



基于液相芯片“黄海芯 1 号”55K SNP 信息开展凡纳滨对虾收获体重的 ssGWAS 和 FarmCPU 分析

夏岩^{1,2}, 孔杰², 王平³, 傅强², 罗坤², 陈宝龙²,
隋娟², 周浩², 孟宪红², 代平², 曹家旺²,
谭建², 强光峰², 刘绵宇², 栾生^{2*}

(1. 上海海洋大学水产与生命学院, 上海 201306;

2. 中国水产科学研究院黄海水产研究所, 海水养殖生物育种与可持续产出全国重点实验室,
青岛海洋科技中心海洋渔业科学与食物产出过程功能实验室, 山东 青岛 266071;

3. 海南中正水产科技有限公司, 海南 东方 572633)

摘要: 为鉴定出与凡纳滨对虾生长性状相关的 SNP 位点和功能基因, 本研究针对来自具有快速生长特性的选育和引进品系共 74 个家系进行生长和存活测试, 利用液相芯片“黄海芯 1 号”(55K SNP) 对其中 39 个家系个体进行高通量基因分型, 通过 ssGBLUP 方法估计了体重性状的基因组遗传参数, 并首次利用 ssGWAS 和 FarmCPU 方法解析了可能与其关联的 SNP 位点及功能基因。遗传参数估计结果显示, 收获体重的遗传力为 0.47, 表现为中高水平, 表明该群体具有较大的遗传改良潜力。ssGWAS 和 FarmCPU 分析结果显示, 两种方法分别筛选到了 2 个和 6 个全基因组水平显著关联 SNP 位点 ($P < 1.26 \times 10^{-6}$), 表型变异解释率为 2.71%~6.76%, 其中共有位点 1 个; 分别筛选到 9 个和 14 个全基因组水平潜在显著关联的 SNP 位点 ($P < 2.523 \times 10^{-5}$), 表型变异解释率为 1.33%~6.76%; QQ 图显示, FarmCPU 较 ssGWAS 能够更好地控制分析结果的假阳性。根据 7 个显著关联的 SNP 位点注释到 15 个候选基因, 与神经系统、肌肉发育和能量代谢等功能有关。其中基于两种方法共有显著关联位点注释到 2 个基因 *LOC113809777* 与 *LOC113809778*, 前者为促进糖原分解的水解酶, 后者作为转录调控因子, 与基因表达调控相关。研究表明, 基于液相芯片“黄海芯 1 号”55K SNP 信息, 利用 ssGWAS 和 FarmCPU 方法, 可以筛选出与凡纳滨对虾收获体重显著关联的 SNP 位点, 研究结果为解析生长性状的遗传基础提供了重要参考。

关键词: 凡纳滨对虾; ssGWAS; FarmCPU; 体重; 显著关联 SNP 位点

中图分类号: S 917.4

文献标志码: A

凡纳滨对虾 (*Litopenaeus vannamei*), 也被称为南美白对虾, 分布在美洲墨西哥、厄瓜多尔、秘鲁等国太平洋沿岸, 具有生长速率快、盐度适

应性广、抗病抗逆性强等优点, 适合低密度土塘生态养殖、高密度工厂化养殖等多种模式, 据 FAO 统计, 2021 年全球养殖产量达到 630 余万 t,

收稿日期: 2023-10-06 修回日期: 2024-01-05

资助项目: 国家重点研发计划 (2022YFD2400202); 中国水产科学研究院科技创新团队 (2020TD26); 现代农业产业技术体系专项 (CARS-48); 广东省“十四五”广东省农业科技创新十大主攻方向“揭榜挂帅”项目 (2022SDZG01); 泰山学者工程; 山东省重点研发计划 (农业良种工程) (2024LZGC038)

第一作者: 夏岩 (照片), 从事对虾遗传育种研究, E-mail: xiayan1207@126.com

通信作者: 栾生, 从事对虾遗传育种方面研究, E-mail: luansheng@ysfri.ac.cn



是当前世界单一经济价值最高的水产养殖种类^[1]。以规模化家系为基础的选择育种技术已经广泛应用于凡纳滨对虾育种项目, 生长速率得到了大幅度改良, 选育出了以泰国正大顶峰等为代表的快速生长良种, 简称为“快大系”良种。然而, 随着高密度工厂化养殖规模的不断扩大, 产业对苗种的生长速率提出了更高的要求。因此, 利用高通量基因组学信息, 筛选出与收获体重显著关联的 SNP 位点, 鉴定出关键基因及通路, 开展基因组选择等分子辅助育种, 是进一步提高养殖产量的重要途径。

全基因组关联分析 (genome-wide association study, GWAS) 旨在对全基因组范围内的遗传标记进行测序和统计分析, 以鉴定出影响复杂性状相关的遗传标记, 是解析收获体重等数量性状遗传基础的重要方法。利用基于混合线性模型 (mixed linear model, MLM) 的单位点 GWAS 分析方法, 王全超^[2]筛选出 47 个与凡纳滨对虾体重可能相关的 SNP 位点。孙坤^[3]利用单性状混合线性模型对凡纳滨对虾体重进行 GWAS 分析, 筛选出 33 个显著关联的 SNP 位点, 解释了 33.52%~53.92% 的表型变异。然而, 已有研究多采用单位点 GWAS 分析方法, 不但样本量较小 (<300 尾), 且未充分利用表型信息和系谱信息, 这可能会对 GWAS 分析结果的可信度产生较大影响。

在对虾规模化家系育种项目中, 通常会积累大量的多代表型和系谱信息。Garrick 等^[4]建立了基因分型个体的去回归育种值 (deregressed estimated breeding value, DEBV) 的计算方法, 同时将更多的表型和系谱信息并入模型估计分型个体的 DEBV。由于在构建 DEBV 的预测模型时, 可以纳入多个随机效应和固定效应, 能有效地剔除环境、性别等固定效应的影响, 将其作为因变量进行 GWAS 分析时比表型值更为准确。DEBV 代替表型已用于约克夏猪^[5]、荷斯坦奶牛^[6]和羊^[7]等重要生产性状的 GWAS 分析中, 鉴定出与社会效应、乳制品、乳腺炎症相关性状显著关联的 SNP 标记。

