



基于不同密度 SNP 面板的凡纳滨对虾 AHPND 抗性基因组预测准确性分析

刘 杨^{1,2}, 栾 生², 刘绵宇², 李旭鹏², 孟宪红², 罗 坤², 隋 娟²,
谭 建², 代 平², 曹家旺², 陈宝龙², 孔 杰^{2*}

(1. 湖州师范学院生命科学学院, 浙江 湖州 313000;

2. 中国水产科学研究院黄海水产研究所, 农业农村部海洋渔业可持续发展重点实验室, 海洋科学与技术试点国家实验室海洋渔业科学与食物产出过程功能实验室, 山东 青岛 266071)

摘要: 为评估不同 SNP 标记密度对凡纳滨对虾 AHPND 抗性基因组预测准确性的影响, 本实验对 26 个全同胞家系进行 Vp_{AHPND} 侵染, 收集 686 尾个体的存活时间数据, 对其中 242 尾个体利用液相芯片“黄海芯 1 号”(55.0 K SNP) 进行基因分型, 基于 A、G 和 H 亲缘关系矩阵估计 Vp_{AHPND} 侵染后存活时间的遗传参数; 采用随机和等距抽取方式, 基于 55.0 K SNP 构建了 8 个低密度 SNP 面板 (40.0、30.0、20.0、10.0、5.0、1.0、0.5 和 0.1 K), 利用 GBLUP 和 ssGBLUP 等方法预测 Vp_{AHPND} 侵染后存活时间的基因组育种值, 利用交叉验证方法计算其预测准确性, 并与 BLUP 方法进行对比分析。遗传参数估计结果显示, Vp_{AHPND} 侵染后存活时间表现为高遗传力水平, 估计值为 0.68~0.79。在 55.0 K SNP 密度下, 针对 242 尾基因分型个体数据集 (G242), 利用 BLUP、GBLUP 和 ssGBLUP 方法获得的预测准确性分别为 0.424、0.450 和 0.452, GBLUP 和 ssGBLUP 比 BLUP 分别提升了 6.13% 和 6.60%; 针对 686 尾表型测定个体数据集 (P686), 利用 BLUP 和 ssGBLUP 方法获得的预测准确性分别为 0.510 和 0.535, 后者比前者提升了 4.90%。对于 8 个低密度 SNP 面板, 当 SNP 密度 ≥ 10.0 K 时, 基因组预测准确性变化幅度在 G242 和 P686 数据集中均较小 (1.1%~1.8%); 随着 SNP 密度自 10.0 K 不断降低, 基因组预测准确性在 2 个数据集中也不断降低, 其中 5.0 K 密度降幅为 0.6%~2.6%、1.0 K 密度降幅为 5.8%~11.0%、0.5 K 密度降幅为 11.4%~17.2%、0.1 K 密度降幅为 38.8%~41.6%。10.0 K 与 55.0 K SNP 密度间基因组亲缘系数、GEBV 的相关系数均高于 0.99, 表明利用 10.0 K SNP 面板可以准确地预测同胞个体间的亲缘关系及其 GEBV。研究表明, 使用 10.0 K SNP 面板对 Vp_{AHPND} 侵染后存活时间进行基因组遗传评估可以得到与 55.0 K SNP 芯片近似的预测准确性, 为低密度 SNP 分型芯片设计提供了参考。

关键词: 凡纳滨对虾; AHPND; 低密度 SNP 面板; 基因组预测准确性

中图分类号: S 961.6

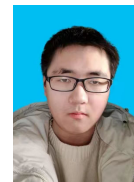
文献标志码: A

收稿日期: 2022-11-01 修回日期: 2022-12-15

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项目 (2020TD26); 海南省院士创新平台科研专项 (YSPTZX202104); 恒兴南美白对虾育种中心 (2021E05032); 现代农业产业技术体系专项 (CARS-48)

第一作者: 刘杨 (照片), 从事对虾遗传育种研究, E-mail: 740107180@qq.com

通信作者: 孔杰, 从事水产遗传育种研究, E-mail: kongjie@ysfri.ac.cn



凡纳滨对虾 (*Litopenaeus vannamei*), 俗称南美白对虾, 具有生长速率快、抗逆性强、出肉率高、盐度适应性广、适合多种养殖模式等特点, 是我国最重要的水产经济养殖物种之一^[1]。据《中国渔业统计年鉴》统计, 2021年我国凡纳滨对虾养殖产量超过 197 万 t^[2]。然而, 随着养殖规模的不断扩大, 病害威胁已成为制约我国凡纳滨对虾产业可持续发展的重要因素。其中, 急性肝胰腺坏死病 (acute hepatopancreatic necrosis disease, AHPND) 是威胁对虾养殖产业的主要疾病之一。该病是由带有毒素质粒的致急性肝胰腺坏死病副溶血性弧菌 (*Vibrio parahaemolyticus*, Vp_{AHPND}) 感染引起^[3-4], 多发于养殖初期, 传播迅速, 死亡率极高, 并且难以通过养殖管理等措施从根本上抑制弧菌的传播^[5-6]。因此, 通过选择育种手段培育具有 AHPND 抗性的新品种是解决该病对产业威胁的重要途径^[7-9]。目前, 基于系谱信息的规模化家系选择已经广泛应用于凡纳滨对虾育种项目^[10-11], 生长速率等性状得到了大幅度改良。但对于难以直接和无损测量的性状如抗病、肉质等, 仅能进行家系间选择, 因此只能利用一半的加性遗传变异^[12], 选择准确性和遗传增益受到较大限制。

