

长毛对虾(*Penaeus penicillatus*)SNP 标记 开发及亲子鉴定技术研究*

游淞元¹ 杨 乐¹ 曾庆民² 张 静¹ 黄良敏¹ 张东玲^{1①} 刘贤德^{1①}

(1. 福建省海洋渔业资源与生态环境重点实验室 集美大学水产学院 福建厦门 361021; 2. 福建省海洋生物增殖与高值化利用重点实验室 福建省水产研究所 福建厦门 361013)

摘要 长毛对虾是我国南方沿海地区的重要经济虾类。近年来,由于过度捕捞、病害频发以及海洋环境恶化,长毛对虾的自然资源量大幅减少。为补充自然资源,改善种群结构,1980 年我国开始对长毛对虾开展增殖放流工作。然而,增殖放流的效果如何,目前尚缺乏有效的评估手段。以长毛对虾为实验材料,通过对其基因组进行 *de novo* 测序和重测序开发 SNP (single nucleotide polymorphism) 标记,最终建立长毛对虾的亲子鉴定技术。具体研究结果如下:长毛对虾基因组大小为 1 335.76 Mb, GC (guanine cytosine) 含量为 40.86%, N50 为 1 221 bp; 以组装的长毛对虾基因组为参考基因组通过重测序进行 SNP 挖掘,共获得 12 579 780 个原始 SNP 位点,经筛选,优化及模型选择,最终建立了基于 192 个 SNP 标记的长毛对虾亲子鉴定技术,亲子鉴定成功率为 100% (95% 的置信度)。研究结果可为下一步长毛对虾增殖放流效果评估提供参考。

关键词 长毛对虾; SNP; 亲子鉴定; 增殖放流

中图分类号 Q953; Q789; S968.2 **doi:** 10.11693/hyhz20221100287

长毛对虾(*Penaeus penicillatus*)又称红虾、大虾、明虾等,隶属于节肢动物门(Arthropoda),甲壳纲(Crustacea),十足目(Decapoda),对虾科(Penaeidae),对虾属(*Penaeus*),属暖水性大型虾类,广泛分布于我国福建、海南、广东和台湾等沿海地区,是我国水产养殖中重要的经济虾类(王渊源,1981;王春忠等,2014)。长毛对虾具有生长速度快、存活率高、抗逆性强等特点。长毛对虾的温度的适应性较广,生存温度为 10~40 °C,在低温条件下,长毛对虾的生长速度比其它虾种的生长速度快。自 20 世纪 50 年代长毛对虾人工苗种技术取得突破之后,长毛对虾一度成为我国南方沿海越冬时的首选换季虾种之一(张桂玲,2009;张桂玲等,2010;袁艺,2016)。20 世纪 90 年代,仅福建省长毛对虾养殖规模就已超过 1.33 万 hm²。然而近年来随着海洋环境恶化、病害频发以及过度捕捞等原因,长毛对虾资源量急剧下降(陈炎辉等,1996;

刘胜男等,2019)。

增殖放流是指向自然水体中投放人工繁育的苗种以增加种群数量、优化群落结构,是恢复渔业资源的重要手段之一(Li, 1999)。自 1980 年起,我国各地区相继开展长毛对虾增殖放流工作(张桂玲,2009)。福建省最早于 1982 年在东吾洋进行长毛对虾的增殖放流,放流总量超过 700 万尾(关金藏,1984)。2003~2009 年,福建省连续 6 年在罗源湾等海域放流长毛对虾共计 11.4 亿尾,通过增殖放流不仅使渔业资源得到恢复性增长,而且还能产生明显的社会、生态、经济效益,据统计,罗源湾海域的长毛对虾捕捞量平均以每年 9.4% 的速度快速增长,放流效果显著(董海,2009;杨爽等,2014)。

随着增殖放流活动开展以及规模不断扩大,如何对放流效果进行准确有效的评估也变得尤为重要(张崇良等,2022)。目前,标志回捕法是增殖放流效

* 国家重点研发计划项目,2018YFD0901404 号。游淞元,硕士研究生, E-mail: 1042607980@qq.com

通信作者: 刘贤德,硕士生导师,教授, E-mail: xdliu@jmu.edu.cn; 张东玲,硕士生导师,副教授, E-mail: 342312933@qq.com

收稿日期: 2022-11-01, 收修改稿日期: 2023-02-10

果评估的主要方法(李小芳, 2012), 传统的标记主要包括物理标记和化学标记。物理标记是利用标志挂牌、剪鳍等方式对个体进行区分, 其缺点是对幼体伤害较大、标志易脱落等(王陌桑, 2016)。化学标记是对幼体注射荧光标志物等, 但化学标记具有成本较高、检测复杂, 有可能对环境造成污染等缺陷(Phinney *et al*, 1967; Catalano *et al*, 2001)。因此, 传统的标记方法难以对放流效果进行有效的评估。

