

DOI: 10.3724/SP.J.1118.2015.14361

三疣梭子蟹生长相关 SNP 位点的鉴定

张德宁^{1,2}, 吕建建¹, 刘萍¹, 冯艳艳¹, 高保全¹, 李健¹

1. 中国水产科学研究院 黄海水产研究所, 山东 青岛 266071;

2. 上海海洋大学 水产与生命学院, 上海 201306

摘要: 为了发掘三疣梭子蟹(*Portunus trituberculatus*)与生长相关的 SNP, 本研究基于三疣梭子蟹比较转录组学数据, 筛选出生长相关候选基因片段 19 条, 利用该基因序列设计引物进行 PCR 扩增, 得到总长为 17 387 bp 的 DNA 片段; 通过直接测序法发现 74 个 SNP 位点, 分布频率为 0.43/100 bp, 其中转换突变的比例为 80%, 颠换突变的比例为 20%, 转换的比例远远大于颠换, 符合“transition bias”原理。C/T(G/A)突变所占比例为 60%, G/T(C/A)突变占 20%, A/T 突变占 9.33%, G/C 突变占 10.67%, C/T(G/A)突变所占比例最大。内含子上突变发生的频率为 1.34/100 bp, 外显子上仅为 0.17/100 bp, 说明外显子上的碱基更加保守。通过飞行质谱法, 在三疣梭子蟹生长性状分离群体中成功分型了 10 个 SNP 位点; 通过卡方检验和多元方差分析的方法进行性状关联分析, 发现 3 个 SNP 位点与生长性状显著相关($P<0.05$), 并且 *comp58070-R31* 位点与生长性状极显著相关($P<0.01$); 一般线性模型的多元方差分析显示, *comp58070-R31* 位点与体重、全甲宽、甲宽 3 个性状极显著相关($P<0.01$), 与体长、体高 2 个性状为显著相关($P<0.05$); *comp46623-F49* 位点与甲宽显著相关($P=0.05$); *comp49193-R333* 位点与甲宽显著相关($P<0.05$)。通过对 SNP 位点的多态性分析发现, 三疣梭子蟹 SNP 标记观测杂合度(H_o)范围为 0.162 8~0.833 3, 期望杂合度(H_e)范围为 0.189 6~0.591 2, 并且显示 SNP 标记的信息量低于微卫星标记。该研究的结果将有助于推进三疣梭子蟹分子标记辅助育种进程。

关键词: 三疣梭子蟹; 生长性状; 转录组; SNP; 关联分析

中图分类号: S917

文献标志码: A

文章编号: 1005-8737-(2015)03-0393-09

三疣梭子蟹(*Portunus trituberculatus*)是重要的大型海洋甲壳类经济动物, 隶属甲壳纲(Crustacea)、十足目(Decapoda)、梭子蟹科(Portunidae), 分布于中国、朝鲜、日本等海域^[1-3]。由于其生长快、个体大、肉味鲜美, 已成为我国重要的渔业捕捞对象和海水养殖对象^[4]。生长性状是影响三疣梭子蟹经济价值最主要的因素。随着三疣梭子蟹养殖规模的不断扩大, 亟需对三疣梭子蟹进行遗传性状的改良, 选育出一种快速生长的品系。在分子标记辅助育种技术蓬勃发展的今天, 人们可以通

过分子标记技术和育种的经济性状进行关联分析来达到良种选育和性状改良的目的。在一系列的分子标记当中, 由于操作方便、信息量多等原因, 微卫星 DNA 分子标记曾被广泛使用, 但仍不能满足研究人员的实际需求。近年来, 分子标记单核苷酸多态性标记(single nucleotide polymorphisms, SNP)的出现, 弥补了微卫星 DNA 分子标记的不足^[5]。

SNP 是指单个核苷酸发生改变而引起的 DNA 序列的多态性, 最早在 1996 年由 Lander^[6]提出。

收稿日期: 2014-08-28; 修订日期: 2014-10-10.

基金项目: 国家 863 计划项目(2012AA10A409); 农业科技成果转化资金项目(2013GB23260589); 山东省自主创新专项(2013CX80202); 国家自然科学基金项目(41306177); 中国博士后科学基金特别资助项目(2014T70668)。

作者简介: 张德宁(1988-), 男, 硕士研究生, 研究方向为三疣梭子蟹生长和低盐适应相关 SNP 筛选及功能验证研究. E-mail: youzaiyouzaihapp@163.com