基因组选择 (genomic selection, GS) 是 Meuwissen 等^[13] 在 2001 年提出的育种值 (estimated breeding value, EBV) 预测方法。与基于系谱信息的选择方法相比, GS 利用覆盖全基因组的单核苷酸多态性 (single nucleotide polymorphism, SNP) 标记信息, 预测未进行性状测定个体的基因组 EBV (genomic EBV, GEBV), 进一步提升选择准确性^[14]。基因组选择已开始应用在水产动物抗病性状的遗传改良研究中, Wang 等^[7] 对凡纳滨对虾 AHPND 抗性进行了基因组最佳线性无偏预测 (genomic best linear unbiased prediction, GBLUP) 评估, 其选择准确性高于基于系谱信息的 BLUP 方法。Vallejo 等^[15] 利用 BLUP 和单步 GBLUP (single-step GBLUP, ssGBLUP) 方法对虹鳟 (*Oncorhynchus mykiss*) 细菌性冷水病抗性进行遗传评估, ssGBLUP 的预测准确性比 BLUP 提升了约 85.3%。因此, 通过基因组选择方法可以进一步加快凡纳滨对虾 AHPND 抗性的选育速度。

与奶牛、猪等大型畜牧动物相比, 凡纳滨对虾个体经济价值较低、世代周期短, 基因分型成本相对较高^[16], 难以组建大参考群和候选群, 这严重限制了基因组选择在对虾育种中的应用进程。

因此, 针对不同物种的育种群体和目标性状, 优化 SNP 面板的密度, 是降低基因分型成本的一个重要方法。Al-Tobasei 等^[17] 针对虹鳟育种群体的鱼片质量与硬度, 将 SNP 密度由 50.0 K 降低至 1.0 K, ssGBLUP 的预测准确性仅降低了 30%。Wang 等^[18] 针对凡纳滨对虾育种群体生长性状, 将 SNP 密度由 32.0 K 降低至 3.2 K 时, 使用贝叶斯 A (BayesA) 及贝叶斯最小绝对收缩和选择算子 (Bayesian least absolute shrinkage selection operator, Bayesian LASSO) 的预测准确性几乎没有下降。Tsai 等^[19] 对养殖大西洋鲑 (*Salmo salar*) 海虱抗性的 GS 研究表明, 当标记密度由 33.0 K 降低至 5.0 K 时, 使用 GBLUP 方法的预测准确性几乎没有下降。Vallejo 等^[20] 对虹鳟细菌冷水病抗性的基因组评估表明, 当标记密度由 35.0 K 降低至 3.0 K 时, 使用贝叶斯 B (BayesB) 方法的预测准确性仅降低了 7%。然而, 针对凡纳滨对虾育种群体 AHPND 抗性, 基于不同密度 SNP 面板的基因组预测准确性研究较少报道。

本研究对 26 个凡纳滨对虾全同胞家系进行了 Vp_{AHPND} 侵染实验, 收集到了 686 尾个体的存活时间, 对其中 242 尾个体利用液相芯片“黄海芯 1 号”(55.0 K SNP) 进行基因分型, 利用 BLUP、GBLUP 及 ssGBLUP 模型对 AHPND 抗性进行遗传评估; 分析 SNP 筛选方式、SNP 面板密度对基因组预测准确性的影响, 为开展凡纳滨对虾 AHPND 抗性的低成本基因组选择育种提供关键技术支撑。

1 材料与方法

1.1 实验材料

实验用凡纳滨对虾家系材料来自于邦普种业科技有限公司。从 2021 年 8 月构建的家系中选取 26 个家系, 每个家系随机抽取 30 尾平均体长为 3 cm 的个体, 共 780 尾个体进行 AHPND 抗性测试。所有测试动物均按照中国水产科学研究院黄海水产研究所动物实验伦理审查委员会制定的实验动物福利伦理原则进行养殖管理及实验操作。

1.2 AHPND 性状测试

2021 年 10 月至 11 月, 在位于即墨鳌山卫镇的农业农村部海水养殖遗传育种中心开展 AHPND 抗性测试。对测试个体进行 Vp_{AHPND} 侵染

前, 确保其不携带白斑综合征病毒 (WSSV)、肝肠胞虫 (EHP)、 Vp_{AHPND} 、传染性皮下及造血组织坏死病毒 (IHHNV)、十足目虹彩病毒 1 (DIV1) 等病原。将每个家系 30 尾测试个体移入带有 10 L 海水 (盐度 30、 Vp_{AHPND} 浓度 1×10^7 CFU/mL) 的塑料小桶中浸浴 15 min 后, 转移至一个网箱 (26 cm \times 18 cm \times 16 cm) 中养殖, 该网箱放置在 100 L 海水 (盐度 30、 Vp_{AHPND} 浓度 1×10^5 CFU/mL) 的玻璃钢水槽中。每个玻璃钢水槽放置 6 个网箱。在转移至玻璃钢水槽 8 h 后开始观察个体的死亡情况, 每隔 4 小时观察 1 次, 记录死亡个体的存活时间、家系编号、水温等信息。每天换水 1 次, 每次换水量 100%, 每次换水后添加菌液致养殖水体内含 1×10^5 CFU/mL Vp_{AHPND} 。全部个体死亡后停止实验, 共收集到 686 尾个体的存活时间。将死亡个体的肌肉组织取样并浸泡在 95% 的乙醇中, 放置在 -20°C 的环境下保存。

1.3 基因分型

每个家系按照个体存活时间均匀抽取 8~10 尾虾, 共 243 尾虾进行基因分型。使用高通量 DNA 提取试剂盒获得 243 尾对虾肌肉组织的高质量 DNA, 利用 GenoBaits 探针捕获技术构建液相芯片“黄海芯 1 号”55.0 K SNP 标记的靶向测序文库, 使用 DNBSEQ-T7 测序平台进行测序。最终获得 243 尾对虾的 56 214 个 SNP 标记基因分型信息。使用 Plink V1.9 进行质控, 删除次要等位基因频率 (MAF) < 0.05 和 SNP 缺失率 > 0.1 的标记, 剔除样本标记缺失率 > 0.2 的个体, 最终得到 242 尾样本, 48 674 个 SNP 标记用于后续分析。根据凡纳滨对虾基因组图谱^[21-22], 共计将 42 445 个标记定位到 44 个连锁群上。