系谱信息是水生动物良种选育的关键, 主要体现在群体遗传力的计算、亲子关系的鉴定以及育种方案的制定等。然而我国多数水生动物并没有建立起一套完整、准确的系谱信息(殷彬等, 2017)。错误的系谱信息往往会增加近亲繁殖的可能, 导致种质资源衰退、遗传多样性降低, 影响良种选育的进程(朱克诚等, 2020)。因此, 开展系谱信息的相关研究具有十分重要的意义。目前, 应用较多的为微卫星标记和 SNP 标记等分子标记。微卫星标记具有基因座小、多态性高、共显性、分辨力高等优点而被广泛应用在水产动物的系谱分析和亲子鉴定研究(Chowdhury *et al*, 2021), 但基于微卫星标记的亲子鉴定方法在基因分型过程中会遇到基因分型错误、等位基因缺失等问题, 影响其鉴定结果(Hoffman *et al*, 2005; Tokarska *et al*, 2009)。近些年发展起来的单核苷酸多态性(SNP)是一种共显性的双等位基因分子标记, 具有遗传稳定性高、位点分布广泛、易实现高通量检测以及基因分型错误率低等优点(Yue *et al*, 2014; Nguyen *et al*, 2018; Van Deventer *et al*, 2020), 是一种更准确可靠的分子标记。相比于 SSR 标记, SNP 标记的亲子鉴定效力更高, 更能够准确有效的评估增殖放流效果。目前, SNP 标记正迅速取代亲子鉴定研究中的微卫星标记, 成为各种水生生物系谱追踪的主流标记之一, 并逐步用于增殖放流效果的评估。

通过 SNP 标记对群体进行亲子鉴定以及遗传多样性等分析, 能够高效地选择繁育材料, 对物种的良种选育、遗传分化等研究具有重要的作用(申淑慧等, 2020)。近些年来, SNP 标记在亲子鉴定、遗传多样性分析以及育种等方面的已有较多报道。如在陆生动物中, Sölzer 等(2022)基于 SNP 标记对荷斯坦奶牛(*Bos taurus*)遗传力进行估计, 并成功找到了两个与牛爪病关联的基因。余国春(2014)利用 120 个 SNP 标记成功对猪(*Sus scrofa f. domestica*)的 24 个样本进行亲缘关系鉴定, 鉴定结果与真实系谱关系完全一致。在水生生物中, Zhao 等(2018)使用 58 个 SNP 标记在大口黑

鲈(*Micropterus salmoides*)群体中进行亲子鉴定, 刘峻宇等(2021)使用 37 个 SNP 标记对 22 个凡纳滨对虾(*Litopenaeus vannamei*)进行了遗传多样性和亲子鉴定分析。然而, 目前长毛对虾尚缺乏 SNP 标记, 也没有看到基于 SNP 标记对长毛对虾进行亲子鉴定技术的相关报道。

为此, 本研究首先初步组装了长毛对虾的基因组, 接着利用长毛对虾的基因组重测序数据开发出大量 SNP 标记, 然后对这些标记进行筛选、评估, 最终建立了基于 SNP 标记长毛对虾的亲子鉴定技术。研究结果可为长毛对虾的增殖放流效果评估、遗传多样性分析提供参考。

1 材料与方法

1.1 实验样品与 DNA 提取

实验于 2021 年 8 月 1 日在福建省虹海水产开发有限公司进行。已抱卵的雌虾由渔民在海区采捕, 运回养殖厂后置于一个产卵池中自然产卵。当雌虾产卵后, 将卵用 200 目过滤网收集后转入到新的水泥池进行孵化、养殖。待子代虾苗养殖至体长 1 cm 以上时, 随机取 49 尾仔虾保存于 80%乙醇中备用。同时, 剪取 91 尾亲虾的腹部肌肉组织保存于 80%乙醇中, 备用。

使用海洋动物组织基因组 DNA 提取试剂盒(天根, 北京)提取 91 尾亲虾和 49 尾仔虾的 DNA。使用 1%的琼脂糖凝胶电泳和酶标仪对 DNA 质量和浓度进行检测。提取的基因组 DNA 储存在-20 °C 备用。

1.2 长毛对虾基因组初步组装

选取 1 尾海区捕捞的长毛对虾进行基因组 *De novo* 测序, 获得原始数据后, 使用 Trimmomatic 软件(Bolger *et al*, 2014)对原始数据进行过滤, 清除低质量、接头污染以及 N 碱基含量过高的序列后获得 clean data。随后, 使用 SOAPdenovo2 软件(Xie *et al*, 2014)对 clean data 进行拼接、组装, 获得长毛对虾的参考基因组。

