

基于 SNP 标记的团头鲂、黑尾近红鲌和 鲌鲂“先锋 2 号”群体遗传分析



李佩, 陈见, 李清, 王贵英, 孙艳红*

武汉市农业科学院水产研究所, 湖北武汉 430207

摘要: 研究应用 SLAF-seq 开发了团头鲂、黑尾近红鲌以及鲌鲂“先锋 2 号”的 SNP 标记并研究了群体遗传多样性和群体遗传结构, 结果显示 Rsa-I 的酶切效率为 90.11%, 三种鱼共得到 181.13M reads。通过生物信息学分析, 共获得 SLAF 标签 1,339,007 个, 共得到 446,680 个 SNP, 测序平均 Q30 为 93.03%, 平均 GC 含量 37.79%, 测得观察等位基因数 (N_o) 为 2.0, 期望等位基因数 (N_e) 为 1.63, 观察杂合度 (H_o) 为 0.08, 期望杂合度 (H_e) 为 0.36, 遗传多样性指数 (Genetic diversity index) 为 0.41, 香浓指数 (Shannon Wiener index) 为 0.54, 多态信息含量 (PIC) 为 0.29, 次要等位基因型频率 (MAF) 为 0.28, 表明三种鱼群体遗传多样处于中等水平, 这可能跟三种鱼长期的人工选育有关。三种鱼的遗传距离显示鲌鲂“先锋 2 号”与团头鲂最近、与黑尾近红鲌次之、团头鲂和黑尾近红鲌最远, 这也与形态学基本保持一致。

关键词: 团头鲂; 黑尾近红鲌; 鲌鲂“先锋 2 号”; SNP 标记; SLAF-seq; 遗传多样性; 遗传结构

DOI: 10.57237/j.jaf.2023.03.003

Genetic Analysis of *Megalobrama amblycephala*, *Ancherythroculter nigrocauda* and Bo Fang “Xianfeng No. 2” Based on SNP Markers

Li Pei, Chen Jian, Li Qing, Wang Gui-Ying, Sun Yan-Hong*

Fisheries Research Institute, Wuhan Academy of Agricultural Sciences, Wuhan 430207, China

Abstract: The SNP markers of *Megalobrama amblycephala*, *Ancherythroculter nigrocauda* and Bo Fang “Xianfeng No. 2” were developed by SLAF-seq in this study. Further analysis on genetic diversity and genetic population were also done in this paper. The results showed that the efficiency of Enzyme digestion was 90.11% and 181.13M reads were obtained from three fishes. 1,339,007 SLAF labels and 446,680 SNP markers were through bioinformatics analysis. And the Q30 of mean sequencing was 93.03%, the mean GC content was 37.79%. The results showed that the observed allele number (N_o) was 2.0, and the expected allele number (N_e) was 1.63. The observed heterozygosity (H_o) was 0.08, while the expected heterozygosity (H_e) was 0.36. The genetic diversity index was 0.41, the Shannon Wiener index was 0.54, and the PIC was 0.29, the MAF was 0.28. These results suggest that the genetic diversity of three populations were in the medium level. It might be related to artificial selection. The nearest genetic distance was between Bo Fang “Xianfeng No. 2” and *Megalobrama amblycephala*, and the furthest distance between was *Megalobrama amblycephala* and

基金项目: 湖北省科技创新专项 (2018ABA105); 湖北洪山实验室开放课题 (2021hskf015).

*通信作者: 孙艳红, 176570314@qq.com

收稿日期: 2023-08-10; 接受日期: 2023-09-20; 在线出版日期: 2023-10-09

<http://www.agrforestry.com>

Ancherythroculter nigrocauda, which were in accordance with their shapes.