1.4 不同密度 SNP 面板构建

使用 R 语言结合 data.table 包, 分别按照随机和等距 2 种方法抽取 SNP 构建不同密度的 SNP 面板: (1) 随机抽取。从质控后的 SNP 标记中, 不返回随机抽取标记构建不同密度的 SNP 面板; (2) 等距抽取。根据拟构建的 SNP 面板密度、每个连锁群长度占全部连锁群体长度的比例等参数, 确定每个连锁群上所需抽取的 SNP 标记数量, 然后将每个连锁群等距离划分为同等数量的多个窗口, 在每个窗口内随机抽取 1 个 SNP 标记。如窗口内未抽取到 SNP 标记, 在后续窗口中抽取补齐; 未定位至连锁群上 SNP 标记, 随机抽取。利用上

述 2 种抽取方法, 分别构建 0.1 K、0.5 K、1.0 K、5.0 K、10.0 K、20.0 K、30.0 K、40.0 K 等 8 个 SNP 面板, 每个 SNP 面板重复抽取 10 次。

1.5 遗传参数评估

利用 BLUP、GBLUP 和 ssGBLUP 3 种模型估计 AHPND 抗性的遗传参数。利用平均信息约束最大似然法, 通过 ASReml-R V4.1 包估计 Vp_{AHPND} 感染后存活时间的遗传参数^[23]。育种分析模型:

$$y_i = \mu + a_i + e_i$$

式中, y_i 表示第 i 尾个体的存活时间 (h), μ 表示总体均值, a_i 表示第 i 尾个体的加性遗传效应, $a \sim N(0, A\sigma_a^2)$, $N(0, G\sigma_a^2)$ 或 $N(0, H\sigma_a^2)$, 其中 A 为加性遗传相关矩阵, G 为基因组亲缘关系矩阵, H 为复合系谱信息和基因组信息的亲缘关系矩阵, e_i 表示第 i 尾个体的随机残差。利用 ASReml-R 软件包中的 ainverse 函数构建 A 矩阵; 利用 BLUPF90 软件包中的 preGSf90 模块构建 G 矩阵和 H 矩阵。遗传力计算公式:

$$h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2)$$

式中, h^2 为遗传力, σ_a^2 为加性遗传方差, σ_e^2 为残差。

1.6 EBV(GEBV) 的预测准确性计算

使用随机交叉验证方法, 针对 242 个基因分型个体数据集 (G242) 和 686 个表型测定个体数据集 (P686), 分析 BLUP、GBLUP、ssGBLUP 方法预测 EBV(GEBV) 的准确性。P686 数据集不但包括了 242 个基因分型个体的表型, 还包括了 444 个未分型个体的表型。在 G242 数据集中, 将基因分型个体随机均分成 5 组, 将其中 1 组的存活时间设为缺失作为验证群体, 其余 4 组数据作为参考群体预测验证群体的 EBV (GEBV)。在 P686 数据集中, 将 242 尾基因分型个体随机均分成 5 组, 将其中 1 组的存活时间设为缺失作为验证群体, 其余 4 组数据加上 444 尾未分型个体的表型数据作为参考群体, 预测验证群体的 EBV (GEBV)。预测准确性定义为验证群体预测 EBV (GEBV) 与其表型值之间的 Pearson 相关系数。随机交叉验证重复 10 次, 取平均值作为预测准确性。

1.7 不同密度 SNP 面板间基因组亲缘系数及 GEBV 的相关系数计算

为了分析 SNP 密度对基因组亲缘系数及 GEBV 的影响程度, 针对 8 个低密度 SNP 面板,

计算基于 G 矩阵的 242 个基因分型个体间的基因组亲缘系数, 预测基因分型个体的 GEBV, 然后计算 8 个低密度 SNP 面板与 55.0 K SNP 芯片数据间基因组亲缘关系、GEBV 的 Pearson 相关系数。

2 结果

2.1 表型数据描述性统计参数

凡纳滨对虾测试个体 Vp_{AHPND} 侵染后存活时间的描述性统计参数见表 1。全部表型测定个体数量为 686 尾, 平均存活时间为 456.49 h, 变异系数为 75.38%; 26 个家系平均存活时间为 470.03 h, 变异系数为 42.02%。基因型分体个体数量为 243 尾, 平均存活时间为 485.55 h, 变异系数为 70.73%; 26 个家系平均存活时间为 488.16 h, 变异系数为 42.69%。家系间存活时间变异幅度远小于个体间的变异幅度。

2.2 方差组分和遗传参数

基于不同亲缘关系矩阵获得的凡纳滨对虾 Vp_{AHPND} 侵染后存活时间的方差组分与遗传力估计值(表 2)。其中, 基于 G 矩阵获得 G242 数据集 Vp_{AHPND} 侵染后存活时间的遗传力估计值 (0.68) 最低; 基于 H 矩阵获得 P686 数据集 Vp_{AHPND} 侵染后存活时间的遗传力估计值 (0.75) 居中; 基于 A

矩阵获得 P686 数据集 Vp_{AHPND} 侵染后存活时间的遗传力估计值 (0.79) 最高。

2.3 基于 55.0 K SNP 芯片和较低密度 SNP 面板获得的 EBV(GEBV) 预测准确性

基于 55.0 K SNP 芯片获得的 EBV(GEBV) 预测准确性 利用 GBLUP 和 ssGBLUP 方法, 通过交叉验证获得凡纳滨对虾基因分型个体和表型测定个体 Vp_{AHPND} 侵染后存活时间 GEBV 的预测准确性 (图 1, 图 2)。针对 G242 数据集, 利用 BLUP、GBLUP 和 ssGBLUP 方法获得的预测准确性分别为 0.424、0.450 和 0.452, GBLUP 和 ssGBLUP 比 BLUP 提升了 6.13% 和 6.60%。针对 P686 数据集, 利用 BLUP 和 ssGBLUP 方法获得的预测准确性分别为 0.510 和 0.535, 后者比前者提升了 4.90%。在 G242 和 P686 数据集中, 利用基因组信息均可进一步提升 EBV(GEBV) 的预测准确性。