1.3 SNP 标记开发与过滤

对 140 尾长毛对虾进行全基因组重测序, 通过 BWA+GATK 的 Pipeline (McKenna *et al*, 2010)进行 SNP 挖掘。首先使用 BWA、SAMtools (Li *et al*, 2009)以及 GATK 构建基因组索引, 利用 BWA 软件将每个样品序列 Mapping 到参考基因组上, 通过 SAMtools 软件去除重复的 reads 并排序得到 bam 文件, 最后运行 GATK 软件进行 SNP Calling, 开发长毛对虾的 SNP 标记。为了减少低质量和假阳性位点对后续分析

准确性造成干扰, 使用 VCFtools 软件(Danecek *et al.*, 2011)对原始 SNP 位点过滤, 过滤条件为: (1) 最低的质量分数(minQ) ≥ 30 ; (2) 过滤次要等位基因深度 < 3 的 SNP; (3) 删除缺失率 $> 5\%$ 的 SNP; (4) 仅保留二等位基因; (5) 保留次要等位基因频率 > 0.05 的 SNP。

1.4 亲子鉴定技术的建立

使用 CERVUS 软件(Kalinowski *et al.*, 2007)进行亲子鉴定分析。CERVUS 是基于最大似然法的原理来推断亲子关系。在 CERVUS 软件中, LOD 值(log-likelihood ratio)是进行亲子关系判定的重要指标。在进行鉴定时, 当 LOD 值为正数时表明假设亲本与子代之间可能存在亲子关系, 且 LOD 越大, 可靠性越强。实验使用的长毛对虾亲本是来自于海区捕获的抱卵雌虾, 所以孵化出的长毛对虾子代只能确定其母本, 无法确定其父本。因此, 本研究选择单亲母权亲子鉴定模型对长毛对虾进行亲子关系鉴定, 对于当前的研究分析, 只有置信度达到 95%以上时, 子代才会被视为正确的分配给亲本。进行亲子鉴定时, 模拟参数如下: 91 个候选母本产生 10 000 个模拟后代, 100%的母本被抽样, “最小分型基因座”设置为 SNP 总数的一半, 1%的分型错误率, 设置 80%和 95%两个置信度。

本研究从次要等位基因频率(minor allele frequency, MAF)、SNP 数量两个方面对亲子鉴定结果进行优化、分析。首先, 在保持 SNP 个数不变的情况下, 对 MAF 值的重要性进行分析。根据 MAF 的大小, 设置若干组 MAF, 每组随机挑选相同的 SNP 位点进行亲子鉴定分析; 接着, 根据上一步的结果, 在保持 MAF 值相对恒定的情况下, 对 SNP 个数进行优化, 形成若干

组 SNP 组合, 进行亲子鉴定分析; 然后再综合其它参数对 SNP 个数进行进一步的优化, 最终确定适合长毛对虾亲子鉴定的 SNP 标记个数。

2 结果

2.1 长毛对虾基因组初步组装

利用 Platanus 软件(Pryszcz *et al.*, 2016)计算组装长毛对虾基因组最佳 K 值, 结果显示当 $K=27$ 时, 组装效果较好。基于 $K=27$, 利用 GenomeScope 软件(Vurture *et al.*, 2017)评估长毛对虾基因组, 大小为 1 445.70 Mb, 杂合度为 0.65%, 重复率为 71.3%。通过 SOAPdenovo2 ($K=27$)进行组装, 组装结果为: 长毛对虾基因组大小为 1 335.76 Mb, GC 含量为 40.86%, N50 为 1 221 bp。此次长毛对虾基因组基于纯二代测序组装, 且基因组本身重复序列多, 杂合度高, 故 N50 长度比预期短。经 BUSCO 评估结果显示, 215 (21.2%)条单拷贝基因序列被覆盖, 1 (0.1%)条多拷贝被覆盖, 377 (37.2%)条基因序列覆盖不完全, 412 (41.6%)条基因序列缺失, 表明基因组组装完整性偏低, 但用于 SNP 标记开发已经足够。

2.2 长毛对虾 SNP 标记开发与过滤

采用 BWA+GATK 的策略对 140 尾长毛对虾的重测序数据进行 SNP 挖掘, 共获得 12 579 780 个原始 SNP 位点。按照表 1 标准筛选 SNP, 过滤质量值 < 30 、次要等位基因深度 < 3 的 SNP, 得到 111 213 个 SNP。随后, 删除缺失率 $> 5\%$ 的 SNP, 剩余 2 717 个 SNP; 最后, 仅保留双等位基因以及保留 MAF > 0.05 的 SNP 后, 共得到 630 个可用于亲子鉴定的候选位点。