通信作者: 刘萍, 研究员. E-mail: liuping@ysfri.ac.cn

由于具有数量多、密度大、两个等位基因多态、可以引起氨基酸突变和易于高通量分析等优点, SNP 标记被广泛应用于遗传连锁图谱的构建、基因定位、遗传多样性研究、品种鉴定和动植物育种等领域^[7-10]。在大西洋鲑(*Clupea harengus*)^[9]、大鳞大麻哈鱼(*Oncorhynchus tshawytscha*)^[7]、舌鳎(*Cymoglossus robustus*)和大西洋鲱(*Clupea harengus*)^[11]、北极嘉鱼(*Salvelinus alpinus*)^[12]、嘉鱼(*alpinus*)^[13]、罗非鱼(*Tilapia*)^[14]、鲤(*Cyprinus carpio*)^[15]等鱼类中均有 SNP 相关研究。在甲壳类的凡纳滨对虾(*Litopenaeus vannamei*)^[16]、拟穴青蟹(*Scylla paramamosain*)^[17]、三疣梭子蟹^[18]等物种中也有应用 SNP 标记的研究报道。

在缺乏全基因组测序的物种中, 通过比较转录组学方法筛选性状相关基因已经成为研究某一机理的普遍方法^[16-18]。本实验基于实验室前期工作^[19], 在生长差异个体中获得差异表达基因 117 对, 这些基因很可能参与三疣梭子蟹生长调控进程。通过性状决定基因筛选到性状相关 SNP 的几率更大^[16-17, 20]。本研究采用直接测序法筛选 SNP 位点, 并利用飞行质谱的方法, 在性状分离群体中对 SNP 位点进行分型, 通过关联分析寻找与生长性状相关的 SNP 位点, 研究结果将有助于三疣梭子蟹分子标记辅助育种工作, 最终加快育种进程。

1 材料和方法

1.1 实验材料

实验用三疣梭子蟹取自“黄选 1 号”核心群体, 饲养于昌邑海丰水产有限公司。160 日龄时测量生长性状(体重、全甲宽、甲宽、体长、体高), 从同一池塘中挑选体重具显著差异个体各 45 只, 分为小个体组 SG 与大个体组 LG, 建立生长性状差异群体。根据体重大小顺序排列, 取肌肉组织, -20℃ 保存, 用于提取基因组 DNA。

1.2 实验方法

1.2.1 三疣梭子蟹基因组 DNA 的提取及检测 采用“酚-氯仿法”提取肌肉中的基因组 DNA。基因组 DNA 利用 1% 琼脂糖凝胶电泳进行检测。

1.2.2 三疣梭子蟹 SNP 位点初步筛选 基于本实

验室之前的研究^[20], 根据比较转录组学所分析的结果, 筛选出表达差异较大的 19 条序列, 运用引物设计软件 Primer Premier 5.0 在各个基因序列的保守区域设计引物, 共设计引物 20 对, 由上海生工技术公司合成, 见表 1。

1.2.3 PCR 反应和 PCR 产物检测 取 SG 全甲宽最小的 15 个样品的 DNA, 与 LG 全甲宽最大的 15 个样品的 DNA 进行混合, PCR 反应利用混合 DNA 作为模板, PCR supermix kit 试剂盒(TaKaRa)进行扩增, PCR 产物用 1% 琼脂糖凝胶电泳检测, 产物送上海桑尼公司测序。

1.2.4 测序峰图分析及 SNP 位点分型 寻找峰图清晰、有明显双峰的位点, 并定其为 SNP 位点。并应用飞行质谱(TOF-MS)方法对 SNP 位点进行分型。

1.2.5 SNP 位点与 5 个生长性状的关联分析 利用 SPSS 17.0 对分型结果进行卡方检验和多元方差分析检验, $P < 0.05$ 为显著相关, $P < 0.01$ 为极显著相关。

2 结果与分析

2.1 生长性状数据分析

2.1.1 三疣梭子蟹各生长性状测量数据 SG 组体重 79.9~171.9 g, 全甲宽 103.13~140.37 mm, 甲宽 52.39~67.58 mm, 体长 55.90~70.62 mm, 体高 27.98~39.67 mm。LG 体重 174.0~289.0 g, 全甲宽 140.6~166.42 mm, 甲宽 71.33~87.03 mm, 体长 72.47~82.57 mm, 体高 35.42~41.57 mm。统计结果见表 2。

2.1.2 各性状间的相关系数 三疣梭子蟹各形态性状和体重两两之间的相关系数见表 3。统计分析结果表明, 各形态性状间的表型相关均达到极显著水平($P < 0.01$), 所有的形态参数之间均为显著正相关, 表明将所选的指标用来进行相关分析具有很重要的意义。其中甲宽与体重的相关系数最大, 为 0.945; 体高与体重的相关系数最小, 为 0.841。三疣梭子蟹各形态性状与体重间的相关关系由大到小的顺序依次为甲宽、全甲宽、体长、体高。