与 G242 数据集相比, 使用 P686 数据集后, BLUP 和 ssGBLUP 的预测准确性分别提升了 20.28% 和 18.36%。这表明, 在 242 尾基因分型个体表型信息基础上加入未分型个体的表型信息, 能够进一步提升预测准确性。

基于较低密度 SNP 面板获得的 EBV (GEBV) 预测准确性 利用 GBLUP 和 ssGBLUP 方法,

表 1 凡纳滨对虾 Vp_{AHPND} 侵染后存活时间的描述性统计参数

Tab. 1 Descriptive statistical parameters of survival time after Vp_{AHPND} infection in *L. vannamei*

数据集 data set	个体/家系 individual/family	数量/尾 no.	存活时间/h survival time				
			均值 mean	最小值 min	最大值 max	标准差 standard deviation	变异系数/% coefficient of variation
表型测定个体 phenotypic individuals	个体 individual	686	456.49	8.00	1 010	344.11	75.38
	家系 family	26	470.03	56.73	803.29	197.50	42.02
基因分型个体 genotyped individuals	个体 individual	243	485.55	8.00	1 010	343.42	70.73
	家系 family	26	488.16	86.90	909.56	208.40	42.69

表 2 基于不同亲缘关系矩阵的凡纳滨对虾 Vp_{AHPND} 侵染后存活时间的方差组分与遗传力

Tab. 2 Variance components and heritability of survival time after Vp_{AHPND} infection in *L. vannamei* based on different relationship matrices

亲缘关系矩阵 relationship matrix	加性遗传方差 additive genetic variance	残差方差 residual variance	表型方差 phenotypic variance	遗传力 heritability
A	107 192.19	28 021.66	135 213.85	0.79±0.15
G	87 047.53	40 164.72	127 212.25	0.68±0.13
H	98 451.91	32 323.67	130 775.58	0.75±0.11

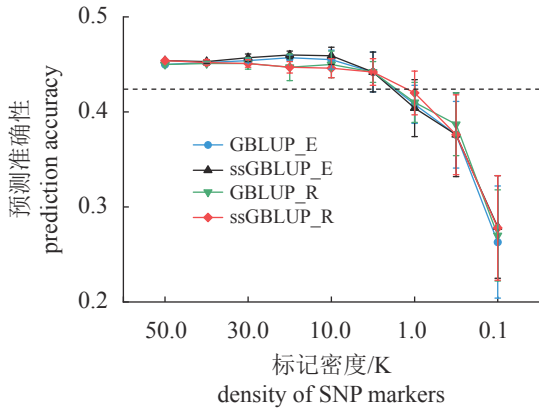


图 1 基于不同密度 SNP 面板获得的凡纳滨对虾 242 尾基因分型个体 Vp_{AHPND} 侵染后存活时间 EBV(GEBV) 的预测准确性

E 表示等距抽取法, R 表示随机抽取法, 虚线表示 PBLUP 模型的预测准确性; 下同

Fig. 1 Prediction accuracies of EBV (GEBV) of survival time after Vp_{AHPND} infection in 242 genotyped individuals of *L. vannamei* using SNP panels with different densities

E refers to select SNPs at a fixed genome interval, R refers to selecting SNPs at random, the dash line refers to the prediction accuracy of the PBLUP model; the same below

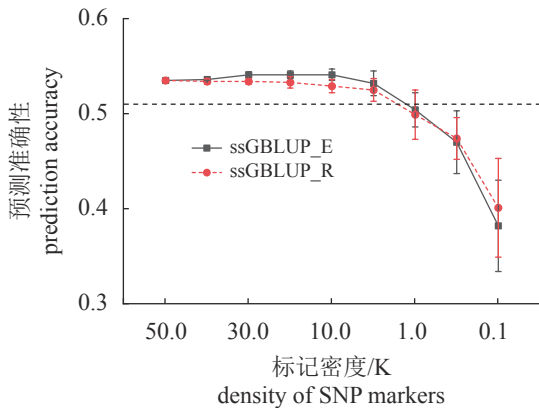


图 2 基于不同密度 SNP 面板获得的凡纳滨对虾 686 尾表型测定个体 Vp_{AHPND} 侵染后存活时间 EBV(GEBV) 的预测准确性

Fig. 2 Prediction accuracies for EBV (GEBV) for survival time after Vp_{AHPND} infection in 686 individuals of *L. vannamei* with phenotypes using SNP panels with different densities

通过交叉验证获得在较低 SNP 密度下凡纳滨对虾 G242 数据集 Vp_{AHPND} 侵染后存活时间 GEBV 的预测准确性 (图 1)。无论是随机抽取还是等距抽取方法, 当 SNP 标记密度大于等于 10.0 K 时, 预测准确性变化幅度较小, 为 1.1%~1.8%。随着 SNP 标

记密度自 10.0 K 不断减小, 预测准确性也开始不断降低, 随机抽取 (等距抽取) 方法的降低幅度分别为 1.8%~2.6% (1.8%~2.6%)、8.8%~7.4% (9.3%~11.0%)、14.0%~17.2% (16.4%~17.2%)、40.0%~38.8% (41.6%~38.5%)。