表 1 SNP 过滤过程
Tab.1 The SNP filtering process

步骤	1	2	3	4	5
过滤指令	--minQ 30	--mac 3	--max-missing 0.95	--min-alleles 2 --max-alleles 2	--maf 0.05
剩余 SNP 个数	12 579 780	9 876 172	111 213	2 717	630

注: --minQ 30、--mac 3、--max-missing 0.95、--min-alleles 2、--max-alleles 2、--maf 0.05 为过滤过程中的过滤指令, “--”为连接上一参数的指令

2.3 长毛对虾亲子鉴定技术建立

2.3.1 次要等位基因频率(MAF)的重要性评估 根据 MAF 的大小, 设置三组 MAF, 分别为 0.05~0.20、0.20~0.35 和 0.35~0.50 (梯度为 0.15), 每组随机挑选 95 个 SNP 位点进行亲子鉴定分析, 鉴定结果如表 2 所示。在 95%的置信度下, 三组 SNP 标记的亲子鉴定成功率随着 MAF 的增加而增加, 鉴定成功率从

表 2 SNP 个数相同时, 不同 MAF 范围亲子鉴定结果
Tab.2 Parentage assignment under different MAF ranges and the same number of SNP

MAF 范围	SNP 个数	匹配上的子代	亲子鉴定成功率
0.05~0.20	95	20	40.8%
0.20~0.35	95	34	69.3%
0.35~0.50	95	40	81.6%

40.8%上升到 81.6%，说明在 SNP 个数保持不变的情况下，亲子鉴定成功率与 MAF 值呈正比。

2.3.2 SNP 个数的重要性评估 根据上一步得出的结论，当 MAF 在 0.35~0.50 之间时，亲子鉴定成功率最高。随后在此范围下形成六组 SNP，每组的 SNP 个数在 50 到 300 之间变化(梯度为 50)，使用 CERVUS 分析 50、100、150、200、250 和 300 个 SNP 的遗传信息，结果如表 3 所示。总体来看，随着 SNP

个数的增加，基因分型比例以及各项遗传信息也随之增加。在 95%的置信度下，使用母权亲子鉴定的模型进行亲子鉴定，鉴定结果如图 1 所示。当 SNP 个数为 50 时，亲子鉴定成功率最低，仅为 24.4%。但随着 SNP 数量的增加，鉴定成功率也逐渐提高。当 SNP 个数达到 200 个时，亲子鉴定成功率达到 100%。随后，继续增加 SNP 个数，亲子鉴定成功率也没有发生改变。

表 3 六组 SNP 组合(MAF=0.35~0.50)的遗传信息
Tab.3 Genetic information of six SNP combinations (MAF=0.35~0.50)

MAF	SNP 个数	基因组分型比例	平均观测杂合度	平均期望杂合度	平均多态信息含量
0.35~0.50	50	0.965 4	0.853 9	0.487 5	0.367 7
	100	0.966 9	0.852 6	0.487 8	0.367 8
	150	0.966 9	0.856 9	0.488 8	0.368 3
	200	0.966 9	0.859 0	0.489 0	0.368 5
	250	0.967 4	0.885 4	0.491 5	0.369 7
	300	0.967 4	0.902 9	0.493 2	0.370 4

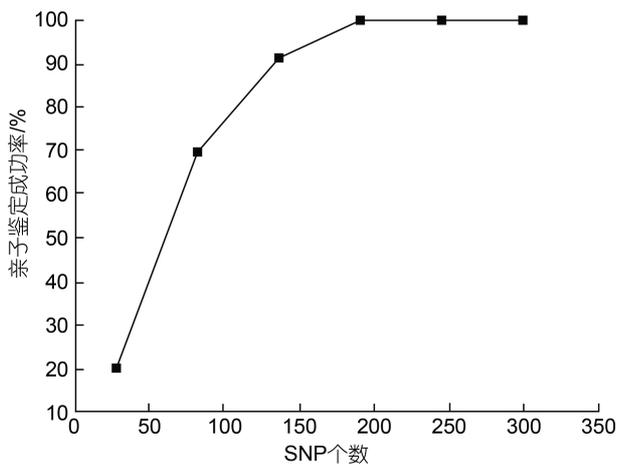


图 1 MAF 值保持恒定下不同 SNP 个数的亲子鉴定成功率
Fig.1 Percentage assignment ratio under different numbers of SNPs and a constant MAF value

2.3.3 最优 SNP 个数的选择 图 1 显示，当 SNP 个数为 150 时，亲子鉴定成功率为 91.8%，SNP 个数为 200 时，亲子鉴定成功率为 100%，因此我们在 150~200 个 SNP 的范围内寻找亲子鉴定成功率达到 100%的最少 SNP 个数。对 200 个 SNP 的 H_o 值按照从小到大排序，依次删除 H_o 最小的 SNP 位点进行亲子鉴定分析。结果显示，当使用 192 个 SNP 位点时，亲子鉴定成功率达到 100%。因此，本研究 192 个位点为最佳 SNP 个数。