表 1 引物设计列表
Tab. 1 List of primers

基因编号 gene code	引物序列(5'-3') primer sequence(5'-3')	设计长度/bp designed length	产物长度/bp product length	是否有内含子 intron or not
<i>comp 54958</i>	F: AACTCCCCACCTGATAATGTCCTCG R: CGTAACTTCGGGATAAGGATTGGC	1000	1200	是 yes
<i>comp 54958</i>	F: ATTGAACTGGTCCCGAGGTTGGC R: CCCGAAAGATGGTGAATATGCCTG	1275	1275	否 no
<i>comp 55883</i>	F: AGGCTTTCAGGCATAATCCACG R: ACTTACTCGGTGAGACGGAATCGG	700	700	否 no
<i>comp 58070</i>	F: CCTTCGTGGGCAACGCTGCTTT R: CTCCTTGCTTGAATGCCACAGTAT	1531	1700	是 yes
<i>comp 58769</i>	F: GAAGTTGGAGACGCCAGTGATGATG R: GAGCACAAGGACAGTCTTACCCCC	843	843	否 no
<i>comp 48505</i>	F: ATCTTGATTTTGATGGTGGAGGGAGC R: TGGACTCTGGTGACGGCGTAACTC	528	600	否 no
<i>comp 45541</i>	F: GCCTTAGAACTTGATGCCAGCCTG R: TGGCTCCCGCCTGTGTATTGTG	741	800	否 no
<i>comp 55595</i>	F: AAAGCCCCAGACACGACAACAGG R: CTCCTGTGGTCCAAAAGGTGCG	794	—	—
<i>comp 56676</i>	F: CCAGAATGTGTTGTCACCGAGCC R: GACACCTGGATGGGCTGCTCAA	1319	800	否 no
<i>comp 58769</i>	F: CTGGGAAACCTTCGCTGGCAAC R: ACGCCGTCCACATCCTCAAAGT	566	700	是 yes
<i>comp 57303</i>	F: CTCGGTTGGTTGGTGTCTTCCTC R: TTTGTCAGCATCTCCAGATAGCG	670	1200	是 yes
<i>comp 50558</i>	F: GTAGTCCATCTTGATCTTGAGGCTGTC R: CTAACGGTTATCGCCGGTTTGTG	833	—	—
<i>comp 17448</i>	F: AGGATACCCAAGGGCTTCTCAAT R: TCTACGCTCATGTCCAACCAG	736	1600	是 yes
<i>comp 40819</i>	F: TGTAGACGGTCTGCCAGCCTTG R: GCGATACCATCCCGCTTACTCCC	1208	2500	是 yes
<i>comp 52896</i>	F: AGTTGGCTCGGTGGAATCTTGTG R: ATCGCTGTAGTCTATCCTGAACTTGG	760	—	—
<i>comp 32720</i>	F: AAGGGCAAGCAAATACCAGAAAAG R: ACAGGCACGGATACCGCATCGTC	569	569	否 no
<i>comp 57375</i>	F: TTGGAAACCGAGGCTTCTGGACC R: CCCACCGAAGTATCCTCCGACAT	815	900	否 no
<i>comp 46623</i>	F: GCCGCTCGTGTTCAACCTCTGTC R: CGAATCCGCTAACATGACTTTACT	761	800	否 no
<i>comp 49193</i>	F: ATTCATTATTATTCAAGGGCAATC R: GTGACAGTCGCCTTGTTGGTGCT	522	1200	是 yes
<i>comp 54937</i>	F: AGGCATCATGGTTGGAATGGACC R: CATTTGCGATGAACGATAGTGGG	999	—	—

注：“—”表示未扩增出条带。

Note: “—”denotes no amplification result.

表 2 三疣梭子蟹生长性状参数
Tab. 2 The growth trait parameters of *Portunus trituberculatus*

n=45; $\bar{x} \pm SD$

性状 trait	分组 group	
	小个体组 SG	大个体组 LG
体重/g body weight	121.966 7±24.932 77	216.855 1±29.527 14
全甲宽/mm full carapace width	125.112 7±9.136 15	150.054 0±7.010 56
甲宽/mm carapace width	61.368 0±4.868 59	77.160 9±4.0413 8
体长/mm body length	62.949 8±3.967 00	73.815 1±3.897 18
体高/mm body height	32.863 6±2.671 19	37.987 3±.907 50

表 3 三疣梭子蟹各形态性状之间的相关系数
Tab. 3 Pearson correlation coefficient between morphological traits in *Portunus trituberculatus*

性状 trait	体重 body weight	全甲宽 full carapace width	甲宽 carapace width	体长 body length	体高 body height
体重 body weight	1				
全甲宽 full carapace width	0.927**	1			
甲宽 carapace width	0.945**	0.923**	1		
体长 body length	0.907**	0.885**	0.882**	1	
体高 body height	0.841**	0.840**	0.832**	0.873**	1

注: **表示极显著相关($P<0.01$).
Note: ** denotes extremely significant correlation ($P<0.01$).