然而, 基于 DEBV 的 GWAS 分析需要分两步执行, 相对复杂。Aguilar 等^[8]提出了一步 GWAS (single-step GWAS, ssGWAS) 方法, 利用一步基因组最佳线性无偏预测法 (single-step genomic best linear unbiased prediction, ssGBLUP) 复合表型信息、系谱信息和基因型信息进行 GWAS 分析。蓝桉树球囊霉病的 ssGWAS 分析结果发现, 与 Single-

SNP 和 GBLUP-GWAS 相比, 该方法在 2、3 和 11 号染色体上鉴定到了更多的 SNP 位点 (66 vs. 13, 16)。为了消除在 MLM 中引入亲缘关系矩阵和群体结构协变量带来的假阴性问题^[9], Segura 等^[10]提出了多位点混合模型 (multi-locus mixed-model, MLMM) 方法。Liu 等^[9]提出了一种交替使用固定效应和随机效应模型优化多位点 GWAS 分析的 FarmCPU 方法 (fixed and random model circulating probability unification), 既完全消除了测试标记与亲缘关系的混淆, 又减少了分析结果的假阳性。针对拟南芥开花期性状的 GWAS 分析结果表明, FarmCPU 不仅有效控制了 *P* 值的膨胀, 还成功地鉴定出了与开花时间相关的新位点。目前使用 ssGWAS 和 FarmCPU 方法针对凡纳滨对虾体重性状开展 GWAS 分析的研究尚未见报道。

本实验以构建的凡纳滨对虾“快大系”育种核心群体多家系个体为研究对象, 利用液相芯片“黄海芯 1 号”(55K SNP) 对其进行高通量基因分型, 复合系谱和基因型信息构建 H 矩阵, 估计收获体重的遗传参数; 使用 ssGWAS 和 FarmCPU 方法针对收获体重执行 GWAS 分析, 筛选可能的关联位点。本研究结果将为凡纳滨对虾收获体重性状的分子辅助育种提供候选的关联标记和基因。

1 材料与方法

1.1 实验材料

本实验所使用的凡纳滨对虾来自海南中正水产科技有限公司, 包括 2021 年 9 月构建的选育系 G02 的 64 个家系, 以及引进的泰国 C 群体 10 个家系, 共计 74 个家系。其中, 育种核心群体自 2019 年开始构建, 奠基者群体主要包括具有快速生长特性的中正公司多代自繁群体、引进泰国 A 和 B 群体。

1.2 性状测试

生长速率和存活测试在直径为 4.5 m 的 2 个蓝色塑料圆桶 (B1、B2) 中进行, 水位高度 0.5 m, 水体体积约为 7.95 m³。2022 年 1 月, 对 74 个家系进行荧光标记, 每个家系每个桶标记 25 尾虾, 共计标记 3 700 尾, 分别放入 2 个圆桶中混合养殖测试, 养殖密度为 233 尾/m³。每天换水量 80%, 使用蛋白含量为 35% 的人工配合饲料进行投喂, 每天投喂 5 次, 每天投喂量为体重的 10%。养殖 57 d 后记录收获体重, 统计养殖存活率。

1.3 基因型鉴定

使用高通量 DNA 提取试剂盒对育种核心群体 91 尾亲本和 724 尾子代个体的肌肉组织进行 DNA 提取。其中, 子代 724 尾, 来自 74 个家系中生长速率最快的 39 个家系。利用 GenoBaits 探针捕获技术构建液相芯片“黄海芯 1 号”55K SNP 标记的靶向测序文库, 使用 MGISEQ-T7 测序平台进行测序。使用 PLINK^[11] 进行质控, 质控参数为“--geno 0.1 --maf 0.05 --mind 0.2”。对拟在遗传评估中使用的 815 尾亲本和子代个体的基因型信息进行质控, 保留了 808 尾个体和 45 649 个 SNP 标记。对拟在 GWAS 分析中使用的 724 尾子代个体的基因型信息进行质控 (参数同上), 保留了 717 尾个体和 45 697 个 SNP 标记, 进一步定位到 44 个连锁群后保留 39 640 个 SNP 标记。

1.4 遗传参数估计

利用 R 语言计算每个池中对虾收获体重的平均值、最小值、最大值、标准差以及变异系数。结合两代的基因型信息, 使用 ssGBLUP 模型估计体重性状各方差组分。采用 2 336 尾虾的表型数据, 其中 808 尾为基因分型个体。使用 BLUPF90^[12] 程序构建联合系谱-基因组关系的逆矩阵 (H^{-1}):

$$H^{-1} = A^{-1} + \begin{pmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{pmatrix} \quad (1)$$

式中, A^{-1} 为基于系谱信息构建的加性遗传相关矩阵的逆矩阵, A_{22}^{-1} 为基于系谱信息构建的基因分型个体的加性遗传相关矩阵的逆矩阵, G^{-1} 为基于基因组亲缘关系矩阵的逆矩阵, G 矩阵使用 Vanraden^[13] 建立的方法 1 构建:

$$G = \frac{ZZ'}{2 \sum P_i(1 - P_i)} \quad (2)$$

式中, Z 矩阵为中心化后的次等位基因含量矩阵, P_i 表示每个位点的次等位基因频率。

使用 ASReml-W V 4.2^[14] 估计凡纳滨对虾收获体重的遗传参数, 个体动物模型:

$$y_{ijk} = \mu + \text{Tank}_i + \text{Sex}_j + \beta b_{wk}(\text{Tank}_i : \text{Sex}_j) + a_k + e_{ijk} \quad (3)$$

式中, y_{ijk} 表示第 k 尾个体的收获体重 (g), μ 表示总体均值, Tank_i 表示第 i 个测试池固定效应, Sex_j 表示第 j 个性别固定效应, $b_{wk}(\text{Tank}_i : \text{Sex}_j)$ 表示拟合在第 i 个测试池与第 j 个性别交互固定效应内的第 k 尾个体的初始体重协变量, β 为初始体