2.3.4 遗传多样性分析 利用 192 个 SNP 标记对

140 个长毛对虾个体进行遗传多样性分析，结果如表 4 显示，192 个 SNP 标记在 140 个长毛对虾个体中的平均观测杂合度(H_o)为 0.478，平均期望杂合度(H_e)为 0.359，平均多态信息含量(Polymorphism information content, PIC)为 0.292。

表 4 长毛对虾 140 个个体在 192 个 SNP 位点上的遗传多样性

Tab.4 Genetic diversity of 140 individuals on 192 SNPs in *P. penicillatus*

SNP 位点个数	平均值					
	H_o	H_e	PIC	E-1P	E-2P	E-PP
192	0.478	0.359	0.292	0.934	0.854	0.769

注: NE-1P 指第一个亲本的非亲权排除概率; NE-2P 指另一亲本的非亲权排除概率; NE-PP 指亲本对非亲权排除概率

3 讨论

随着现代生物技术的发展，SNP 标记已被逐渐应用于水产动物的系谱分析、亲子鉴定等研究(盖超伟等, 2021)。本研究在初步组装长毛对虾基因组的基础上，通过基因组重测序方法开发出大量的 SNP 标记，经过筛选优化最终建立了长毛对虾的亲子鉴定技术，为长毛对虾增殖放流效果的评估提供了有效的技术手段。

本研究通过对 140 尾长毛对虾进行全基因组测序共开发了 192 个可用于亲子关系鉴定的 SNP 标记，192 个 SNP 位点的平均 PIC 含量为 0.292， H_o 平均值

为 0.478, H_e 平均值为 0.359, 根据 Botstein 等(1980)提出的多态性划分标准, $0.25 < PIC < 0.5$ 为中度多态性。低于微卫星标记分析长毛对虾群体所得到的 PIC 值($PIC=0.690$) (刘胜男等, 2019)。造成 SNP 标记 PIC 含量低于 SSR 标记的主要原因是由于 SNP 标记主要为二等位基因(刘峻宇等, 2021)。

亲子鉴定的成功率受到多种因素的影响, 即 MAF、基因分型错误率和 SNPs 数目(Premachandra *et al*, 2019)。次等位基因频率(MAF)是指在特定群体中, 第二常见的等位基因出现的频率。MAF 值是影响亲子鉴定效率的关键, 通常随着 MAF 值的增加, 亲子鉴定成功率也会随之增加(Vu *et al*, 2021), 在许多水产生物中都有鉴定成功率随 MAF 值增加而增加的相关报道。在鞍带石斑鱼(*Epinephelus lanceolatus*)中, 选用高 MAF 值的 SNP 标记的非亲本排除概率要高于选用低 MAF 值的 SNP 标记的非亲本排除概率(Weng *et al*, 2021)。在大西洋鲑(*Salmo salar*)中, 当 MAF 值从 0.10 增加到 0.40 时, 亲子鉴定成功率从 98.6% 增加到 100%; 在黄条魮(*Seriola lalandi*)中, 在保持 SNP 个数相同的情况下, 随着 MAF 的增加, 亲子鉴定成功率从 87.36% 提高到 94.58% (Dussault *et al*, 2018; Premachandra *et al*, 2019)。在本研究中, 在保持 SNP 数量不变的条件下, 亲子鉴定的成功率随着 MAF 的增加而增加, MAF 值从 0.05 增加到 0.50, 成功率从 40.8% 上升到 81.6%, MAF 在 0.35~0.50 时, 鉴定成功率最高, 说明亲子鉴定的成功率与 MAF 值呈正比, 这与上述研究结果相符。

基因分型错误率也是影响亲子鉴定成功率的原因之一, 当基因分型错误率较低时(0%~10%)对鉴定的成功率影响不大, 当基因分型错误率较高时, 亲子鉴定的成功率则会大幅降低。当有分型错误的个体分配到遗传群体中, 则有可能会影响同胞个体的亲缘关系鉴定(Wang, 2004)。通常情况下, 实验室的基因分型准确率很少超过 99%, 因此我们在计算机模拟过程中, 基因分型错误率默认为 1% (Kalinowski *et al*, 2007)。