Keywords: *Megalobrama amblycephala*; *Ancherythroculter nigrocauda*; Bo Fang “Xianfeng No. 2”; SNP Marker; SLAF-seq; Genetic Diversity; Genetic Population

1 引言

团头鲂 (*Megalobrama amblycephala*), 隶属鲤形目 (Cypriniformes)、鲤科 (Cyprinidae)、鲃亚科 (Culterinae)、鲂属 (*Megalobrama*), 是中国重要的淡水经济鱼类, 广泛分布于各流域湖泊, 因其食性广、易捕捞、生长快、肉质细嫩而深受喜爱; 黑尾近红鲃 (*Ancherythroculter nigrocauda*), 隶属鲤形目 (Cypriniformes)、鲤科 (Cyprinidae)、鲃亚科 (Culterinae)、近红鲃属 (*Ancherythroculter*), 是长江上游特有经济鱼类; 其人工养殖成本低、性情温驯、易捕捞、易活鱼上市。远缘杂交是目前育种的重要手段之一, 可以把亲本的优良性状结合在一起, 笔者团队将选育的 4 代团头鲂 (♀) 与选育的 4 代黑尾近红鲃 (♂) 进行杂交, 再经过 4 代以上的选育获得了体型与团头鲂相近, 生长快、抗应激能力强且在肉质方面上具有优势的鲃“先锋 2 号”, 已于 2019 年被全国水产原种和良种审定委员会审定为新品种。

单核苷酸多态性 (SNPs, single nucleotide polymorphisms) 是基因组中常见的遗传变异形式, 是指在基因组水平上由单个核苷酸的变异引起的一种 DNA 序列多态性, 是基因组中分布最广泛而稳定的点突变; SNP 具有数量多、分布广泛、遗传稳定性好、便于高通量、高度自动化检测分析等优点, 通过研究其在基因组中的分布能够全面的反映群体的遗传变异水平。SLAF-seq (Specific-Locus Amplified Fragment Sequencing) 技术是一种高通量、高分辨率的 SNPs 位点识别和分型技术, 可在短时间内以低成本筛选出大量多态性 SNPs 标记, 准确性高。

本文采用 SLAF-seq 技术开发了三种鱼的 SNP 标记, 并进一步分析了三者的群体遗传多样性水平和遗传结构, 对三者作出遗传特性评价。

2 材料与方法

2.1 实验材料

团头鲂 (*Megalobrama amblycephala*, TTF) 群体引

自湖北省梁子湖野生群体, 黑尾近红鲃 (*Ancherythroculter nigrocauda*, HWB) 群体引自长江上游野生群体, 两者均经过了武汉市农科院水产所和武汉先锋水产科技有限公司 (先锋公司) 4 代选育, 将选育的团头鲂 (♀) 与选育的黑尾近红鲃 (♂) 进行杂交, 再经过 4 代以上选育获得的鲃“先锋 2 号” (Bo Fang “Xianfeng No. 2”, XF2)。目前三种鱼均在先锋公司基地进行人工养殖。三种鱼各取 10 尾, 分别剪尾鳍保存于无水乙醇中, 带回实验室于 4℃ 冰箱中备用。

2.2 基因组 DNA 制备

采用传统酚-氯仿方法提取基因组 DNA [1], 采用琼脂糖凝胶电泳和 Nanodrop2000 检测抽提并测算基因组 DNA 的浓度。抽提的基因组 DNA 置于 -20℃ 冰箱备用。

2.3 SLAF-seq 文库构建及高通量测序

根据 Sun [2] 等的方法构建三种鱼的 SLAF-seq 文库。选择翘嘴鲃基因组作为参考基因组进行基因组电子酶切预测, 内切酶的选择原则为: 位于重复序列的酶切片段比例尽可能低, 酶切片段均匀分布于基因组上, 并使酶切片段数符合预期 SLAF 标签数 [3]。根据选定的最适酶切方案, 对检测合格的各样品基因组 DNA 分别进行酶切实验。对得到酶切的片段进行 3' 端加 A 处理、连接测序接头、PCR 扩增、纯化、混样、切胶选取目的片段, 文库质检合格后用 IlluminaHiSeq™ 进行双末端测序。为评估建库实验的准确性, 选用日本晴水稻 (*Oryza sativa indica*) 作为对照 (Control) 进行相同的处理参与建库和测序。

2.4 SNP 标记开发

根据序列的相似度, 将聚类在一起的 Reads 定义为一个 SLAF 标签。根据 SLAF 标签在不同样品间的等位基因数和基因序列间的差异程度筛选多态性的 SLAF 标签, 并以每个 SLAF 标签中测序深度最高的序

列类型作为参考序列[3]。SNP 的检测工具主要使用 GATK 软件工具包实现。根据测序 Reads 在参考基因组上的定位结果, GATK 进行局部重比对、GATK 变异检测, samtools 变异检测, 取 GATK 和 samtools 两种方法得到的交集变异位点, 以保证检测得到的 SNP 的准确性, 并得到最终的 SNP 位点集。