重协变量的回归系数, a_k 表示第 k 尾个体的加性遗传效应 (随机效应), $a \sim N(0, H\sigma_a^2)$, σ_a^2 为加性遗传方差; e_{ijk} 表示第 k 尾个体的随机残差, $e \sim N(0, I\sigma_e^2)$, σ_e^2 是残差方差; 表型方差 $\sigma_p^2 = \sigma_a^2 + \sigma_e^2$; 遗传力 $h^2 = \sigma_a^2 / \sigma_p^2$ 。

1.5 DEBV 计算

利用系谱信息构建 A^{-1} , 使用“遗传参数估计”中的模型 1, 预测基因分型个体的 EBV。为了计算 DEBV, 首先根据以下公式计算个体 EBV 的预测准确性 (r):

$$r = \sqrt{1 - \frac{se(BLUP)^2}{\sigma_a^2}} \quad (4)$$

式中, $se(BLUP)$ 表示 EBV 的标准误, σ_a^2 是加性遗传方差。获得 r 后, 利用以下公式计算 DEBV^[4]:

$$DEBV = \frac{EBV}{r^2} \quad (5)$$

1.6 ssGWAS 分析

使用 BLUPF90^[12] 程序进行 ssGWAS 分析。主要步骤: 按照“遗传参数估计”中方法构建 H^{-1} ; 构建 ssGBLUP 的线性混合效应模型方程, 同模型 1; 分解和计算系数矩阵的稀疏逆矩阵; 求解 (β_μ) 线性混合效应模型方程; 计算育种值; 计算 SNP 效应值; 计算 SNP 的标准误; 计算 SNP 的 P 值。每个 SNP 的 P 值计算公式:

$$pval_i = 2 \left[1 - \varphi \left(\frac{\widehat{a}_i}{sd(\widehat{a}_i)} \right) \right] \quad (6)$$

式中, φ 为累积标准正态函数, \widehat{a}_i 为第 i 个 SNP 的效应值, $sd(\widehat{a}_i)$ 为 \widehat{a}_i 的标准差。

1.7 FarmCPU 分析

使用 R 包 rMVP^[15] 中 FarmCPU 方法进行多位点 GWAS 分析。固定效应模型:

$$y_i = PC_1 a_1 + PC_2 a_2 + PC_3 a_3 + M_{i1} b_1 + M_{i2} b_2 \cdots + M_{it} b_t + S_{ij} d_j + e_i \quad (7)$$

式中, y_i 为第 i 尾个体的 DEBV; PC_1 、 PC_2 、 PC_3 是主成分分析的前三列主成分, a_1 、 a_2 、 a_3 代表相应的效应值; M_{i1} , $M_{i2} \dots M_{it}$ 为 t 个加入到模型的可能关联位点的基因型, 第一次迭代的时候这部分为空, b_1 , $b_2 \dots b_t$ 为相应的效应值; S_{ij} 为第 i 尾个体第 j 个检测 SNP 标记的基因型; d_j 为相应的效应值; e_i 为残差, 服从平均数为 0、方差为

σ_e^2 的正态分布。

随机效应模型:

$$y_i = u_i + e_i \quad (8)$$

式中, y_i 为第 i 尾个体的 DEBV, u_i 为第 i 尾个体的总遗传效应, $u_i \sim N(0, 2K\sigma_a^2)$, K 为利用可能关联位点所构建的亲缘关系矩阵, σ_a^2 为未知遗传方差; e_i 为残差, 服从平均数为 0、方差为 σ_e^2 的正态分布。

使用 Bonferroni 校正方法确定全基因组水平显著阈值^[16]。全基因组水平显著阈值设置为 $P = \frac{0.05}{N_1} = \frac{0.05}{39\,640} = 1.26 \times 10^{-6}$, 全基因组水平潜在显著阈值设置为 $P = \frac{1}{N_1} = \frac{1}{39\,640} = 2.523 \times 10^{-5}$, 其中 N_1 是 SNP 标记数量。

1.8 候选基因

参考 Yu 等^[17]报道, 本研究检测位于显著 SNP 位点上下游 ± 40 kb 区域的基因。使用 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov>) 中的凡纳滨对虾基因组注释文件 (GCF_003789085.1_ASM378908v1_genomic.gff) 对两种方法得到的全基因

组水平显著 SNP 位点进行基因注释, 将其中“Gene”类型提取出来, 并使用 NCBI 进行查询, 在“Gene description”上找到相关基因描述。

本研究获得了中国水产科学研究院黄海水产研究所实验动物管理和使用伦理委员会批准, 实验过程中操作人员严格遵守中国水产科学研究院黄海水产研究所伦理规范, 并按照中国水产科学研究院黄海水产研究所伦理委员会制定的规章制度执行。

2 结果

2.1 描述性统计

收获体重观测值的描述性统计参数见表 1。全部 74 个家系 VIE 标记测试个体的总养殖存活率为 74.56%, 平均体重 15.68 g, 变异系数为 26.84%。其中, 在 B1 圆桶中个体的养殖存活率低于 B2 圆桶, 使其养殖密度低于 B2 圆桶, 因此导致在 B1 圆桶中个体的收获体重均值大于 B2 圆桶。基因分型个体来自生长速率较快的 39 个家系, 因此其收获体重均值高于全部个体的收获体重均值。

表 1 凡纳滨对虾收获体重的描述性统计参数

Tab. 1 Descriptive statistical parameters of harvested body weight in different populations of *L. vannamei*

群体 population	家系数量 no. of families	测试池号 tank ID	VIE标记 个体数/尾 no. of VIE-tagged individuals	收获个体数/尾 no. of harvest individuals	存活率/% survival rate	收获体重 weight				
						平均值/g mean	最小值/g minimum	最大值/g maximum	标准差/g standard deviation	变异系数/% coefficient of variation
所有测试个体 all tested individuals	74	B1	1850	1287	69.57	15.95	6.50	35.60	4.16	26.06
	74	B2	1850	1472	79.57	15.45	6.00	35.40	4.24	27.46
	74	全部 total	3700	2759	74.56	15.68	6.00	35.60	4.21	26.84
基因分型个体 genotyped individuals	38	B1	—	347	—	17.61	7.30	35.60	4.41	25.04
	39	B2	—	377	—	16.82	7.20	35.40	4.54	26.99
	39	全部 total	—	724	—	17.20	7.20	35.60	4.50	26.16

注: VIE. 可视嵌入性荧光标记; “—” 该项数据已经涵盖在上面“所有测试个体”中, 不需要重复列出。

Notes: VIE. visible implant elastomer; “—” the data for this item is already included in the above test subjects and does not need to be listed again.