SNP 数量是影响亲子鉴定成功率另一重要因素。Weng 等(2021)发现, 在鞍带石斑鱼中亲子鉴定率达到 100% 需要 208 个 SNP 位点。一般情况下, 随着 SNP 数量的增加, 亲子鉴定成功率也会增加, 直至保持不变。本研究中, 在保持 MAF 不变的情况下, 设置六组不同数量的 SNP 进行亲子鉴定分析, 结果显示亲子鉴定成功率随着 SNP 数量的增加而增加, 这也与珍珠贝(*Pinctada maxima*) (Massault *et al*, 2021)、鲤鱼

(*Cyprinus carpio*) (Xu *et al*, 2017)、虹鳟(*Oncorhynchus mykiss*) (Abadía-Cardoso *et al*, 2013)的研究结果一致。Flanagan 等(2019)认为在大多数情况下, 100~500 个 SNP 就足以解决亲子关系问题, 然而在一些物种中, 使用更少的位点数就可完成亲子鉴定。如 Harney 等(2018)使用 60 个信息含量最丰富的 SNP 标记在欧洲鲍螺(*Haliotis tuberculata*)复杂家系中实现了 95% 亲子鉴定成功率; 在东方牡蛎(*Crassostrea virginica*)中, Thongda 等(2018)开发了 58 个 SNP 便达到了 99.37% 的亲子鉴定成功率。在本研究中, 则需要 192 个 SNP 才能在达到 95% 以上鉴定成功率, 用于亲子鉴定的位点数偏多, 可能与长毛对虾基因组较大、基因组的杂合度较高有关。

总之, 亲子鉴定成功率受多方面的影响, 如 SNP 标记的数量、质量、基因分型准确性及群体特征等(Thongda *et al*, 2018)。因此在建立某些物种亲子鉴定技术时, 不仅需要开发出具有多态性高的遗传标记, 也要注重提高分型、分析过程的准确性, 进而确保建立的亲子鉴定技术的可靠性。

4 结论

本研究以长毛对虾为实验材料, 通过对其进行基因组 *De novo* 测序和重测序, 初步组装了长毛对虾基因组; 在此基础上, 通过基因组重测序开发了 12 579 780 个 SNP, 经优化, 最终建立了基于 192 个 SNP 标记的长毛对虾亲子鉴定技术, 鉴定成功率达到 100%。该结果可为下一步长毛对虾增殖放流效果评估、遗传多样性分析提供参考。

参 考 文 献

- 王陌桑, 2016. 中国对虾增殖放流群体溯源分析及迁徙动态研究[D]. 上海: 上海海洋大学: 7-8.
- 王春忠, 林国荣, 严涛, 等, 2014. 长毛对虾海水养殖环境以及虾肠道微生物群落结构研究[J]. 水产学报, 38(5): 706-712.
- 王渊源, 1981. 长毛对虾[J]. 动物学杂志, 16(1): 73-74.
- 申淑慧, 戴习林, 2020. 基于生长和抗逆功能基因 SNP 分子标记的凡纳滨对虾野生及选育群体遗传多样性分析[J]. 南方农业学报, 51(11): 2836-2845.
- 朱克诚, 宋岭, 刘宝锁, 等, 2020. 黄鳍棘鲷家系亲缘关系鉴定[J]. 水产学报, 44(3): 351-357.
- 刘胜男, 孙典荣, 刘岩, 等, 2019. 使用微卫星标记评估北部湾长毛对虾增殖放流效果[J]. 中国海洋大学学报, 49(S2): 37-45.
- 刘峻宇, 刘均辉, 孔杰, 等, 2021. 凡纳滨对虾 SNP 标记开发与家系亲缘关系验证分析[J]. 渔业科学进展, 42(1): 108-116.
- 关金藏, 郝尧兰, 柯才焕, 1984. 福建省东吾洋东方对虾和长毛对虾的标志放流[J]. 海洋渔业(06): 243-248.