2.5 群体遗传多样性分析

采用 Genemarker 录入每个微卫星位点各个个体的等位基因数据。

对获取的 SNP 位点进行筛选, 标准为次要等位基因频率 (Minor allele frequency, MAF) 大于 0.05、完整度大于 0.8。使用符合条件的 SNP 用于三个群体遗传多样性和遗传结构分析。

使用 Power-Marker V3.25 软件计算观测等位基因数 (Observed allele number, N_o)、期望等位基因数 (Expected allele number, N_e)、观测杂合度 (Observed heterozygosity, H_o)、期望杂合度 (Expected heterozygosity, H_e)、遗传多样性指数 (Genetic diversity index)、香浓指数 (Shannon Wiener index)、多态性信息含量 (Polymorphism information content, PIC)、次要等位基因频率 (MAF)。

2.6 群体遗传结构分析

使用 Admixture [4]软件分析三种鱼的群体遗传结构, 使用 MEGA5 与 Neighbor-Joining 算法构建个体间的进化关系分析以及群体遗传距离[5]。

采用软件 Poptree V2 和 MEGA X 基于等位基因频率与遗传距离建立代表群体遗传关系的聚类树; PCoA-GenALEx v6.5 基于遗传距离来分析群体间的遗传关系。

3 结果

3.1 SLAF 酶切方案及建库分析

通过对翘嘴鲃基因组进行酶切预测, 选择 *RsaI* 内切酶组合酶切鲃鲂“先锋 2 号”基因组、团头鲂、黑尾近红鲂, 目标酶切片长度为 414-494bp, 预测可获得 120304 个 SLAF 标签。测序获得水稻 (Control) 数据量为 382M reads, 与参考基因组比对, Paired-end mapped reads、single-end mapped reads 和 unmapped reads 占总 reads 的

比例分别为 87.77%、4.87%与 7.36%; 由此可知, 双端比对效率较高, 比对结果较好。另外, 水稻测序 reads 插入片段中残留酶切位点的比列为 9.89%, 酶切效率为 90.11%, 表明酶切效果较好。上述数据显示本研究的 SLAF-seq 测序文库构建正常。从所有样本中共获得 181.13 百万序列 (MReads) 数据, 平均 Q30 (测序正确率为 99.9%的碱基比例) 为 93.03%, 平均 GC 含量为 37.79%, 表明测序质量较高, 测序结果可靠。

3.2 SNP 标记开发

从 30 个样本中共获得 1,339,007 个 SLAF 标签, 平均测序深度为 25.91x, 从 SLAF 标签中共鉴别 446,680 个 SNP 标记, 每个样本获得 85,159-225,683 个 SNP 标记, 平均每个样本获得 163,570 个 SNP 标记, 在团头鲂群体中得到 168,242 个 SNP, 在鲃鲂“先锋 2 号”群体中得到 161,395 个 SNP, 在黑尾近红鲂群体中得到 161,072 个 SNP。样本中 SNP 位点的杂合率为 9.48%-17.34%, 平均杂合率为 12.01%。根据选择标准, 从获得 SNP 标记中选择 446,680 个有效 SNP 进行群体遗传多样性分析。

3.3 群体多样性分析

根据 SNP 计算各遗传多样性参数的平均值, 得到的统计结果是观测等位基因数 (N_o) 为 2.00, 期望等位基因数 (N_e) 为 1.63, 观测杂合度 (H_o) 为 0.08, 期望杂合度 (H_e) 为 0.36, 多样性指数 (Genetic diversity index) 为 0.41, 香浓指数 (Shannon Wiener index) 为 0.54, 多态性信息含量 (PIC) 为 0.29, 次要等位基因频率 (MAF) 为 0.28。每个群体的多样性参数见表 1, 其中黑尾近红鲂的多样性最高, 其次是鲃鲂“先锋 2 号”, 团头鲂最低。

3.4 群体聚类与遗传距离分析

三个群体间遗传距离见表 2, 鲃鲂“先锋 2 号”和团头鲂的遗传距离小于鲃鲂“先锋 2 号”和黑尾近红鲂的遗传距离, 表明杂交鲃鲂“先锋 2 号”与母本团头鲂的亲缘关系更近, 遗传上更偏向于母本, 这与杂交选育过程中表现出来的形态分析结果相一致, 表明团头鲂对杂交子代的贡献率大于父本黑尾近红鲂。