2.2 方差组分和遗传力

收获体重的基因组遗传参数显示, 收获体重的表型方差 (σ_p^2) 为 11.47 ± 0.46 , 加性遗传方差 (σ_a^2) 为 5.29 ± 0.75 , 残差方差 (σ_e^2) 为 5.88 ± 0.42 , 遗传力 (h^2) 为 0.47 ± 0.05 , 表现为中高遗传力水平。

2.3 GWAS 分析结果

根据 Bonferroni 校正方法, 以 $P < 1.26 \times 10^{-6}$ 为全基因组水平显著阈值, ssGWAS 和 Farm-CPU 分别筛选到了 2 个和 6 个与体重显著关联的 SNP 位点, 对表型变异的解释率为 2.71%~6.76%, 其中 NW_020869456.1_73385 是共有位点 (表 2)。

表 2 凡纳滨对虾收获体重全基因组水平显著和潜在显著关联位点

Tab. 2 Significant and potentially significant genomic loci associated with harvest weight in *L. vannamei* at the whole genome level

方法 method	SNP名称 SNP_ID	连锁群 LG	物理位置 position	P值 P-value	方差解释率/% variance explained rate	
ssGWAS	NW_020868823.1_262749	15	45 830 727	1.57×10^{-7}	4.52	
	NW_020869050.1_828958	17	40 031 518	3.25×10^{-6}	2.83	
	NW_020868409.1_600910	1	46 571 994	1.02×10^{-5}	5.11	
	NW_020872939.1_365265	9	31 554 516	1.02×10^{-5}	2.26	
	NW_020868589.1_80129	31	2 929 316	1.14×10^{-5}	6.76	
	NW_020870488.1_36739	27	16 235 930	1.29×10^{-5}	6.47	
	NW_020868678.1_971813	29	16 654 473	1.58×10^{-5}	1.33	
	NW_020868678.1_963153	29	16 645 813	2.39×10^{-5}	2.31	
FarmCPU	NW_020868589.1_80129	31	2 929 316	9.52×10^{-8}	3.13	
	NW_020872939.1_365265	9	31 554 516	1.28×10^{-7}	3.18	
	NW_020872875.1_402024	36	30 733 600	2.03×10^{-7}	3.08	
	NW_020870683.1_49796	2	4 794 078	3.16×10^{-7}	2.92	
	NW_020869299.1_140699	31	34 152 939	6.68×10^{-7}	2.71	
	NW_020869724.1_535534	15	43 260 769	3.14×10^{-6}	2.46	
	NW_020870598.1_440389	4	17 921 466	3.99×10^{-6}	2.19	
	NW_020869886.1_173557	32	38 098 216	4.44×10^{-6}	2.21	
	NW_020870936.1_86686	34	4 366 596	6.76×10^{-6}	1.91	
	NW_020869666.1_1125718	42	35 666 845	8.57×10^{-6}	1.89	
	NW_020872498.1_16758	20	32 334 319	9.60×10^{-6}	2.10	
	NW_020868730.1_199761	27	9 117 447	1.75×10^{-5}	2.07	
	NW_020869923.1_247472	5	10 692 424	2.52×10^{-5}	1.92	
	ssGWAS	NW_020869456.1_73385	42	51 499 354	3.15×10^{-7}	6.76
					1.97×10^{-7}	2.97

注: 物理位置. 以连锁群上5'端第1个碱基作为初始位置。

Notes: Position. using the first base at the 5' end of the linkage group as the starting position.

以 $P < 2.523 \times 10^{-5}$ 为全基因组水平潜在显著阈值标准, ssGWAS 和 FarmCPU 分别筛选到了 9 个和 14 个与体重关联的 SNP 位点, 表型变异解释率为 1.33%~6.76%。ssGWAS 和 FarmCPU 分析的曼哈顿图和 QQ 图结果分别如图 1 和图 2 所示。

2.4 基因注释

针对 ssGWAS 和 FarmCPU 筛选到的 2 个和 6 个全基因组水平显著关联的 SNP 位点进行基因注释, 分别鉴定出 3 个和 14 个候选基因, 共计注释到 15 个候选基因(表 3), 其中, *LOC113809777* 与 *LOC113809778* 是共同注释到的基因, 这 2 个

基因的功能是 glycogen debranching enzyme-like 与 protein groucho-like。glycogen debranching enzyme 是一种促进糖原分解的水解酶^[18], protein groucho 是一种转录调控因子, 与基因表达调控相关^[19]。这些基因主要参与了发育、DNA 复制修复、能量代谢、细胞增殖、细胞分化等功能调控, 在机体分化发育方面可能发挥了重要作用。

3 讨论

3.1 收获体重的遗传参数估计

在选育过程中, 获得准确的遗传参数是开展

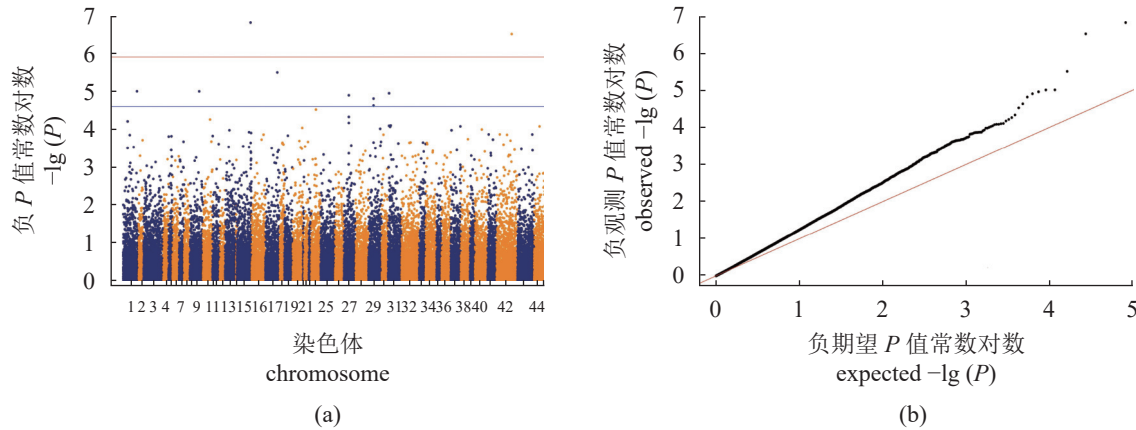


图 1 凡纳滨对虾收获体重 ssGWAS 分析的曼哈顿图 (a) 和 QQ 图 (b)