利用 ssGBLUP 方法, 通过交叉验证获得在较低 SNP 密度下凡纳滨对虾 P686 数据集 Vp_{AHPND} 侵染后存活时间 EBV(GEBV) 的预测准确性 (图 2)。无论是随机抽取还是等距抽取方法, 当 SNP 标记密度大于等于 10.0 K 时, 预测准确性变化幅度较小, 为 1.1%~1.8%。随着 SNP 标记密度自 10.0 K 不断减小, 预测准确性也开始不断降低, 随机抽取 (等距抽取) 方法的降低幅度分别为 0.6% (1.9%)、5.8% (6.7%)、12.1% (11.4%)、25.6% (25.0%)。随着 SNP 密度的不断降低, 预测准确性的变化趋势同基因分型个体的变化趋势。此外, 由于 ssGBLUP 方法可以利用更多的未分型个体的表型数据, 随着 SNP 密度的不断降低, 利用 P686 数据集获得的 GEBV 预测准确性的降低幅度 (0.2%~28.6%) 要小于利用 G242 数据集获得的 GEBV 预测准确性的降低幅度 (0.7%~38.6%)。

2.4 基于不同密度 SNP 面板获得的基因组亲缘系数和 GEBV 的相关性

基因组亲缘系数在 8 个密度 SNP 面板与 55.0 K SNP 芯片之间的相关系数见表 3。无论是随机抽取还是等距抽取方法, 当 SNP 标记密度大于 10.0 K 时, 相关系数变化非常小, 降低幅度仅为 0.1%。随着 SNP 标记密度自 10.0 K 不断降低, 相关系数也开始不断降低, 随机抽取 (等距抽取) 方法的降低幅度分别为 0.4% (0.3%)、0.7% (0.7%)、3.8% (3.9%)、7.6% (7.7%)、27% (29%)。

GEBV 在 8 个密度 SNP 面板与 55.0 K SNP 芯片之间的相关系数见表 4。无论是随机抽取还是等距抽取方法, 当 SNP 标记密度大于 10.0 K 时, 相关系数变化非常小, 降低幅度仅为 0.1%。随着 SNP 标记密度自 10.0 K 不断降低, 相关系数也开始不断降低, 随机抽取 (等距抽取) 方法的降低幅度分别为 0.2% (0.2%)、0.4% (0.4%)、3.1% (2.0%)、4.9% (4.3%)、17.2% (17.9%)。

3 讨论

3.1 方差组分与遗传力

在本研究中, 基于 H 矩阵获得 P686 数据集

表 3 不同密度 SNP 面板与 55.0 K SNP 芯片之间的基因组亲缘系数的相关系数

Tab. 3 Correlation coefficients of genomic relationship coefficients between SNP panels with different densities and the 55.0 K SNP chip

标记密度/K marker density	SNP抽取方式 method for selecting SNPs	
	随机抽取 selecting SNPs at random	等距抽取 selecting SNPs at a fixed genome interval
40.0	0.999±7.858 9×10 ⁻⁶	0.999±6.436 1×10 ⁻⁶
30.0	0.999±1.286 6×10 ⁻⁵	0.999±2.262 3×10 ⁻⁵
20.0	0.999±5.509 3×10 ⁻⁵	0.999±5.257 7×10 ⁻⁵
10.0	0.996±0.000 2	0.997±0.000 1
5.0	0.993±0.000 3	0.993±0.000 3
1.0	0.962±0.001 6	0.961±0.001 3
0.5	0.924±0.004 0	0.923±0.003 3
0.1	0.730±0.018 0	0.710±0.016 9

表 4 不同密度 SNP 面板与 55.0 K SNP 芯片之间的 GEBV 相关系数

Tab. 4 Correlation coefficients of GEBV between SNP panels with different densities and the 55.0 K SNP chip

标记密度/K marker density	SNP抽取方式 method for selecting SNPs	
	随机抽取 selecting SNPs at random	等距抽取 selecting SNPs at a fixed genome interval
40.0	0.999±7.441 0×10 ⁻⁶	0.999±1.490 2 ⁻⁵
30.0	0.999±4.603 8×10 ⁻⁵	0.999±0.000 1
20.0	0.999±0.000 2	0.999±0.000 3
10.0	0.998±0.001 1	0.998±0.002 8
5.0	0.996±0.001 1	0.996±0.003 4
1.0	0.979±0.006 5	0.980±0.016 0
0.5	0.961±0.014 6	0.957±0.017 0
0.1	0.828±0.030 2	0.821±0.038 8

$V_{P_{AHPND}}$ 侵染后存活时间的遗传力估计值 (0.75) 相比 A 矩阵 (0.79) 降低了 5.1%，但预测准确性提升了 6.6%，表明与 A 矩阵相比，利用包括基因组信息的 H 矩阵能够精准估计个体间的亲缘关系，考虑全同胞个体间的孟德尔抽样，从而更准确地估计遗传力^[24]。此外，基于 G 矩阵的遗传力估计值低于基于 A 和 H 矩阵的遗传力估计值，主要是因为前者仅针对 232 尾基因型个体，而后者针对 686 尾表型测定个体 (包括了 454 尾未基因分型个体)，因此其遗传变异更为丰富。本研究获得的 $V_{P_{AHPND}}$ 侵染后存活时间的遗传力估计值 (0.68~0.79) 明显高于其他已报道的研究结果 (0.24±0.09,

0.26±0.10)^[25-27]，具有群体特异性。本研究所评估的群体是由 AHPND 抗性存在较大差异的 24 个家系组成，分为高、中、低 3 个抗性组，是根据亲本的 $V_{P_{AHPND}}$ 侵染后存活时间的预测 GEBV 设计优化配种方案产生的^[28]。因此，本研究获得的高遗传力估计值，主要是由于家系间 AHPND 抗性的显著差异导致的。本研究及已有研究报道均表明，AHPND 抗性具有中高遗传力水平，因此通过选择育种手段可以有效改良育种群体的抗性^[7]。在对虾其他抗病性状的研究中，Argue 等^[26]的研究表明对虾抗 TSV 存活时间性状的遗传力为 0.28±0.14；孙坤等^[29]的研究表明对虾抗 WSSV 存活时间性状的遗传力在 0.10~0.14。综上表明多数对虾抗病性状均具有较大的遗传改良潜力。