Neighbor-joining (Kimura 2-parameter 模型, 1000 次 bootstrap) [6]进化树显示 30 尾鱼样本分为 3 个类群 (图 1)。

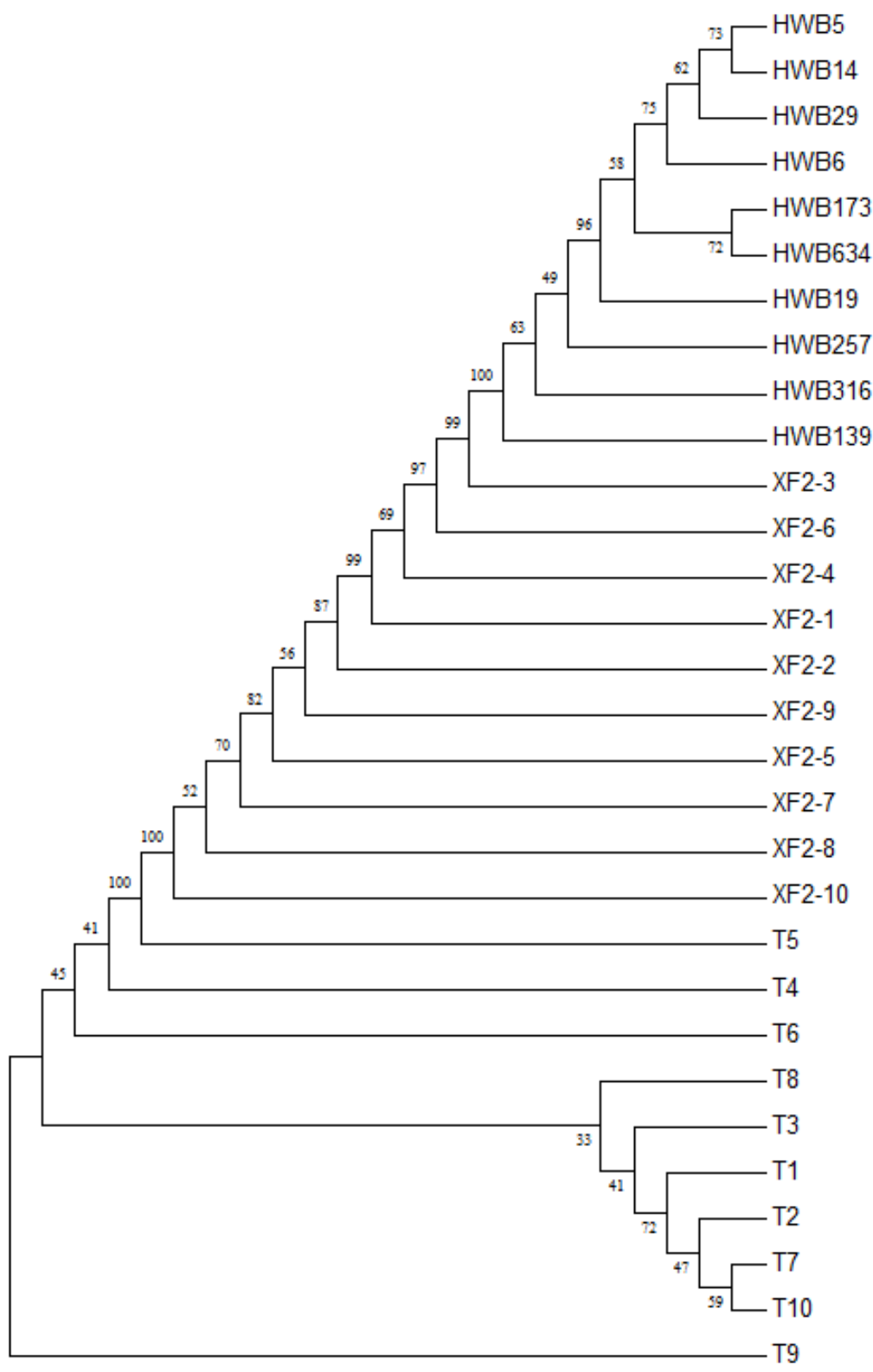


图 1 遗传进化树，进化树上的数字代表置信度，T 代表团头鲂，XF2 代表先锋鲂，HWB 代表黑尾近红鲌

Figure 1 The phylogenetic tree, the numbers on phylogenetic tree represent confidence level, T represents *Megalobrama amblycephala*, XF2 represents Bo Fang “Xianfeng No. 2”, HWB represents *Ancherythroculter nigrocauda*