Fig. 1 Manhattan plot (a) and QQ plot (b) for genome-wide association study on harvest weight by ssGWAS in *L. vannamei*

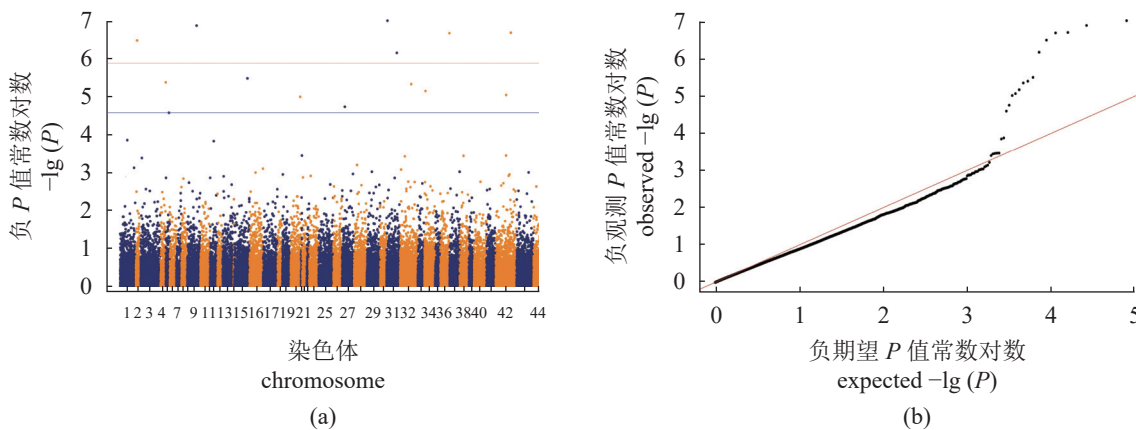


图 2 凡纳滨对虾收获体重 FarmCPU 分析的曼哈顿图 (a) 和 QQ 图 (b)

Fig. 2 Manhattan plot (a) and QQ plot (b) for genome-wide association study on harvest weight by FarmCPU in *L. vannamei*

育种方案规划、育种值估计等工作的重要前提条件^[20]。在本实验中, 利用 H 矩阵复合了亲本、子代的基因型和表型信息, 获得收获体重的遗传力估计值为 0.47。在栾生等^[21]研究中, 获得收获体重的遗传力估计值范围为 0.19~0.43。在 Sui 等^[22]的研究中, 综合两代系谱和表型数据获得收获体重的遗传力估计值为 0.335。与之相比, 本实验获得的遗传力估计值偏高, 推测其原因, 可能是由于受限于家系数据结构, 在模型中无法加入共同环境效应, 加性效应和共同环境效应混淆在一起所致。Ibarra 等^[23]的研究同样表明, 由于在模型中忽略了共同环境效应, 凡纳滨对虾收获体重的遗传力估计值高达 0.61。在 Dai 等^[24]针对剩余采食量的遗传评估研究中, 由于家系之间的遗传联系较弱, 在模型中没有包括共同环境效应, 导致其遗传力估计值被高估, 范围为 0.641~0.707。在

后续研究中, 降低不同家系单独养殖环境间的差异, 增加半同胞家系数量, 预期可以有效地避免遗传力被高估^[25]。

3.2 收获体重的单位点和多位点 GWAS 分析

对虾单尾价值低, 且基因组偏大, 重测序测序成本高, 较难开展大样本量的 GWAS 分析。因此在本研究中, 利用 ssGWAS 方法进一步复合 2 000 多尾未分型个体的表型值, 筛选出 2 个全基因组水平显著相关的 SNP 位点。孙坤^[3]利用 ssGWAS 方法, 通过复合 1 100 尾虾的收获体重和 292 尾个体的基因分型数据, 成功鉴定出 33 个全基因组水平显著相关的 SNP 位点。另外, 在对虹鳟 (*Oncorhynchus mykiss*) 立克次氏体病抗性^[26]、大西洋鲑 (*Salmo salar*) 和虹鳟海虱抗性^[27]等性状的研究中, 利用 ssGWAS 等方法也鉴定出了全基

表 3 凡纳滨对虾与收获体重性状显著相关的 SNPs 及候选基因

Tab. 3 Genome-wide significant SNPs and candidate genes associated with harvest weight in *L. vannamei*

方法 method	SNP名称 SNP_ID	起始 start	结束 end	基因 gene	基因描述 description
ssGWAS	NW_020868823.1_262749	284 193	308 088	LOC113803942	neurobeachin-like
FarmCPU	NW_020868589.1_80129	45 064	46 419	LOC113825361	piggyBac transposable element-derived protein 4-like
	NW_020868589.1_80129	79 934	80 881	LOC113825371	protein O-mannose kinase-like
	NW_020868589.1_80129	79 452	86 022	LOC113825223	SRR1-like protein
	NW_020868589.1_80129	92 397	116 466	LOC113825379	uncharacterized LOC113825379
	NW_020869299.1_140699	125 110	184 645	LOC113808389	myb-like protein X
	NW_020870683.1_49796	73 444	76 352	LOC113820817	serine/arginine repetitive matrix protein 5-like
	NW_020872875.1_402024	385 476	39 1803	LOC113802171	uncharacterized LOC113802171
	NW_020872875.1_402024	347 399	37 1422	LOC113802200	ATP-dependent RNA helicase A-like
	NW_020872875.1_402024	376 026	38 5098	LOC113802170	jmjC domain-containing protein 8-like
	NW_020872875.1_402024	415 206	485 007	LOC113802196	uncharacterized LOC113802196
	NW_020872875.1_402024	392 655	408 616	LOC113802178	ubiquitin carboxyl-terminal hydrolase 36-like
	NW_020872939.1_365265	270 761	438 536	LOC113802654	transcription factor collier-like
ssGWAS FarmCPU	NW_020869456.1_73385	25 597	55 233	LOC113809777	glycogen debranching enzyme-like
	NW_020869456.1_73385	61 337	70 053	LOC113809778	protein groucho-like