3.2 不同遗传评估方法的预测准确性

在本研究中，GBLUP 和 ssGBLUP 的预测准确性比 BLUP 分别提升了 6.1% 和 4.9%~6.6%。Wang 等^[7]对凡纳滨对虾 AHPND 抗性 (存活状态) 进行了遗传评估，GBLUP 预测准确性比 BLUP 高 6.3%；Yoshida 等^[30]利用 ssGBLUP 和 BLUP 评估了虹鳟传染性胰脏坏死病毒抗性，前者的预测准确性比后者提升了 7%，与本研究结果相似。然而，Vallejo 等^[15]对虹鳟细菌性冷水病的遗传评估研究表明，ssGBLUP 的预测准确性比 BLUP 高了 83.3%~85.3%。斑点叉尾鲷 (*Ictalurus punctatus*) 和大西洋鲑等收获体重的基因组选择研究表明，ssGBLUP 的预测准确性比 BLUP 提高 19.00%~42.42%^[31-32]。参考群体大小可能是影响选择准确性增加幅度的一个主要因素。受限于基因分型成本，在本研究中使用的参考群体较小，仅为 194 尾；而在虹鳟细菌性冷水病的遗传评估研究中，构建的参考群体大小为 1 473 尾，后者是前者的 7.59 倍。当参考群个体数量降低至大约 500 尾时，预测准确性降低了 20.90%^[15]。

3.3 不同密度 SNP 面板的预测准确性

本研究表明，无论使用随机抽取法还是等距抽取法，当 SNP 标记密度保持在 10.0 K 及以上时，GBLUP 及 ssGBLUP 的预测准确性未发生明显变化；当 SNP 标记密度自 10.0 K 分梯度降低至 0.1 K 时，逐渐出现较为明显的降幅，最高达到 41.6%。陈美佳^[16]针对凡纳滨对虾收获体重的遗传评估研究表明，仅使用 10.0 K SNP 面板就可以获得与基因组重测序水平相当的预测准确性，这一结论与本

研究结果相同。在大西洋鲑海虱抗性的基因组选择研究中, 当标记密度从 33.0 K 降低至 5.0 K 时, GBLUP 方法的预测准确性几乎没有变化^[19]。在虹鳟细菌冷水病抗性的基因组遗传评估研究中, 当标记密度由 35.0 K 降低至 3.0 K 时, BayesB 方法的预测准确性仅降低了 7%^[20]。在奶牛的基因组遗传评估研究中, 研究学者发现当 SNP 标记密度达到一定程度后, 增加更多的标记并不能取得更大的遗传进展^[33-35]。分析其原因, 推测主要有以下两点: 第一, 利用当前标记密度已经可以准确地地区分同胞个体间的亲缘关系, 如在本实验中, 基于 10.0 K 与 55.0 K SNP 芯片获得的 2 组基因组亲缘系数间的相关系数高达 $0.996\ 0 \pm 0.000\ 2$; 第二, 利用当前标记密度已经可以准确地预测同胞个体的 GEBV, 如在实验中, 基于 10.0 K 与 55.0 K SNP 芯片获得的 2 组 GEBV 间的相关系数高达 $0.998\ 0 \pm 0.001\ 1$ 。此外, 为了进一步降低成本, 可以将 SNP 密度降低至 5.0 K, 此时 GBLUP 的预测准确性仍然高于 BLUP 方法。

4 结论

在凡纳滨对虾 AHPND 抗性的遗传评估中加入覆盖基因组的液相芯片“黄海芯 1 号”55.0 K SNP 标记信息, 可以进一步增加选择的准确性。当 SNP 标记密度保持在 10.0 K 及以上时, 基因组预测的准确性未发生明显变化; 当 SNP 标记密度低于 10.0 K 时, 基因组预测的准确性随着 SNP 密度的降低而降低。利用 10.0 K SNP 面板可以较为准确地预测同胞个体间的亲缘关系及其 GEBV, 因此, 利用 10.0 K SNP 面板对 *Vp_{AHPND}* 侵染后存活时间进行基因组遗传评估, 可以得到与 55.0 K SNP 芯片近似的预测准确性。

(作者声明本文无实际或潜在的利益冲突)

参考文献 (References):

- [1] 张伟权. 世界重要养殖品种——南美白对虾生物学简介[J]. 海洋科学, 1990, 14(3): 69-73.
Zhang W Q. Biological profile of important world aquaculture species-*Penaeus vannamei*[J]. Marine Science, 1990, 14(3): 69-73 (in Chinese).
- [2] 农业农村部渔业渔政管理局, 全国水产技术推广总站, 中国水产学会. 中国渔业统计年鉴 2022[M]. 北京: 中国农业出版社, 2022.
- [3] Soto-Rodriguez S A, Gomez-Gil B, Lozano-Olvera R, et al. Field and experimental evidence of *Vibrio parahaemolyticus* as the causative agent of acute hepatopancreatic necrosis disease of cultured shrimp (*Litopenaeus vannamei*) in Northwestern Mexico[J]. *Applied and Environmental Microbiology*, 2015, 81(5): 1689-1699.
- [4] Han J E, Tang K F, Tran L H, et al. *Photobacterium* insect-related (Pir) toxin-like genes in a plasmid of *Vibrio parahaemolyticus*, the causative agent of acute hepatopancreatic necrosis disease (AHPND) of shrimp[J]. *Diseases of Aquatic Organisms*, 2015, 113: 33-40.
- [5] 郝景伟, 高保全, 王崇, 等. 致急性肝胰腺坏死病副溶血弧菌(*Vp_{AHPND}*)自然感染三疣梭子蟹[J]. 水产学报, 2019, 43(7): 1647-1660.
Hao J W, Gao B Q, Wang C, et al. Natural infection of *Portunus trituberculatus* with acute hepatopancreas necrosis disease causing by *Vibrio parahaemolyticus* (*Vp_{AHPND}*)[J]. *Journal of Fisheries of China*, 2019, 43(7): 1647-1660 (in Chinese).
- [6] 孙明玉, 冯博, 张昭寰, 等. 引起凡纳滨对虾急性肝胰腺坏死病的副溶血弧菌 MLST 新序列型[J]. 水产学报, 2018, 42(3): 410-418.
Sun M Y, Feng B, Zhang Z H, et al. New sequence type isolates of AHPND-causing *Vibrio parahaemolyticus* from *Litopenaeus vannamei* by multilocus sequence typing[J]. *Journal of Fisheries of China*, 2018, 42(3): 410-418 (in Chinese).
- [7] Wang Q C, Yu Y, Zhang Q, et al. Evaluation on the genomic selection in *Litopenaeus vannamei* for the resistance against *Vibrio parahaemolyticus*[J]. *Aquaculture*, 2019, 505: 212-216.
- [8] Gjedrem T, Robinson N. Advances by selective breeding for aquatic species: a review[J]. *Agricultural Sciences*, 2014, 5(12): 1152-1158.
- [9] Gjedrem T, Robinson N, Rye M. The importance of selective breeding in aquaculture to meet future demands for animal protein: a review[J]. *Aquaculture*, 2012, 350-353: 117-129.
- [10] 唐扬, 孟小菲, 沈瑞福, 等. 凡纳滨对虾家系选育的研