2.2 基因组 DNA 提取结果

从图 1 可见, 基因组 DNA 主带清晰, 表明基因组 DNA 提取质量可靠, 完整性良好, 满足后续实验条件。

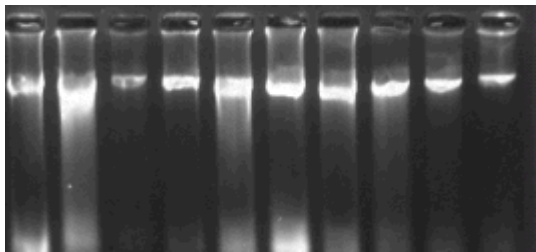


图 1 三疣梭子蟹基因组 DNA 电泳图谱
Fig. 1 Gel electrophoresis of *Portunus trituberculatus* genome DNA

2.3 SNP 位点筛选与分型

本研究通过直接测序的方法, 共挑选了 30 个位点送分型分析, 利用飞行质谱法成功分型了 10 个 SNP 位点。为了验证分型的准确率, 本实验采用直接测序的方法检验了一部分分型结果。

2.4 SNP 多态性分析

本研究使用 POPGENE 软件统计分析各 SNP 位点的观测杂合度(observed heterozygosity, H_o)、期望杂合度(expected heterozygosity, H_e)、最小等位基因频率(minor allele frequency, MAF)以及哈

迪-温伯格平衡(Hardy-Weinberg equilibrium, HWE)检验(表 4)。

根据表 4 可知, 三疣梭子蟹 SNP 标记观测杂合度(H_o)范围为 0.162 8~0.833 3, 期望杂合度(H_e)范围为 0.189 6~0.591 2; 通过本实验室之前的研究发现^[21], 三疣梭子蟹微卫星标记的观测杂合度(H_o)范围为 0.222 2~1.000 0, 期望杂合度(H_e)范围为 0.436 7~0.909 9; 两种标记多态性的差异显示 SNP 标记的信息量低于微卫星标记^[22]。

*comp57303-F54*和*comp57303-F352*是位于同一基因上两个连锁的 SNP 位点, 所以多态性分析数据完全一致。*comp55883-R42* 极显著($P<0.01$)偏离 Hardy-Weinberg 平衡, *comp32720-F318* 位点显著($P<0.05$)偏离 Hardy-Weinberg 平衡(HWE), 在 SG 和 LG 内的基因型分布都不符合 HWE。推测原因, 一是“黄选 1 号”三疣梭子蟹并非自然随机交配群体, 而是经过人为设计交配方案的选育群体; 二是 SG 和 LG 是人为挑选的, 不符合 HWE 的结果可作为 SNP 与生长性状关联的初步判断。除 *comp55883-R42* 和 *comp32720-F318* 两个位点, 其他 SNP 位点在两组内都符合 HWE。上述 SNP 的 MAF 远大于 0.05, 说明这些 SNP 位点的多态性较高, 适合进行关联分析。

表 4 SNP 位点基因多态性分析
Tab. 4 The analysis of gene polymorphism of SNPs

SNP	观测杂合度 H_o		期望杂合度 H_e		最小等位基因频率 MAF		哈迪—温伯格平衡 $HWE(p)$	
	SG	LG	SG	LG	SG	LG	SG	LG
<i>comp55883-R42</i>	0.756 8	0.756 8	0.476 9	0.476 9	0.378 4	0.378 4	0.000 3	0.000 3
<i>comp58070-R31</i>	0.255 8	0.431 8	0.320 9	0.499 2	0.197 7	0.443 2	0.172 1	0.364 8
<i>comp57303-F54</i>	0.162 8	0.250 0	0.189 6	0.286 1	0.104 7	0.170 5	0.327 9	0.389 0
<i>comp57303-F352</i>	0.162 8	0.250 0	0.189 6	0.286 1	0.104 7	0.170 5	0.327 9	0.389 0
<i>comp32720-F129</i>	0.219 5	0.250 0	0.197 8	0.221 3	0.109 8	0.1250	0.459 2	0.368 6
<i>comp32720-F318</i>	0.476 2	0.545 5	0.367 2	0.401 3	0.238 1	0.272 7	0.049 3	0.015 3
<i>comp46623-F49</i>	0.348 8	0.386 4	0.396 4	0.499 2	0.267 4	0.443 2	0.422 5	0.129 2
<i>comp46623-F358</i>	0.2750	0.162 8	0.3085	0.189 6	0.187 5	0.104 7	0.478 8	0.327 9
<i>comp46623-F511</i>	0.232 6	0.295 5	0.2430	0.367 6	0.139 5	0.238 6	0.770 6	0.184 2
<i>comp49193-R333</i>	0.5833	0.8333	0.476 1	0.591 2	0.069 4	0.111 1	0.266 3	0.005 6

2.5 SNP 关联分析

将获得的分型数据利用 SPSS 17.0 软件对不同等位基因、基因型和生长性状之间的相关性进行卡方检验。经卡方检验分析, *comp58070-R31* 位点与生长性状为极显著相关($P<0.01$), 其他位点与生长性状的相关度不显著($P>0.05$); 一般线性模型的多元方差分析结果显示, *comp58070-R31* 位点与体重、全甲宽、甲宽 3 个性状呈现极显著相关关系($P<0.01$), 与体长、体高 2 个性状为