基因组水平显著相关的 SNP 位点。

但在 ssGWAS 等 MLM 模型中引入群体结构 Q 矩阵或主成分以及亲缘关系矩阵, 尽管可以降低 GWAS 分析结果的假阳性, 但是由于与待检测标记之间存在混淆问题, 因此也会导致 GWAS 分析结果的假阴性^[28]。FarmCPU 方法交替使用固定效应和随机效应模型, 可以有效避免群体结构以及亲缘关系矩阵与待检测标记的混淆问题, 降低分析结果的假阴性。在针对皱纹盘鲍 (*Haliotis discus hannai*) 的饲料效率^[29]、点带石斑鱼 (*Epinephelus coioides*) 氨的耐受性^[30] 以及乔奇拟鲈 (*Pseudocaranx georgianus*) 生长^[31] 等性状的研究中, 利用 FarmCPU 方法也鉴定出全基因组水平显著 SNP 位点。在本研究中, 实验利用 FarmCPU 方成功鉴定出 6 个全基因组水平显著 SNP 位点。

经过对 SNP 位点筛选和 QQ 图结果进行评估后发现, ssGWAS 的结果发生了偏离, *P* 值被高估, 而 FarmCPU 在分离与性状相关位点方面比 ssGWAS 表现得更出色。这种差异可能归因于两种方法在处理遗传信息时的主要区别: ssGWAS 通常只聚焦于单个位点的分析, 这可能导致其忽视了由多基因与环境因素引起的复杂交互作用, 使某些间接与性状相互关联的位点 *P* 值被高估。

相反, FarmCPU 的优势可能在于它能更全面地考虑和处理这些复杂的遗传信息, 从而更准确地识别出与性状显著关联的位点。Fu 等^[32] 使用 MLM、FarmCPU 方法针对凡纳滨对虾耐氨氮性状进行全基因组关联分析, 结果显示, 在全基因组水平显著阈值标准下, MLM 中未找到显著 SNP, 而 FarmCPU 找到 2 个显著 SNP。除此之外, Liu 等^[9] 分别使用 MLM 与 FarmCPU 针对拟南芥 (*Arabidopsis thaliana*) 的开花期性状进行全基因组关联分析, 结果显示, 在全基因组水平显著阈值周围, MLM 中没有找到显著 SNP, 而 FarmCPU 找到 5 个显著 SNP。这些研究都表明多位点模型比单位点模型表现更好。

3.3 收获体重的候选基因鉴定

在本次的关联分析中找到了 7 个显著性位点, 共注释到了 15 个基因, 大部分基因在凡纳滨对虾中尚未报道。其中, Neurobeachin^[33], piggyBac transposable element-derived protein 4^[34] 和 serine/arginine repetitive matrix protein 5 (SRRM5)^[35] 都与神经系统发育有关, 特别是 Neurobeachin 基因通过影响与进食相关的神经回路的活性, 进而对体重有重要影响, 强调了神经机制在体重控制

中的关键角色^[33]。同时, protein O-mannose kinase (POMK)^[36]与 transcription factor collier^[37]都与肌肉发育有关,尤其在生长过程中起着重要作用。例如, POMK 在斑马鱼 (*Danio rerio*) 胚胎和仔鱼的肌肉发育过程中都会表达^[36]。还有一些基因如 SRR1 protein、myb protein X^[38]、jmc domain-containing protein 8^[39]、ubiquitin carboxyl-terminal hydrolase 36^[40]和 ATP-dependent RNA helicase A^[41]等,它们与 DNA 复制、修复,细胞增殖、分化以及维持细胞功能等过程密切相关。值得注意的是, ssGWAS 与 FarmCPU 共同定位到了 glycogen debranching enzyme 和 protein groucho 这 2 个基因。其中, glycogen debranching enzyme 是一种促进糖原分解的水解酶^[18],与生长发育过程密切相关。

4 结论

本实验针对凡纳滨对虾“快大系”种质资源群体,利用液相育种芯片“黄海芯 1 号”(55K SNP)对部分家系个体进行高通量基因分型,通过 ssGB-LUP 方法复合基因型、系谱和表型信息,获得了体重性状的基因组遗传参数,遗传力表现为中高水平;首次使用 ssGWAS 和 FarmCPU 方法对收获体重进行 GWAS 分析,成功鉴定出 7 个全基因组水平显著 SNP 位点;基于显著 SNP 位点注释到候选基因 15 个,可能与神经系统、肌肉发育、能量代谢等功能有关。这些基因大部分在凡纳滨对虾中均未有报道,为开展生长性状的分子辅助选育提供了重要参考。

(作者声明本文无实际或潜在的利益冲突)

参考文献 (References):