表 1 三个群体的遗传多样性

Table 1 The genetic diversity of three populations

	SNP 个数	平均杂合率	观察杂合度	期望杂合度	多样性指数	香浓指数	多态信息含量	平均次要等位基因型频率
	SNP num	Heter ratio (%)	H_o	H_e	Nei_diversuty_index	Shnnon_Wiener_index	PIC	Average_MAF
TT	168,242	9.29	0.07	0.19	0.22	0.29	0.15	0.25
XF2	161,395	11.65	0.09	0.22	0.26	0.33	0.18	0.3
HWB	161,072	15.08	0.12	0.24	0.27	0.35	0.19	0.3

表 2 三种鱼间的遗传距离

Table 1 Genetic distance among the three populations

群体 Population	团头鲂 TTF	鲃鲂“先锋 2 号”XF2	黑尾近红鲂 HWB
团头鲂 TTF			
鲃鲂“先锋 2 号”XF2	0.34		
黑尾近红鲂 HWB	5.99	0.69	

3.5 群体遗传结构分析

群体遗传结构分析能够提供个体的血统来源及其组成信息，是一种重要的遗传关系分析工具。基于 SNP，通过 admixture 软件，分析样品的群体结构，分别假设样品的分群数（K 值）为 1-10，进行聚类。并对聚类

结果进行交叉验证，根据交叉验证错误率的谷值确定最优分群数。K 值为 1-10 的聚类情况及各个 K 值对应的交叉验证错误率见下图 2 和图 3，结果显示 K=2 时交叉验证错误率的数字小，表明最优分群数为 2，推测鲃鲂“先锋 2 号”与亲本群体间存在混合现象。