- 李小芳, 2012. 鲢(*Hypophthalmichthys molitrix*)亲本增殖放流遗传效果评估[D]. 重庆: 西南大学: 8-9.
- 杨爽, 宋娜, 张秀梅, 等, 2014. 基于线粒体控制区序列的三疣梭子蟹增殖放流亲蟹遗传多样性研究[J]. 水产学报, 38(8): 1089-1096.
- 余国春, 2014. 微卫星与 SNP 标记技术在猪亲子鉴定中的有效性研究[D]. 成都: 四川农业大学: 1-2.
- 张桂玲, 2009. 我国东南沿海长毛明对虾群体遗传多样性与分化的研究[D]. 厦门: 集美大学: 4-5.
- 张桂玲, 黎中宝, 王展林, 等, 2010. 长毛明对虾的研究现状与展望[J]. 现代渔业信息, 25(2): 7-10.
- 张崇良, 徐宾铎, 薛莹, 等, 2022. 渔业资源增殖评估研究进展与展望[J]. 水产学报, 46(8): 1509-1524.
- 陈炎辉, 洪心, 1996. 养殖长毛对虾虾体长体重关系式研究[J]. 福建水产, 18(4): 11-14.
- 袁艺, 2016. 长毛明对虾微卫星的开发及雌雄群体遗传多样性的研究[D]. 厦门: 集美大学: 1-2.
- 殷彬, 岳书俭, 俞英, 等, 2017. 奶牛分子系谱构建或亲权鉴定的微卫星标记筛选[J]. 畜牧兽医学报, 48(4): 595-604.
- 盖超伟, 李旭鹏, 曹宝祥, 等, 2021. 中国对虾 4 代人工选育群体与 1 个野生群体遗传多样性分析及差异 SNP 位点筛查[J]. 中国水产科学, 28(12): 1505-1514.
- 董海, 2009. 福建省渔业增殖放流取得明显生态与经济效益[J]. 现代渔业信息, 24(5): 31.
- ABADÍA-CARDOSO A, ANDERSON E C, PEARSE D E, *et al*, 2013. Large-scale parentage analysis reveals reproductive patterns and heritability of spawn timing in a hatchery population of steelhead (*Oncorhynchus mykiss*) [J]. *Molecular Ecology*, 22(18): 4733-4746.
- BOLGER A M, LOHSE M, USADEL B, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data [J]. *Bioinformatics*, 30(15): 2114-2120.
- BOTSTEIN D, WHITE R L, SKOLNICK M, *et al*, 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms [J]. *American Journal of Human Genetics*, 32(3): 314-331.
- CATALANO M J, CHIPPS S R, BOUCHARD M A, *et al*, 2001. Evaluation of injectable fluorescent tags for marking centrarchid fishes: retention rate and effects on vulnerability to predation [J]. *North American Journal of Fisheries Management*, 21(4): 911-917.
- CHOWDHURY L M, KATHIRVELPANDIAN A, DIVYA P R, *et al*, 2021. Genetic characterization of Kiddi shrimp, *Parapenaeopsis styliifera* (H. Milne Edwards, 1837) along the Indian coast using microsatellite markers [J]. *Fisheries Research*, 244: 106128.
- DANECEK P, AUTON A, ABECASIS G, *et al*, 2011. The variant call format and VCFtools [J]. *Bioinformatics*, 27(15): 2156-2158.
- DUSSAULT F M, BOULDING E G, 2018. Effect of minor allele frequency on the number of single nucleotide polymorphisms needed for accurate parentage assignment: a methodology illustrated using Atlantic salmon [J]. *Aquaculture Research*, 49(3): 1368-1372.
- FLANAGAN S P, JONES A G, 2019. The future of parentage analysis: from microsatellites to SNPs and beyond [J]. *Molecular Ecology*, 28(3): 544-567.
- HARNEY E, LACHAMBRE S, ROUSSEL S, *et al*, 2018. Transcriptome based SNP discovery and validation for parentage assignment in hatchery progeny of the European abalone *Haliotis tuberculata* [J]. *Aquaculture*, 491: 105-113.
- HOFFMAN J I, AMOS W, 2005. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion [J]. *Molecular Ecology*, 14(2): 599-612.
- KALINOWSKI S T, TAPER M L, MARSHALL T C, 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment [J]. *Molecular Ecology*, 16(5): 1099-1106.
- LI J, 1999. An appraisal of factors constraining the success of fish stock enhancement programmes [J]. *Fisheries Management and Ecology*, 6(2): 161-169.
- LI H, HANDSAKER B, WYSOKER A, *et al*, 2009. The sequence alignment/map format and SAMtools [J]. *Bioinformatics*, 25(16): 2078-2079.
- MASSAULT C, JONES D B, ZENGER K R, *et al*, 2021. A SNP parentage assignment panel for the silver lipped pearl oyster (*Pinctada maxima*) [J]. *Aquaculture Reports*, 20: 100687.
- MCKENNA A, HANNA M, BANKS E, *et al*, 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data [J]. *Genome Research*, 20(9): 1297-1303.
- NGUYEN N H, PREMACHANDRA H K A, KILIAN A, *et al*, 2018. Genomic prediction using DArT-Seq technology for yellowtail kingfish *Seriola lalandi* [J]. *BMC Genomics*, 19(1): 107.
- PHINNEY D E, MILLER D M, DAHLBERG M L, 1967. Mass-marking young salmonids with fluorescent pigment [J]. *Transactions of the American Fisheries Society*, 96(2): 157-162.
- PREMACHANDRA H K A, NGUYEN N H, KNIBB W, 2019. Effectiveness of SNPs for parentage and sibship assessment in polygamous yellowtail kingfish *Seriola lalandi* [J]. *Aquaculture*, 499: 24-31.
- PRYSZCZ L P, GABALDÓN T, 2016. Redundans: an assembly pipeline for highly heterozygous genomes [J]. *Nucleic Acids Research*, 44(12): e113.
- SÖLZER N, MAY K, YIN T, *et al*, 2022. Genomic analyses of claw disorders in Holstein cows: genetic parameters, trait associations, and genome-wide associations considering interactions of SNP and heat stress [J]. *Journal of Dairy Science*, 105(10): 8218-8236.
- THONGDA W, ZHAO H G, ZHANG D D, *et al*, 2018. Development of SNP panels as a new tool to assess the genetic diversity, population structure, and parentage analysis of the eastern oyster (*Crassostrea virginica*) [J]. *Marine Biotechnology*, 20(3): 385-395.
- TOKARSKA M, MARSHALL T, KOWALCZYK R, *et al*, 2009. Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison [J]. *Heredity*, 103(4):

