	细胞骨架	83

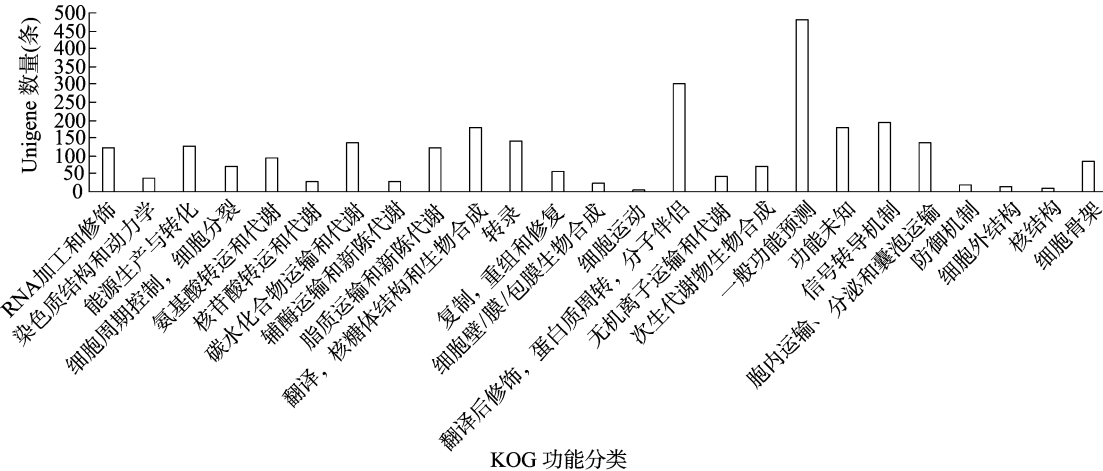


图4 SNP-Unigenes 的 KOG 分类

证和实践应用,这些免疫候选标记将可能为虾类培育出具有强抗病性的新品种。

参考文献:

[1] Holthuis L B. An annotated catalogue of species of interest to fisheries [J]. FAO Fisheries Synopsis,1980,125(1):103.

[2] 马 劳. 罗氏沼虾养殖手册[M]. 罗马:联合国粮食及农业组织,2005:1-11.

[3] Sahul H S,Yoganandhan K,Sri W J,et al. Experimental transmission and tissue tropism of *Macrobrachium rosenbergii* nodavirus (MrNV) and its associated extra small virus(XSV) [J]. Diseases of Aquatic

Organisms,2004,62(3):191.

[4] Bonami J R, Shi Z, Qian D, et al. White tail disease of the giant freshwater prawn, *Macrobrachium rosenbergii*: separation of the associated virions and characterization of Mr NV as a new type of nodavirus[J]. Journal of Fish Diseases,2005,28(1):23-31.

[5] 唐立群,肖层林,王伟平. SNP 分子标记的研究及其应用进展 [J]. 中国农学通报,2012,28(12):154-158.

[6] 谭 新,童金苟. SNPs 及其在水产动物遗传学与育种学研究中的应用[J]. 水生生物学报,2011,35(2):348-354.

[7] Vignal A,Milan D,Sancristobal M,et al. A review on SNP and other types of molecular markers and their use in animal genetics [J]. Genetics Selection Evolution,2002,34(3):275-305.