- 究与应用[J]. *水产科学*, 2018, 37(4): 555-563.
- Tang Y, Meng X F, Shen R F, *et al.* Research and application of family selective breeding in culture of pacific white shrimp *Litopenaeus vannamei*[J]. *Fisheries Science*, 2018, 37(4): 555-563 (in Chinese).
- [11] 黄永春, 艾华水, 潘忠诚, 等. 凡纳滨对虾抗WSSV选育家系的建立及其抗病特性[J]. *水产学报*, 2013, 37(3): 359-366.
- Huang Y C, Ai H S, Pan Z C, *et al.* Establishment and WSSV resistant characteristics of selective breeding families for resistance to the white spot syndrome virus of *Litopenaeus vannamei*[J]. *Journal of Fisheries of China*, 2013, 37(3): 359-366 (in Chinese).
- [12] Nirea K G, Sonesson A K, Woolliams J A, *et al.* Strategies for implementing genomic selection in family-based aquaculture breeding schemes: double haploid sib test populations[J]. *Genetics Selection Evolution*, 2012, 44(1): 30.
- [13] Meuwissen T H, Hayes B J, Goddard M E. Prediction of total genetic value using genome-wide dense marker maps[J]. *Genetics*, 2001, 157(4): 1819-1829.
- [14] Castillo-Juárez H, Campos-Montes G R, Caballero-Zamora A, *et al.* Genetic improvement of Pacific white shrimp [*Penaeus (Litopenaeus) vannamei*]: perspectives for genomic selection[J]. *Frontiers in Genetics*, 2015, 6: 93.
- [15] Vallejo R L, Leeds T D, Gao G T, *et al.* Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture[J]. *Genetics Selection Evolution*, 2017, 49(1): 17.
- [16] 陈美佳. 不同密度 SNP 和低密度填充面板对凡纳滨对虾收获体重基因组预测准确性的影响 [D]. 上海: 上海海洋大学, 2021.
- Chen M J. Effects of SNPs with different densities and low-density imputation panels on genomic prediction accuracy of harvest weight trait in *Litopenaeus vannamei*[D]. Shanghai: Shanghai Ocean University, 2021 (in Chinese).
- [17] Al-Tobasei R, Ali A, Garcia A L S, *et al.* Genomic predictions for fillet yield and firmness in rainbow trout using reduced-density SNP panels[J]. *BMC Genomics*, 2021, 22(1): 92.
- [18] Wang Q C, Yu Y, Yuan J B, *et al.* Effects of marker density and population structure on the genomic prediction accuracy for growth trait in Pacific white shrimp *Litopenaeus vannamei*[J]. *BMC Genetics*, 2017, 18(1): 45.
- [19] Tsai H Y, Hamilton A, Tinch A E, *et al.* Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations[J]. *Genetics Selection Evolution*, 2016, 48(1): 47.
- [20] Vallejo R L, Silva R M O, Evenhuis J P, *et al.* Accurate genomic predictions for BCWD resistance in rainbow trout are achieved using low-density SNP panels: evidence that long-range LD is a major contributing factor[J]. *Journal of Animal Breeding and Genetics*, 2018, 135(4): 263-274.
- [21] Zhang X J, Yuan J B, Sun Y M, *et al.* Penaeid shrimp genome provides insights into benthic adaptation and frequent molting[J]. *Nature Communications*, 2019, 10(1): 356.
- [22] Yu Y, Zhang X J, Yuan J B, *et al.* Genome survey and high-density genetic map construction provide genomic and genetic resources for the Pacific white shrimp *Litopenaeus vannamei*[J]. *Scientific Reports*, 2015, 5(1): 15612.
- [23] Gilmour A R, Gogel B J, Cullis B R, *et al.* ASReml user guide release 4.1 structural specification[M]. Hemel Hempstead: VSN international ltd, 2015.
- [24] 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.
- [25] Gitterle T, Salte R, Gjerde B, *et al.* Genetic (co)variation in resistance to white spot syndrome virus (WSSV) and harvest weight in *Penaeus (Litopenaeus) vannamei*[J]. *Aquaculture*, 2005, 246(1-4): 139-149.
- [26] Argue B J, Arce S M, Lotz J M, *et al.* Selective breeding of Pacific white shrimp (*Litopenaeus vannamei*) for growth and resistance to Taura Syndrome Virus[J]. *Aquaculture*, 2002, 204(3-4): 447-460.
- [27] Fjalestad K T, Gjedrem T, Carr W H, *et al.* Final report: the shrimp breeding program, selective breeding of *Penaeus vannamei*[R]. Waimanalo: The Oceanic Institute, 1997.