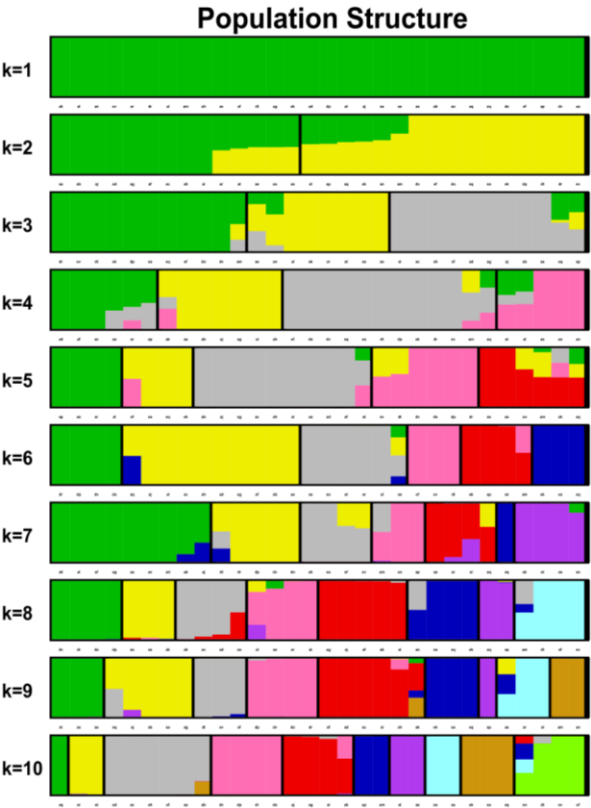


图 2 样品聚类结果

Figure 2 Samples clustering results

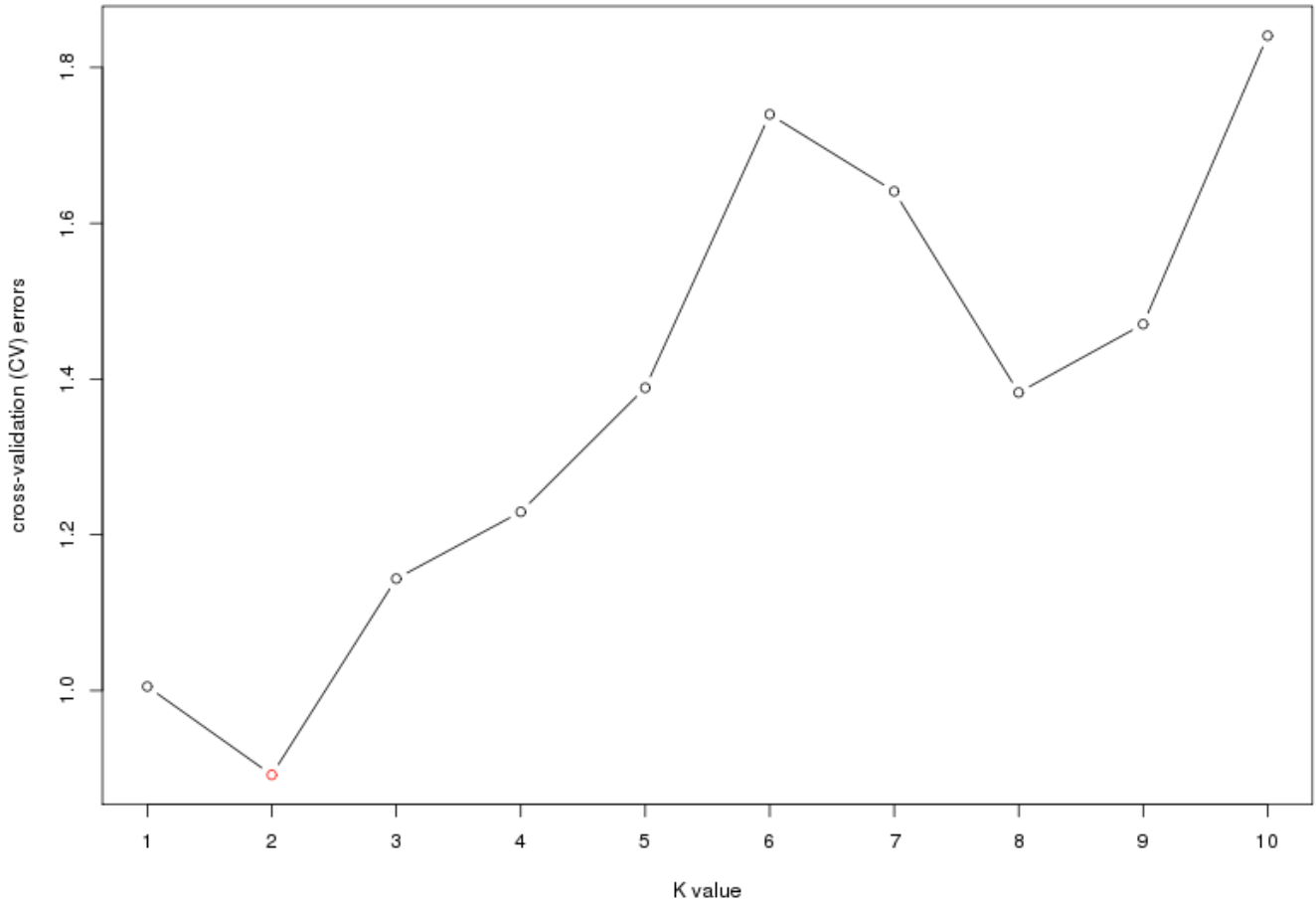


图 3 Admixture 各个 K 值的交叉验证错误率

Figure 3 The error of admixture K value by cross validation

4 讨论

4.1 SLAF-seq 技术在 SNP 分子标记开发中的可行性

SLAF-seq 是一种高通量的 SNPs 位点识别和分型技术，是基于生物信息学进行的高度自动化的对特定酶切片段进行双端测序的一种测序方法。由于 SNP 在 DNA 中呈现出分布均匀、广泛的特点，因此通过设置预酶切方案，对目的片段酶切测序后可实现候选功能区定位，且此方法具有较高的重复性，使研究者们能够轻易的找出所需 SNP。由于该技术具有全自动的特点，大大降低了成本。目前该技术已经高效、低成本的开发出大量的标记用于群体遗传分析[2]，遗传连锁图谱构建和重要性状分子标记定位[7-8]。在水产领域，SLAF-seq 技术已经应用到了多种鱼类的 SNP 开发，如