表 2 4 个通路中免疫基因的 SNP

信号通路名称	Unigenes	SNP 类型(基因结构中的位置)	SNP 数量(个)
Toll 信号通路 Spatzle	comp21497_c0	2000TC(CDS)	1
IMD 信号通路 Dredd/Casp	comp20643_c0	201AG(5′)、205GT(5′)、686GA(CDS)、1226TA(CDS)、1937AG(3′)、1971AG(3′)、2053CA(3′)、2142AC(3′)、2171TC(3′)、2592TC(3′)	10
JAK/STAT 信号通路 CytokinR	comp20403_c0	1607GA(3′)、1710CT(3′)	2
JAK	comp16930_c0	3085TC(CDS)	1
CBP	comp21736_c0	5336TC(CDS)、5423GT(CDS)、6036AG(CDS)	3
CycD	comp22321_c0	135TC(5′)、199TA(CDS)、475TC(CDS)、560TC(CDS)、745CT(CDS)、781TA(CDS)、949TG(CDS)、1028AG(CDS)、1259TG(CDS)、1286AG(CDS)、1399CT(3′)、1626AG(3′)、1661TG(3′)、1756CT(3′)、1842GA(3′)、1868CT(3′)、1916TG(3′)、1936GC(3′)	18
BelXL	comp9195_c0	177GA(5′)	1
Spred	comp18313_c0	233TC(3′)、250TC(3′)	2
MAPK 信号通路 CACN	comp21974_c0	460TC(CDS)、514TC(CDS)、1093GA(CDS)、1210AG(CDS)、2173AT(CDS)、2323AG(CDS)、2491TC(3′)	7
	comp4736_c0	736TG(CDS)、1144GA(5′)	2
G12	comp24564_c0	173AG(5′)	1
p120GAP	comp14947_c0	2427TC(CDS)	1
Rap1	comp10157_c0	2232TA(3′)	1
ERK	comp19937_c0	185GA(CDS)	1
RSK2	comp4632_c0	237CG(CDS)	1
	comp10815_c0	1912GA(CDS)	1
SRF	comp21780_c0	157AG(CDS)、161TC(CDS)、318TC(CDS)、922GA(CDS)、1088AC(CDS)	5
MKP	comp18590_c0	1897GA(CDS)、1985TC(CDS)、2689CT(CDS)、2791AG(CDS)、2847AG(CDS)	5
CASP	comp20643_c0	201AG(5′)、205GT(5′)、686GA(CDS)、1226TA(CDS)、1937AG(3′)、1971AG(3′)、2053CA(3′)、2142AC(3′)、2171T(3′)、2592TC(3′)	10
cdc42/Ra	comp20825_c0	197CT(3′)、202TG(3′)、248GA(3′)、253CG(3′)、255AT(3′)、373CT(3′)、621TG(CDS)、649AG(CDS)、808GA(CDS)、982GA(CDS)	10
PAK1/3	comp16131_c0	159GA(3′)	1
MST1/2	comp22322_c0	1078TC(CDS)、2008TC(CDS)、2489TC(3′)、2911GA(3′)、3947CT(3′)、4234CA(3′)	6
HSP72	comp20130_c0	268AG(CDS)、562TC(3′)、676CT(3′)	3
	comp20130_c1	58CA(CDS)	1
	comp20130_c2	64GA(CDS)、163CT(CDS)、180GA(CDS)	3
	comp20130_c4	85TC(CDS)、163AT(CDS)、166GA(CDS)、301GA(CDS)、469GA(5′)、298AG(CDS)、370TA(CDS)、412TA(CDS)	8
	comp23028_c1	1221CT(CDS)、1224GA(CDS)、1242GC(CDS)、1254CT(CDS)、1380CT(CDS)、1785GA(CDS)、1817AG(CDS)、1821GC(CDS)	8
	comp4205_c0	433AG(CDS)	1
p53	comp22048_c1	1429TC(CDS)	1
MAX	comp4363_c0	326TC(CDS)	1
MSK1/2	comp10815_c0	1912GA(CDS)	1
MAPK	comp19083_c0	2244CT(3′)、2259AG(3′)、2286CT(3′)、2318CT(3′)、2322GT(3′)、2244CT(3′)、2259AG(3′)、2286CT(3′)、2295TC(3′)、2316GA(3′)、2341AT(3′)、2367AC(3′)	12

[8]Steiner T, McGarrity G J, Phillips D M. Cultivation and partial characterization of spiroplasmas in cell cultures [J]. Infection & Immunity,1982,35(1):296.

[9]Jung H, Lyons R E, Dinh H, et al. Transcriptomics of a giant freshwater prawn (*Macrobrachium rosenbergii*): de novo assembly, annotation and marker discovery [J]. PLoS One, 2011, 6 (12):e27938.

[10]Frazer K A, Ballinger D G, Cox D R, et al. A second generation human haplotype map of over 3.1 million SNPs[J]. Nature,2007, 449(7164):851 – 861.

[11]李希红. 中华绒螯蟹(*Eriocheir sinensis*)高通量转录组解析及免疫通路关键基因筛选[D]. 青岛:中国科学院大学,2014.

[12]Hayes B, Laerdahl J K, Lien S, et al. An extensive resource of single nucleotide polymorphism markers associated with Atlantic salmon (*Salmo salar*) expressed sequences [J]. Aquaculture, 2007, 265 (1/2/3/4):82 – 90.