- [1] 王兴强, 马姓, 董双林. 凡纳滨对虾生物学及养殖生态学进展 [J]. *海洋湖沼通报*, 2004(4): 94-100.
Wang X Q, Ma S, Dong S L. Studies on the biology and cultural ecology of *Litopenaeus vannamei*: a review[J]. *Transactions of Oceanology and Limnology*, 2004(4): 94-100 (in Chinese).
- [2] 王全超. 凡纳滨对虾生长和抗病性状的全基因组关联分析与基因组选择育种研究 [D]. 青岛: 中国科学院大学 (中国科学院海洋研究所), 2017.
Wang Q C. Genome-wide association study and genomic selection of growth and disease resistance traits in *Litopenaeus vannamei*[D]. Qingdao: The University of Chinese Academy of Sciences (Institute of Oceanology), 2017 (in Chinese).
- [3] 孙坤. 凡纳滨对虾重要经济性状的遗传评估及全基因组关联分析 [D]. 上海: 上海海洋大学, 2021.
Sun K. Genetic parameter evaluation and genome wide association analysis of important economic traits in *Litopenaeus vannamei*[D]. Shanghai: Shanghai Ocean University, 2021 (in Chinese).
- [4] Garrick D J, Taylor J F, Fernando R L. Deregressing estimated breeding values and weighting information for genomic regression analyses[J]. *Genetics Selection Evolution*, 2009, 41(1): 55.
- [5] Wu P X, Wang K, Zhou J, et al. A combined GWAS approach reveals key loci for socially-affected traits in Yorkshire pigs[J]. *Communications Biology*, 2021, 4(1): 891.
- [6] Cole J B, Wiggans G R, Ma L, et al. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows[J]. *BMC genomics*, 2011, 12(1): 408.
- [7] Rupp R, Senin P, Sarry J, et al. A point mutation in suppressor of cytokine signalling 2 (Socs2) increases the susceptibility to inflammation of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model[J]. *PLoS genetics*, 2015, 11(12): e1005629.
- [8] Aguilar I, Legarra A, Cardoso F, et al. Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle[J]. *Genetics Selection Evolution*, 2019, 51(1): 28.
- [9] Liu X L, Huang M, Fan B, et al. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies[J]. *PLoS genetics*, 2016, 12(2): e1005767.
- [10] Segura V, Vilhjálmsson B J, Platt A, et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations[J]. *Nature genetics*, 2012, 44(7): 825-830.
- [11] Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based
中国水产学会主办 sponsored by China Society of Fisheries

- linkage analyses[J]. *The American journal of human genetics*, 2007, 81(3): 559-575.
- [12] Lourenco D, Legarra A, Tsuruta S, *et al.* Single-step genomic evaluations from theory to practice: using SNP chips and sequence data in BLUPF90[J]. *Genes*, 2020, 11(7): 790.
- [13] Vanraden P M. Efficient methods to compute genomic predictions[J]. *Journal of dairy science*, 2008, 91(11): 4414-4423.
- [14] Gilmour A R, Gogel B J, Cullis B R, *et al.* User guide release 4.2 structural specification[R]. Hemel Hempstead: VSN International Ltd, Amberside House, Wood Lane, Paradise Industrial Estate, 2021.
- [15] Yin L L, Zhang H H, Tang Z S, *et al.* rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study [J]. *Genomics, Proteomics & Bioinformatics*, 2021, 19(4): 619-628.
- [16] Bland J M, Altman D G. Multiple significance tests: the Bonferroni method[J]. *Bmj*, 1995, 310(6973): 170.
- [17] Yu Y, Wang Q C, Zhang Q, *et al.* Genome scan for genomic regions and genes associated with growth trait in Pacific white shrimp *Litopenaeus vannamei*[J]. *Marine Biotechnology (NY)*, 2019, 21(3): 374-383.
- [18] Biscarini F, Bovenhuis H, Van Der Poel J, *et al.* Across-line SNP association study for direct and associative effects on feather damage in laying hens[J]. *Behavior Genetics*, 2010, 40(5): 715-727.
- [19] Jennings B H, Ish-Horowicz D. The Groucho/TLE/Grg family of transcriptional co-repressors[J]. *Genome Biology*, 2008, 9(1): 205.
- [20] Falconer D S. Introduction to Quantitative Genetics[M]. Harlow: Longman, 1996.
- [21] 栾生, 罗坤, 阮晓红, 等. 凡纳滨对虾 (*Litopenaeus vannamei*) 体重、存活性状的遗传参数和基因型与环境互作效应 [J]. *海洋与湖沼*, 2013, 44(2): 445-452.
- Luan S, Luo K, Ruan X H, *et al.* Genetic parameters and genotype by environment interaction for body weight and survival of Pacific white shrimp *Litopenaeus vannamei*[J]. *Oceanologia et Limnologia Sinica*, 2013, 44(2): 445-452 (in Chinese).
- [22] Sui J, Luan S, Luo K, *et al.* Genetic parameters and response to selection for harvest body weight of Pacific white shrimp, *Litopenaeus vannamei*[J]. *Aquaculture Research*, 2016, 47(9): 2795-2803.
- [23] Ibarra A M, Famula T R. Genotype by environment interaction for adult body weights of shrimp *Penaeus vannamei* when grown at low and high densities[J]. *Genetics Selection Evolution*, 2008, 40(5): 541.
- [24] Dai P, Luan S, Lu X, *et al.* Genetic assessment of residual feed intake as a feed efficiency trait in the Pacific white shrimp *Litopenaeus vannamei*[J]. *Genetics Selection Evolution*, 2017, 49(1): 61.
- [25] Liu J Y, Yang G L, Kong J, *et al.* Using single-step genomic best linear unbiased prediction to improve the efficiency of genetic evaluation on body weight in *Macrobrachium rosenbergii*[J]. *Aquaculture*, 2020, 528: 735577.
- [26] Barria A, Marín-Nahuelpi R, Cáceres P, *et al.* Single-step genome-wide association study for resistance to *Piscirickettsia salmonis* in rainbow trout (*Oncorhynchus mykiss*)[J]. *G3: Genes| Genomes| Genetics*, 2019, 9(11): 3833-3841.
- [27] Cáceres P, Barria A, Christensen K A, *et al.* Genome-scale comparative analysis for host resistance against sea lice between Atlantic salmon and rainbow trout[J]. *Scientific Reports*, 2021, 11(1): 13231.
- [28] 刘小磊. 一种交替运用固定效应和随机效应模型优化全基因组关联分析的算法开发 [D]. 武汉: 华中农业大学, 2016.
- Liu X L. Development of an iterative usage of fixed effect and random effect models for powerful and efficient genome-wide association studies[D]. Wuhan: Huazhong Agricultural University, 2016 (in Chinese).
- [29] Yu W C, Gong S H, Lu Y S, *et al.* Genome sequence-based genome-wide association study of feed efficiency in Pacific abalone[J]. *Aquaculture*, 2022, 561: 738630.
- [30] Xu T F, Zhang X H, Ruan Z Q, *et al.* Genome resequencing of the orange-spotted grouper (*Epinephelus coioides*) for a genome-wide association study on ammonia tolerance[J]. *Aquaculture*, 2019, 512: 734332.
- [31] Valenza-Troubat N, Montanari S, Ritchie P, *et al.* Unraveling the complex genetic basis of growth in New