李光华等采用 SLAF-seq 技术开发了 777082 个短须裂腹鱼的 SNP 标记，证明了 SLAF-seq 技术能够较好的开发 SNP [2]；Qiu 等通过 SLAF-seq 技术对中华绒螯蟹进行了 SNP 位点的筛选，最终筛选出 127,677 个 SNP 候选位点用于后续试验[7]；蒋寿佳对 60 份大黄鱼进行 SLAF-seq 技术测序，获得了 151,253 个 SLAF 标签与 153,406 个 SNP 位点，最终筛选了 6 个与全体长相关的 SNP 用作后续研究[9]。本课题组利用 SLAF-seq 技术成功构建了黑尾近红鲃的遗传连锁图谱并发掘了跟生长相关的多个 QTL 位点[10]。经过大量的相关试验证明，SLAF-seq 技术由于其具备稳定可重复、准确率高、成本较低等特点已经成为了一种开发特异分子标记的理想手段被应用到分子研究领域的各个相关专业当中。

本研究利用 SLAF-seq 技术对鲃“先锋 2 号”、团头鲃、黑尾近红鲃共 30 个样本进行了 SNP 的开发，共获得了 1,339,007 个多态性的 SLAF 标签，以及分别在母本团头鲃群体中开发出 168,242 个 SNP 位点，在父本黑尾近红鲃群体中开发出 161,072 个 SNP 位点，在

鲌鲂“先锋 2 号”群体中开发出了 161,395 个 SNP 位点, 所获得的数据量足够用来进行特异性的 SNP 标记的开发与利用。

4.2 基于 SNP 位点的遗传多样性分析

本研究采用 SNP 标记的方式对鲌鲂“先锋 2 号”及其父母本进行群体遗传分析, 开发出了 446,680 个 SNP 位点, 遗传多样性分析结果表明, 本研究中 PIC 值为 0.29, 表明所选的位点均为中度多态性位点, 多态信息含量 (PIC) 反映了位点在群体中的多样性状况, 一般认为 $PIC > 0.5$ 时为高度多态性位点, $0.25 < PIC < 0.5$ 时为中度多态性位点, $PIC < 0.25$ 时为低度多态性位点 [11, 12]。 N_a 值为 2.00, N_e 值为 1.63, 表明其等位基因分布均匀, 适合遗传多样性研究。杂合度的高低反映了群体遗传一致性的程度, 基因杂合度又称为基因多样性, 是度量群体遗传变异的最合适参数之一, 其位点的平均杂合度近似反映了群体遗传变异程度的高低 [13, 14], 杂合度高的群体变异越大, 对环境的变化和自然选择的适应能力会越强 [15, 16]。本研究 H_o 数值在 0.07 至 0.12 之间, H_e 数值在 0.19 至 0.24 之间, 相对野生群体的鱼类多态性较低, 表明三种鱼都是经过了长期的人工选育, 群体变异降低, 这与许多经过选育的经济鱼类相似 [17, 18]。多样性指数是指群落丰度的指标, 指标越高, 该群落越稳定; 香农指数是用来反应群体的多样性的, 指数越大, 多样性越高。本研究的三种鱼多样性指数在 0.22 至 0.27 之间, 香农指数在 0.29 至 0.35 之间, 说明所选鲌鲂“先锋 2 号”群体、团头鲂群体及黑尾近红鲌群体的变异程度不高, 多样性偏低。

4.3 基于 SNP 位点的亲缘关系分析

从群体结构层次来分析, 由于在传统的选育过程中, 杂交子代的选择以形态学上接近母本团头鲂的个体为主, 通过对样本间遗传距离分析, 能够清晰的显示出三种经济鱼类遗传关系的远近, 并且结合聚类分析结果能够清晰的显示出杂交子代鲌鲂“先锋 2 号”更接近母本团头鲂, 与父本黑尾近红鲌相距较远。前期我们利用微卫星标记也对三种鱼做过遗传分析 [19], 与本研究的分析结果基本一致, 但是基因高通量测序获得的 SNP 标记数量多, 相对于传统分子标记, 能够更好反映全基因的遗传信息, 因此具有更高的准确性和稳定性。