显著相关关系($P<0.05$); *comp46623-F49* 位点与甲宽为显著相关($P=0.05$); *comp49193-R333* 位点与甲宽为显著相关($P<0.05$), 其他位点与生长性状的相关度不显著($P>0.05$)。SG 和 LG 卡方检验和多元方差分析结果表明, *comp55883-R42* 和 *comp32720-F318* 2 个位点在 2 个群体间的基因型分布没有显著差异。但 2 个位点偏离 HWE 的结果提示, 在整个‘黄选 1 号’选育群体中该位点有可能受到了选择 (表 5)。

表 5 SNP 位点与生长性状的关联分析
Tab. 5 The correlation analysis of DNA SNPs with growth traits

SNP 位点 SNP locus	突变碱基 mutation base	位置 location	突变类型 mutation type	卡方检验 $\chi^2(P)$	一般线性模型多元方差分析 multivariate analysis of variance of general linear model MANOVA				
					体重 body weight	全甲宽 full carapace width	甲宽 carapace width	体长 body length	体高 body height
<i>comp55883-R42</i>	C-T	内含子 intron	N	0(1.000)	0.752	0.464	0.985	0.953	0.696
<i>comp58070-R31</i>	G-A	ORF	同义突变 synonymous mutation	9.792(0.007)**	0.001**	5.58E-05**	2.38E-04**	0.011*	0.033*
<i>comp57303-F54</i>	A-G	ORF	同义突变 synonymous mutation	1.453(0.484)	0.167	0.281	0.229	0.494	0.607
<i>comp57303-F352</i>	G-A	内含子 intron	N	1.453(0.484)	0.167	0.281	0.229	0.494	0.607
<i>comp32720-F129</i>	A-C	3'UTR	N	0.110(0.741)	0.487	0.478	0.649	0.964	0.391
<i>comp32720-F318</i>	G-C	3'UTR	N	0.413(0.521)	0.519	0.540	0.622	0.571	0.841
<i>comp46623-F49</i>	C-T	ORF	同义突变 synonymous mutation	4.981(0.083)	0.157	0.073	0.050*	0.366	0.158
<i>comp46623-F358</i>	C-T	ORF	同义突变 synonymous mutation	3.875(0.144)	0.397	0.226	0.360	0.439	0.200
<i>comp46623-F511</i>	C-T	ORF	同义突变 synonymous mutation	2.604(0.272)	0.261	0.358	0.268	0.407	0.323
<i>comp49193-R333</i>	G-A-C	内含子 intron	N	6.090(0.107)	0.155	0.199	0.032*	0.298	0.290

注: *表示显著相关($P<0.05$); **表示极显著相关($P<0.01$)。

Notes: *denotes significant correlation ($P<0.05$); ** denotes extremely significant correlation ($P<0.01$).

3 讨论

本研究成功扩增了 17 387 bp 长的 DNA 片段, 筛选出 74 个 SNP 位点, SNP 分布频率为 0.43/100 bp。SNP 在人类基因组中的分布频率为 0.1/100 bp^[23-28], 在牡蛎(*Ostrea gigas thunberg*)^[29]中为 2.5/100 bp, 在拟穴青蟹^[19]中为 0.3/100 bp, 扇贝(*Chlamys farreri*)^[30]中为 1/100 bp, 抹香鲸(*Physeter catodon*)^[31]中为 0.19/100 bp, 江豚(*Neophocaena phocaenoides*)^[32]中为 0.18/100 bp。各个物种中 SNP 分布频率差距较大, 可能与生长环境的复杂程度有关。另外, 本实验中 SNP 转换突变的比例为 80%, 颠换突变的比例为 20%, 转换的比例远远大于颠换, 符合中文“transition bias”原理。长牡蛎(*Crassostrea gigas*)^[33]SNP 标记的 84%是转换类型, 16%为颠换类型, 与本研究基本一致; C/T(G/A)突变所占比例为 60%, G/T(C/A)突变占 20%, A/T 突变占 9.33%, G/C 突变占 10.67%, C/T (G/A)所占比例最大。在其他物种中同样也发现了上述规律^[33-34]。出现这种规律的原因有两点: 一是 C 碱基容易发生 5-甲基化转换突变为 T 碱基, 突变频率是其他碱基突变的 10 倍; 二是 CpG 岛 5-甲基胞嘧啶非常频繁地发生脱氨基作用^[35]。通过分析突变前后的碱基比例发现, A+T 所占的比例在突变之后增加了, A/T 突变的频率高于 G/C, 这种现象可以用热力学原理来进行解释。在 DNA 中, 碱基之间是通过氢键来进行连接的, G 和 C 碱基之间有 3 个氢键, 而 A 和 T 之间只有 2 个, 所以 GC 碱基对比 AT 更加坚固, 更加不易发生突变。在果蝇(*Drosophila melanogaster*)和哺乳动物中^[36]同样发现, 在碱基替换之后, A+T 的比例增加。另外还发现内含子上突变发生的频率远远高于外显子, 内含子上的突变频率为 1.34/100 bp, 外显子上仅为 0.17/100 bp, 说明外显子上的碱基更加保守。