- [28] 刘绵宇. 液相芯片“黄海芯 1 号”在凡纳滨对虾 AHPND 抗性选择育种中的应用 [D]. 连云港: 江苏海洋大学, 2022.
Liu M Y. Application of the liquid chip “Yellow Sea Chip No. 1” in the selective breeding of resistance on AHPND in *Litopenaeus vannamei*[D]. Lianyungang: Jiangsu Ocean University, 2022 (in Chinese).
- [29] 孙坤, 李旭鹏, 隋娟, 等. 基于微卫星标记评估凡纳滨对虾收获体长、体质量以及抗 WSSV 性状的遗传参数 [J]. 水产学报, 2022, 46(4): 583-593.
Sun K, Li X P, Sui J, *et al.* Evaluation of genetic parameters of body length, body weight and WSSV resistance of *Litopenaeus vannamei* based on microsatellite markers[J]. Journal of Fisheries of China, 2022, 46(4): 583-593 (in Chinese).
- [30] Yoshida G M, Carvalho R, Rodríguez F H, *et al.* Single-step genomic evaluation improves accuracy of breeding value predictions for resistance to infectious pancreatic necrosis virus in rainbow trout[J]. *Genomics*, 2019, 111(2): 127-132.
- [31] Garcia A L S, Bosworth B, Waldbieser G, *et al.* Development of genomic predictions for harvest and carcass weight in channel catfish[J]. *Genetics Selection Evolution*, 2018, 50(1): 66.
- [32] Sae-Lim P, Kause A, Lillehammer M, *et al.* Estimation of breeding values for uniformity of growth in Atlantic salmon (*Salmo salar*) using pedigree relationships or single-step genomic evaluation[J]. *Genetics Selection Evolution*, 2017, 49(1): 33.
- [33] Hidalgo J, Lourenco D, Tsuruta S, *et al.* Changes in genomic predictions when new information is added[J]. *Journal of Animal Science*, 2021, 99(2): skab004.
- [34] Lourenco D, Tsuruta S, Masuda Y, *et al.* 384 Genetic and genomic analysis in livestock with increasing datasets[J]. *Journal of Animal Science*, 2020, 98(S4): 137-138.
- [35] Lourenco D, Tsuruta S, Hidalgo J, *et al.* 294 Increased fluctuations of genetic evaluations with genomic information[J]. *Journal of Animal Science*, 2020, 98(S4): 32-33.

Genomic prediction accuracy analysis of AHPND resistance genome prediction in *Litopenaeus vannamei* using SNP panels with different densities

LIU Yang^{1,2}, LUAN Sheng², LIU Mianyu², LI Xupeng², MENG Xianhong², LUO Kun²,
SUI Juan², TAN Jian², DAI Ping², CAO Jiawang², CHEN Baolong², KONG Jie^{2*}

(1. College of Marine Science and Fisheries, Huzhou University, Huzhou 313000, China;

2. Key Laboratory for Sustainable Development of Marine Fisheries, Ministry of Agriculture and Rural Affairs, Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology (Qingdao), Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China)

Abstract: To evaluate the influence of SNP panels with different densities on the accuracy of genomic prediction of AHPND resistance of *Litopenaeus vannamei*, 26 full-sib families were infected with Vp_{AHPND} , and survival time of 686 individuals were recorded. A total of 242 individuals were genotyped by the liquid chip "Yellow Sea Chip No.1" (55.0 K SNP), and the genetic parameters of survival time after Vp_{AHPND} infection were estimated using the A, G and H relationship matrices. Eight low-density SNP panels (40.0 K, 30.0 K, 20.0 K, 10.0 K, 5.0 K, 1.0 K, 0.5 K, 0.1 K) were generated based on the 55.0 K SNP panel. GBLUP and ssGBLUP methods were used to predict the GEBV of survival time, and prediction accuracy was calculated by cross-validation method and compared with BLUP method. The estimated heritabilities of survival time after Vp_{AHPND} infection were high (0.68-0.79). At 55.0 K SNP density, the prediction accuracies obtained by BLUP, GBLUP and ssGBLUP were 0.424, 0.450 and 0.452, respectively, in the data set with 242 genotyped individuals (G242). The accuracy of GBLUP and ssGBLUP was 6.13% and 6.60% higher than BLUP, respectively. For the data set with 686 phenotyped individuals (P686), the prediction accuracies obtained by BLUP and ssGBLUP methods were 0.510 and 0.535, respectively. ssGBLUP showed 4.90% greater accuracy than BLUP. For 8 low-density SNP panels, genomic prediction accuracies were all low (1.1%-1.8%) in G242 and P686 data sets when the SNP density was greater than or equal to 10.0 K. With the decrease of SNP density since 10.0 K, the accuracy of genomic prediction also decreased in the two data sets, The density of 5.0 K decreased by 0.6%-2.6%, 1.0 K decreased by 5.8% -11.0%, 0.5 K decreased by 11.4%-17.2%, and 0.1 K decreased by 38.8%-41.6%. The correlation coefficients of genomic relationship coefficient and GEBV obtained using 10.0 K and 55.0 K SNP panels were all higher than 0.99, which indicates that the relationship between sibling individuals and their GEBVs can be accurately predicted by using the 10.0 K SNP panel. The results showed that the 10.0 K SNP panel could replace 55.0 K SNP panel to perform genetic evaluation for survival time after Vp_{AHPND} infection, which provided a reference for the design of low density SNP chips.

Key words: *Litopenaeus vannamei*; AHPND; low-density SNP panel; accuracy of genomic prediction

Corresponding author: KONG Jie. E-mail: Kongjie@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); Scientific Research Project of Academician Innovation platform in Hainan Province (YSPTZX202104); *Litopenaeus vannamei* Breeding Center of the Hengxing (2021E05032); China Agriculture Research System (CARS-48)