- 326-332.
- VAN DEVENTER R, RHODE C, MARX M, *et al*, 2020. The development of genome-wide single nucleotide polymorphisms in blue wildebeest using the DArTseq platform [J]. *Genomics*, 112(5): 3455-3464.
- VU S V, PREMACHANDRA H K A, O'CONNOR W, *et al*, 2021. Development of SNP parentage assignment in the Portuguese oyster *Crassostrea angulata* [J]. *Aquaculture Reports*, 19: 100615.
- VURTURE G W, SEDLAZECK F J, NATTESTAD M, *et al*, 2017. GenomeScope: fast reference-free genome profiling from short reads [J]. *Bioinformatics*, 33(14): 2202-2204.
- WANG J L, 2004. Sibship reconstruction from genetic data with typing errors [J]. *Genetics*, 166(4): 1963-1979.
- WENG Z Y, YANG Y, WANG X, *et al*, 2021. Parentage analysis in giant grouper (*Epinephelus lanceolatus*) using microsatellite and SNP markers from genotyping-by-sequencing data [J]. *Genes*, 12(7): 1042.
- XIE Y L, WU G X, TANG J B, *et al*, 2014. SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads [J]. *Bioinformatics*, 30(12): 1660-1666.
- XU J, FENG J Y, PENG W Z, *et al*, 2017. Development and evaluation of a high-throughput single nucleotide polymorphism multiplex assay for assigning pedigrees in common carp [J]. *Aquaculture Research*, 48(4): 1866-1876.
- YUE G H, XIA J H, 2014. Practical considerations of molecular parentage analysis in fish [J]. *Journal of the World Aquaculture Society*, 45(2): 89-103.
- ZHAO H G, LI C, HARGROVE J S, *et al*, 2018. SNP marker panels for parentage assignment and traceability in the Florida bass (*Micropterus floridanus*) [J]. *Aquaculture*, 485: 30-38.

DEVELOPMENT OF SNP MARKERS AND PARENTAGE ASSIGNMENT TECHNIQUES IN *PENAEUS PENICILLATUS*

YOU Song-Yuan¹, YANG Le¹, ZENG Qing-Min², ZHANG Jing¹, HUANG Liang-Min¹,
ZHANG Dong-Ling¹, LIU Xian-De¹

(1. Fujian Provincial Key Laboratory of Marine Fishery Resources and Eco-environment, Fisheries College of Jimei University, Xiamen 361021, China; 2. Key Laboratory of Cultivation and High-value Utilization of Marine Organisms in Fujian Province, Fisheries Research Institute of Fujian, Xiamen 361013, China)

Abstract *Penaeus penicillatus* is an essential marine economic shrimp in the southeastern coastal regions of China. Wild populations of *P. penicillatus* have declined dramatically in recent years due to overfishing, marine environmental pollution, and other causes. Since 1980, China has initiated numerous stock-enhancing programs for *P. penicillatus* to replenish natural resources and improve population structure. However, no efficient evaluation method is available at present. SNP (single nucleotide polymorphism) markers were developed by *De novo* genome sequencing and resequencing in *P. penicillatus*, and parentage assignment techniques were established for *P. penicillatus*. Results show that the genome size of *P. penicillatus* is 1 335.76 Mb, GC (guanine cytosine) content is 40.86%, and the assembly of the genome contigs N50 reaches 1 221 bp. The whole-genome sequencing and SNP marker discovery were performed using the assembled *P. penicillatus* genome as the reference genome, yielding a total of 12 579 780 original SNP markers. Through screening, optimization, and model selection, a parentage assignment technique was developed for *P. penicillatus* based on 192 SNP markers, by which a parentage assignment rate of 100% with 95% confidence was achieved. This study offered a reference for future evaluation on the stock-enhancing effects for *P. penicillatus*.

Key words *Penaeus penicillatus*; SNP; parentage assignment; enhancement and release