利用飞行质谱法分型的 10 个 SNP 位点的位置分别为: 位于 ORF 上的有 5 个, 都为同义突变; 3'UTR 上的突变有 2 个; 内含子上的突变有 3 个。3 个与生长性状显著相关的位点中, 有 1 个是内

含子上的突变, 2 个是 ORF 上的同义突变。众所周知, 大约 20%的错义突变会改变最终合成的蛋白质。错义突变改变性状的机理是通过改变氨基酸从而改变最终合成的蛋白质^[26, 37]。而同义突变改变性状的机理就比较复杂, 它可能通过几种机制来改变性状, 例如: 通过改变蛋白质的二级结构^[38]; 通过改变 mRNA 的翻译速度^[39]; 使常规密码子突变为稀有密码子从而改变合成蛋白质的效率; 产生一个隐藏的 mRNA 剪接位点和尾序列; 如果突变位于重叠基因区域, 突变还可能影响其他的基因表达等。内含子上的 SNP 位点可能通过改变剪接位点的活性来达到改变最终蛋白质的目的^[40]。本研究中的 SNP 位点导致性状改变的机制还需要进一步的研究阐明。

本研究首次在三疣梭子蟹中利用转录组数据开发生长相关 SNP, 为下一步的工作打下了良好的基础。将来工作是将这些位点标记到基因组上, 与功能基因相联系。具体可以分为两部分: 第一部分, 继续开发与性状相关的 SNP 位点, 如免疫相关、低盐胁迫相关等; 另一部分是研究 SNP 影响基因表达的分子机制, 比如, 影响了转录速率、翻译速率或者影响了蛋白质的折叠, 这些分子机制可以通过研究酶活性和信号通路来寻找突破口。本研究为下一步构建高密度连锁图谱、基因图谱、性状追踪、基因功能模型研究、基因表达通路研究等打下了坚实的基础。同时为生长相关分子机制研究以及生长相关新基因的开发提供了一种很好的手段。

参考文献:

- [1] Dai A Y, Zhng Q, Song Y Z, et al. The preliminary investigation of biological resources in *Portunus trituberculatus*[J]. Chinese Journal of Zoology, 1977, 12(2): 30-33.[戴爱云, 钟琪, 宋玉枝, 等. 三疣梭子蟹渔业生物资源的初步调查[J]. 动物学杂志, 1977, 12(2): 30-33.]
- [2] Dai A Y, Yang S L, Song Y Z, et al. Chinese Marine Crabs[M]. Beijing: China Ocean Press, 1986: 213-214.[戴爱云, 杨思谅, 宋玉枝, 等. 中国海洋蟹类[M]. 北京: 海洋出版社, 1986: 213-214.]
- [3] Gao B Q, Liu P, Li J, et al. Heterosis of F₁ *Portunus trituberculatus* from matings and crosses among different