- Zealand silver trevally (*Pseudocaranx georgianus*)[J]. *G3 Genes|Genomes|Genetics*, 2022, 12(3): jkac016.
- [32] Fu S, Liu J Y. Genome-wide association study identified genes associated with ammonia nitrogen tolerance in *Litopenaeus vannamei*[J]. *Frontiers in Genetics*, 2022, 13: 961009.
- [33] Olszewski P K, Rozman J, Jacobsson J A, *et al.* Neurobeachin, a regulator of synaptic protein targeting, is associated with body fat mass and feeding behavior in mice and body-mass index in humans[J]. *PLoS Genetics*, 2012, 8(3): e1002568.
- [34] Raskó T, Pande A, Radschek K, *et al.* A Novel gene controls a new structure: PiggyBac transposable element-derived 1, unique to mammals, controls mammal-specific neuronal paraspeckles[J]. *Molecular Biology and Evolution*, 2022, 39(10): msac175.
- [35] Li Y N, Donmez N, Sahinalp C, *et al.* SRRM4 drives neuroendocrine transdifferentiation of prostate adenocarcinoma under androgen receptor pathway inhibition[J]. *European Urology*, 2017, 71(1): 68-78.
- [36] Di Costanzo S, Balasubramanian A, Pond H L, *et al.* POMK mutations disrupt muscle development leading to a spectrum of neuromuscular presentations[J]. *Human Molecular Genetics*, 2014, 23(21): 5781-5792.
- [37] Carayon A, Bataillé L, Lebreton G, *et al.* Intrinsic control of muscle attachment sites matching[J]. *eLife*, 2020, 9: e57547.
- [38] George O L, Ness S A. Situational awareness: regulation of the myb transcription factor in differentiation, the cell cycle and oncogenesis[J]. *Cancers (Basel)*, 2014, 6(4): 2049-2071.
- [39] Li Y, Zhu S H, Yao J B, *et al.* Genome-wide characterization of the JmjC domain-containing histone demethylase gene family reveals GhJMJ24 and GhJMJ49 involving in somatic embryogenesis process in cotton[J]. *Frontiers in Molecular Biosciences*, 2022, 9: 888983.
- [40] Sharma A, Liu H D, Tobar-Tosse F, *et al.* Ubiquitin carboxyl-terminal hydrolases (UCHs): potential mediators for cancer and neurodegeneration[J]. *International journal of molecular sciences*, 2020, 21(11): 3910.
- [41] Nayak N R, Putnam A A, Addepalli B, *et al.* An *Arabidopsis* ATP-dependent, DEAD-box RNA helicase loses activity upon IsoAsp formation but is restored by protein isoaspartyl methyltransferase[J]. *The Plant Cell*, 2013, 25(7): 2573-2586.

ssGWAS and FarmCPU analysis of harvest body weight using the liquid chip "Yellow Sea Chip No.1" with 55K SNPs in *Litopenaeus vannamei*

XIA Yan^{1,2}, KONG Jie², WANG Ping³, FU Qiang², LUO Kun², CHEN Baolong², SUI Juan²,
ZHOU Hao², MENG Xianhong², DAI Ping², CAO Jiawang², TAN Jian²,
QIANG Guangfeng², LIU Mianyu², LUAN Sheng^{2*}

(1. College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China;

2. State Key Laboratory of Mariculture Biobreeding and Sustainable Goods,

Laboratory for Marine Fisheries Science and Food Production Processes,

Qingdao Marine Science and Technology Center, Yellow Sea Fisheries Research Institute,

Chinese Academy of Fishery Sciences, Qingdao 266071, China;

3. Hainan Zhongzheng Aquatic Science and Technology Co., Ltd., Dongfang 572633, China)

Abstract: Growth and survival tests were conducted on 74 families, comprising both selection and introduced populations with fast growth rates to identify single nucleotide polymorphisms (SNPs) and functional genes associated with growth traits in *Litopenaeus vannamei*. High-throughput genotyping of individuals from 39 families was carried out using the "Yellow Sea Chip No.1" liquid chip, which included 55K SNPs. Genomic genetic parameters for harvest body weight were estimated using the ssGBLUP method. Additionally, the ssGWAS and FarmCPU methods were employed for the first time to discover potentially associated SNPs and candidate genes. The estimation of genetic parameters revealed a heritability of 0.47 for harvest body weight, indicating a moderate to high potential for genetic improvement within the population. The ssGWAS and FarmCPU analyses identified two and six genome-wide significant SNPs ($P < 1.26 \times 10^{-6}$), respectively, explaining phenotypic variation ranging from 2.71% to 6.76%. A common SNP was also identified among them. Furthermore, both methods separately screened nine and 14 suggestive genome-wide SNPs ($P < 2.523 \times 10^{-5}$), explaining phenotypic variation ranging from 1.33% to 6.76%. QQ plots demonstrated that FarmCPU exhibited superior control over false positives compared to ssGWAS. Gene annotation of the seven significantly associated SNPs revealed 15 candidate genes related to functions such as the nervous system, muscle development, and energy metabolism. Notably, the common significantly associated SNP was annotated to two genes: *LOC113809777*, which encoded a hydrolase promoting glycogen hydrolysis, and *LOC113809778*, a transcriptional regulatory factor associated with gene expression regulation. In conclusion, utilizing the "Yellow Sea Chip No.1" liquid chip, the ssGWAS and FarmCPU methods facilitated the identification of significantly associated SNPs with harvest weight in *L. vannamei*, offering valuable insights into the genetic basis of growth traits.

Key words: *Litopenaeus vannamei*; ssGWAS; FarmCPU; body weight; significant SNP

Corresponding author: LUAN Sheng. E-mail: luansheng@ysfri.ac.cn

Funding projects: National Key Research and Development Program of China (2022YFD2400202); Science and Technology Innovation Team Project of Chinese Academy of Fishery Sciences (2020TD26); China Agriculture Research System (CARS-48); The Open Competition Program of Top Ten Critical Priorities of Agricultural Science and Technology Innovation for the 14th Five-Year Plan of Guangdong Province (2022SDZG01); Taishan Scholars Program; Shandong Province Key R & D Program (Agricultural Breeding Project) (2024LZGC038)