5 结论

本文应用 SLAF-seq 开发了鲌鲂“先锋 2 号”、团头鲂及黑尾近红鲌的 SNP 标记并研究了群体遗传多样性和群体遗传结构, 结果表明, 所选鲌鲂“先锋 2 号”群体、团头鲂群体、黑尾近红鲌群体遗传多样性处于中等水平。三种鱼的遗传距离显示鲌鲂“先锋 2 号”与团头鲂最近、与黑尾近红鲌较远, 本研究与前期基于 SSR 标记对鲌鲂“先锋 2 号”、团头鲂、黑尾近红鲌三种鱼进行的群体遗传分析, 结果保持一致。

参考文献

- [1] Sambrook J, Russell DW. Molecular Cloning: A Laboratory Manual, 3rd ed [M]. Cold Spring Harbor Laboratory Press, New York, 2001: 463-446.
- [2] Sun X, Liu D, Zhang X, et al. SLAF-seq: an efficient method of large-scale De novo SNP discovery and genotyping using high-throughput sequencing [J]. PloS one, 2013, 8 (3): e58700.
- [3] 李光华, 金方彭, 周睿, 等. 基于 SNP 标记的短须裂腹鱼自然群体遗传多样性分析 [J]. 水生生物学报, 2018, 42 (2): 271-276.
- [4] Alexander D H, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals [J]. Genome Research, 2009, 19 (9): 1655-1664.
- [5] Tamura K., Peterson D., Peterson N., et al. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods [J]. Molecular Biology and Evolution, 2011, 28 (10): 2731-2739.
- [6] Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences [J]. Journal of Molecular Evolution, 1980, 16: 111-120.
- [7] Qiu G F, Xiong L W, Han Z K, et al. 2017. A second generation SNP and SSR integrated linkage map and QTL mapping for the Chinese mitten crab *Eriocheir sinensis* [J]. Scientific Reports, 2017, 7, 39826.
- [8] 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, 15612.
- [9] 蒋寿佳. 大黄鱼生长性状相关 SNP 标记的研究 [硕士学位论文]. 舟山: 浙江海洋大学图书馆, 2015.

- [10] Sun Y H, Qing Li Q, Wang G Y, et al. Polymorphisms in the Myostatin-1 gene and their association with growth traits in *Ancherythroculter nigrocauda*. Chinese Journal of Oceanology and Limnology, 2017, 35: 597-602.
- [11] 侯宁, 张研, 鲁翠云, 等. 微卫星 DNA 标记分析德国镜鲤的遗传潜力 [J]. 遗传, 2007, 29 (12): 1509-1518.
- [12] Wright. Evolution and the Genetics of Populations [M]. Chicago: University of Chicago Press. 1978.
- [13] Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals [J]. Genetics, 1978, 89 (3): 583-590.
- [14] Nei M, Maruyama T, Chakraborty R. The bottleneck effect and genetic variability in populations [J]. Evolution, 1975, 29 (1): 1-10.
- [15] Beardmore J A, Mair G C, Lewis R I. Biodiversity in aquatic systems in relation to aquaculture [J]. Aquaculture Research, 1997, 28 (10): 829-839.
- [16] Xu Z, Primavera J P, Pena L D, et al. Genetic diversity of wild and cultured Black Tiger Shrimp (*Penaeus monodon*) in the Philippines using microsatellites [J]. Aquaculture, 2001, 199 (1-2): 13-40.
- [17] 赵广泰, 刘贤德, 王志勇, 等. 大黄鱼连续 4 代选育群体遗传多样性与遗传结构的微卫星分析 [J]. 水产学报, 2010, 34 (4): 500-507.
- [18] 张进, 梁旭芳, 杨敏, 等. 2 个鳊鱼选育群体遗传多样性分析 [J]. 水产科学, 2014, 33 (7): 447-450.
- [19] 孙艳红, 马浩, 魏辉杰, 等. 基于 SSR 标记的鲃鲂“先锋 2 号”、团头鲂及黑尾近红鲃的群体遗传分析 [J]. 淡水渔业, 2022, 52 (1): 30-36.

作者简介

李佩

1986 年生, 硕士, 工程师, 研究方向为鱼类育种及健康养殖技术.

E-mail: 305201878@qq.com

孙艳红

1982 年生, 博士, 副研究员, 研究方向为鱼类育种技术.

E-mail: 176570314@qq.com