- geographical populations[J]. *Oceanologia Et Limnologia Sinica*, 2008, 39(3): 291–296.[高保全, 刘萍, 李健, 等. 三疣梭子蟹 (*Portunus trituberculatus*) 不同地理种群内自繁和种群间杂交子一代生长性状的比较[J]. 海洋与湖沼, 2008, 39(3): 291–296.]
- [4] Wu C W, Yu S C, Lv Y L. Swimming Crab Fishery Technology[M]. Shanghai: Shanghai Scientific and Technical Publishers, 1996: 28–31.[吴常文, 虞顺成, 吕永林. 梭子蟹渔业技术[M]. 上海: 上海科学技术出版社, 1996: 28–31.]
- [5] 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]. *Genet Select Evol*, 2002, 34(3): 275–306.
- [6] Lander E S. The new genomics: global views of biology[J]. *Science*, 1996, 274: 536–539.
- [7] Smith C T, Templin W D, Seeb J E, et al. Single nucleotide polymorphisms provide rapid and accurate estimates of the proportions of US and Canadian Chinook salmon caught in Yukon River fisheries[J]. *N Am J Fish Manag*, 2005, 25(3): 944–953.
- [8] Yamamoto K, Narukawa J, Kadono-Okuda K, et al. Construction of a single nucleotide polymorphism linkage map for the silkworm, *Bombyx mori*, based on bacterial artificial chromosome end sequences[J]. *Genetics*, 2006, 173(1): 151–161.
- [9] Rengmark A H, Slettan A, Skaala Ø, et al. Genetic variability in wild and farmed Atlantic salmon (*Salmo salar*) strains estimated by SNP and microsatellites[J]. *Aquaculture*, 2006, 253(1–4): 229–237.
- [10] Du Z Q, Ciobanu D C, Onteru S K, et al. A gene-based SNP linkage map for pacific white shrimp, *Litopenaeus vannamei*[J]. *Anim Genet*, 2010, 41(3): 286–294.
- [11] Stokstad E. To fight illegal fishing, forensic DNA gets local[J]. *Science*, 2010, 330(6010): 1468–1469.
- [12] Tao W J, Boulding E G. Associations between single nucleotide polymorphisms in candidate genes and growth rate in Arctic charr (*Salvelinus alpinus* L.)[J]. *Heredity*, 2003, 91(1): 60–69.
- [13] Huszar D, Lynch C A, Fairchild-Huntress V, et al. Targeted disruption of the melanocortin-4 receptor results in obesity in mice[J]. *Cell*, 1997, 88(1): 131–141.
- [14] Liu F P, Bai J J, Ye X, et al. Cloning of MC4R gene and study on the association between SNPs of MC4R and growth trait in tilapia (*Oreochromis niloticus*)[J]. *Journal of Fishery Sciences of China*, 2009, 16(6): 816–823.[刘福平, 白俊杰, 叶星, 等. 罗非鱼 MC4R 基因克隆及与其生长相关的 SNPs 位点[J]. 中国水产科学, 2009, 16(6): 816–823.]
- [15] Zhang X F, Yang J, Sun X W. Exploitation of SNPs related to growth traits of common carp (*Cyprinus carpio* L.) based on EST sequences[J]. *Chinese Journal of Fisheries*, 2009, 22(4): 1–7.[张晓峰, 杨晶, 孙效文. 基于 EST 序列的鲤鱼生长相关 SNP 发掘[J]. 水产学杂志, 2009, 22(4): 1–7.]
- [16] Yu Y, Wei J, Zhang X, et al. SNP discovery in the transcriptome of white Pacific shrimp *Litopenaeus vannamei* by next generation sequencing[J]. *PLoS ONE*, 2014, 9(1): e87218.
- [17] Ma H, Ma Q, Ma C, et al. Isolation and characterization of gene-derived single nucleotide polymorphism (SNP) markers in *Scylla paramamosain*[J]. *Biochem System Ecol*, 2011, 39(4): 419–424.
- [18] Li X, Cui Z, Liu Y, et al. Polymorphisms of anti-lipopolysaccharide factors in the swimming crab *Portunus trituberculatus* and their association with resistance/susceptibility to *Vibrio alginolyticus*[J]. *Fish Shellfish Immunol*, 2013, 34(6): 1560–1568.
- [19] Lv J J, Liu P, Gao B Q, et al. Transcriptome analysis of the *Portunus trituberculatus*: de novo assembly, growth-related gene identification and marker discovery[J]. *PloS ONE*, 2014, 9(4): e94055.
- [20] Jin Y L, Kong L F, Yu H, et al. Development, inheritance and evaluation of 55 novel single nucleotide polymorphism markers for parentage assignment in the Pacific oyster (*Crassostrea gigas*)[J]. *Genes Genom*, 2014, 36(2): 129–141.
- [21] Li X P, Liu P, Song X F. Construction on enriched microsatellite library and characterization of microsatellite markers from swimming crab[J]. *Journal of Fishery Sciences of China*, 2011, 18(1): 194–201.[李晓萍, 刘萍, 宋协法. 三疣梭子蟹微卫星富集文库的构建与群体遗传分析[J]. 中国水产科学, 2011, 18(1): 194–201.]
- [22] Aitken N, Smith S, Schwarz C, et al. Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach[J]. *Mol Ecol*, 2004, 13(6): 1423–1431.
- [23] Li W H, Sadler L A. Low nucleotide diversity in man[J]. *Genetics*, 1991, 129(2): 513–523.
- [24] Wang D G, Fan J B, Siao C J, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome[J]. *Science*, 1998, 280(5366): 1077–1082.
- [25] Lai E, Riley J, Purvis I, et al. A 4-Mb high-density single nucleotide polymorphism-based map around human APOE[J]. *Genomics*, 1998, 54(1): 31–38.
- [26] Nickerson D A, Taylor S L, Weiss K M, et al. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene[J]. *Nat Genet*, 1998, 19(3): 233–240.
- [27] Harding R M, Fullerton S M, Griffiths R C, et al. Archaic

- African and Asian lineages in the genetic ancestry of modern humans[J]. *Am J Human Genet*, 1997, 60(4): 772.
- [28] Taillon-Miller P, Gu Z, Li Q, et al. Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms[J]. *Genome Res*, 1998, 8(7): 748–754.
- [29] Curole J P, Hedgecock D. High frequency of SNPs in the Pacific oyster genome[C]. *Plant & Animal Genomes XIII Conference*, 2005.
- [30] Arias A, Freire R, Boudry P, et al. Single nucleotide polymorphism for population studies in the scallops *Aequipecten opercularis* and *Mimachlamys varia*[J]. *Conserv Genet*, 2009, 10(5): 1491–1495.
- [31] Morin P A, Aitken N C, Rubio-Cisneros N, et al. Characterization of 18 SNP markers for sperm whale (*Physeter macrocephalus*)[J]. *Mol Ecol Notes*, 2007, 7(4): 626–630.
- [32] Li S Z, Wan H R, Ji H Y, et al. SNP discovery based on CATS and genotyping in the finless porpoise (*Neophocaena phocaenoides*)[J]. *Conserv Genet*, 2009, 10(6): 2013–2019.
- [33] Wang J F. Large-scale development of SNP marker and its potential application in genetic breeding of *Crassostrea gigas*[D]. Qingdao: Institute of Oceanology, Chinese Academy of Sciences, 2012. [王家丰. 长牡蛎基因区 SNP 标记规模开发及其在遗传育种研究中的应用[D]. 青岛: 中国科学院海洋研究所, 2012.]
- [34] Li X P, Meng X H, Kong J, et al. Molecular cloning and characterization of a cathepsin B gene from the Chinese shrimp *Fenneropenaeus chinensis*[J]. *Fish Shellfish Immunol*, 2013, 35(5): 1604–1612.
- [35] Holliday R, Grigg G W. DNA methylation and mutation[J]. *Mutat Res-Fund Mol M*, 1993, 285(1): 61–67.
- [36] Petrov D A, Hartl D L. Patterns of nucleotide substitution in *Drosophila* and mammalian genomes[J]. *Proc Natl Acad Sci USA*, 1999, 96(4): 1475–1479.
- [37] Li W H, Sadler L A. Low nucleotide diversity in man[J]. *Genetics*, 1991, 129(2): 513–523.
- [38] Shen L X, Basilion J P, Stanton V P. Single-nucleotide polymorphisms can cause different structural folds of mRNA[J]. *Proc Natl Acad Sci USA*, 1999, 96(14): 7871–7876.
- [39] Komar A A. SNPs, silent but not invisible[J]. *Science*, 2007, 315(5811): 466–467.
- [40] Tran H T T, Takeshima Y, Surono A, et al. A G-to-A transition at the fifth position of intron-32 of the dystrophin gene inactivates a splice-donor site both *in vivo* and *in vitro*[J]. *Mol Genet Metab*, 2005, 85(3): 213–219.

Identifying SNP markers correlated with growth of swimming crab (*Portunus trituberculatus*) based on a comparative transcriptome

ZHANG Dening^{1,2}, LV Jianjian¹, LIU Ping¹, FENG Yanyan¹, GAO Baoquan¹, LI Jian¹

1. Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China;

2. College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China

Abstract: Comparative transcriptomics data have become one of the most important ways to study single nucleotide polymorphisms (SNP). We have identified 129 gene segments associated with growth. SNP is a third-generation molecular marker and is useful for breeding rapidly growing species, such as the swimming crab, *Portunus trituberculatus*. We developed an efficient method to select SNPs correlated with *P. trituberculatus* growth. Time-of-flight-mass spectrometry was used for SNP typing, and the chi-square test was used for the correlation analysis. Nineteen gene segments were amplified based on comparative transcriptomics, and a total of 17 387 bp DNA fragments were amplified. SNP frequency was estimated to be 0.43/100 bp of DNA sequence, the conversion mutation rate was 80%, and the transversion mutation rate was 20%. The conversion ratio was far greater than the transversion ratio to comply with the principle of “transition bias”. The G/T (C/A) rate was 20%, the A/T rate was 9.33%, the G/C rate was 10.67%, and the C/T (G/A), which is the most common gene mutation type, was 60%. The intron mutation frequency was 1.34/100 bp and that of exons was 0.17/100 bp, indicating that the exon mutation frequency was more conserved. Three SNP markers were significantly correlated with growth traits, and *P*-values of 10 SNPs were < 0.05. The chi-square test result revealed that *comp58070-R31* was closely related to growth traits (*P*<0.01). A multivariate analysis of variance showed that *comp58070-R31* was closely related to body weight, full carapace width, and carapace width (*P*<0.01). A significant correlation was detected between body length and body height (*P*<0.01). Loci *comp46623-F49* and *comp49193-R333* were significantly correlated with carapace width (*P*≤0.05). Observed heterozygosity of the *P. trituberculatus* SNPs was 0.162 8–0.833 3, and expected heterozygosity was 0.189 6–0.591 2, indicating less information content than that from microsatellite markers. These SNPs will be helpful for pedigree analysis, association studies, and marker-associated selection in this crab species.

Key words: *Portunus trituberculatus*; growth trait; transcriptome; SNP; correlation analysis

Corresponding author: LIU Ping. E-mail: liuping@ysfri.ac.cn