

不同生态类型菘蓝 SLAF-seq 分子标记

刘 东, 孟瑾瑾, 王 盼, 陈红刚, 王惠珍, 杜 弢*

甘肃中医药大学药学院, 甘肃 兰州 730000

摘 要: **目的** 探究不同生态类型菘蓝 *Isatis tinctoria* 表型差异和特异位点扩增片段测序 (specific-locus amplified fragment sequencing, SLAF-seq) 分子标记。**方法** 以 25 个不同生态类型菘蓝为研究对象, 利用 SLAF-seq 技术开发 SLAF 标签, 再结合菘蓝叶片叶表观特征的田间观测, 分析不同生态类型菘蓝植株表观差异。**结果** 测序后各样品所获得的读长总量各不相同, 测序质量值 Q30 为 95.66%~96.60%, GC 比例为 38.13%~41.37%。共开发 535 105 个菘蓝 SLAF 标签, 多态性 SLAF 标签有 34 412 个, 共得到 63 002 个群体单核苷酸多态性 (single nucleotide polymorphism, SNP) 标记。测序分析结果与叶表观特征调查结果相结合, 可以将 25 个不同生态类型菘蓝分为 5 类, 分别为倒卵圆形无缺刻皱缩叶、倒卵圆形缺刻皱缩叶、倒卵圆形无缺刻平坦叶、倒卵圆形缺刻平坦叶、披针形无缺刻平坦叶。**结论** 为后期进一步研究不同类型菘蓝的基因来源和遗传进化提供参考, 为菘蓝种质分类、品种鉴定和分子辅助育种提供借鉴。

关键词: 菘蓝; 生态类型; SLAF 标签; SNP; 群体结构

中图分类号: R286.12 **文献标志码:** A **文章编号:** 0253-2670(2024)07-2375-08

DOI: 10.7501/j.issn.0253-2670.2024.07.023

SLAF-seq molecular markers in different ecological types of *Isatis tinctoria*

LIU Dong, MENG Jinjin, WANG Pan, CHEN Honggang, WANG Huizhen, DU Tao

College of Pharmacy, Gansu University of Chinese Medicine, Lanzhou 730000, China

Abstract: Objective To explore the phenotypic differences and specific-locus amplified fragment sequencing (SLAF-seq) molecular markers of *Isatis indigogenes* from different ecological types. **Methods** Using twenty-five different ecological types of *Isatis indigotica* as research objects, SLAF labels were developed by SLAF-seq technique, and the apparent differences of *Isatis indigotica* plants were analyzed by field observation of leaf apparent characteristics. **Results** After sequencing, the total read length of each sample was different, the sequencing quality value Q30 was 95.66%—96.60%, and the GC ratio was 38.13%—41.37%. A total of 535 105 *Isatis indigotica* SLAF labels and 34 412 polymorphic SLAF labels were developed, resulting in 63 002 single nucleotide polymorphism (SNP) markers. Combining the results of sequencing analysis with the investigation of leaf apparent characteristics, the 25 different ecological types of *Isatis indigotica* could be divided into five categories, namely, obovate without engraved wrinkled leaves, obovate with engraved wrinkled leaves, obovate without engraved flat leaves, obovate with engraved flat leaves, lanceolate without engraved flat leaves. **Conclusion** It provides reference for further study on gene origin and genetic evolution of different types of *Isatis indigotica*, and provides reference for germplasm classification, variety identification and molecular assisted breeding of *Isatis indigotica*.

Key words: *Isatis tinctoria* L.; ecological type; SLAF label; SNP; population structure

随着分子生物技术水平的不断提高, 分子标记技术应用越来越广泛, 人们常采用随机扩增多态性 DNA 标记 (random amplified polymorphic DNA, RAPD) [1-3]、相关序列扩增多态性 (sequence related

amplified polymorphism, SRAP) [4-6]、限制性内切酶片段长度多态性 (restriction fragment length polymorphism, RFLP) [7]、扩增片段长度多态性 (amplified fragment length polymorphism, AFLP) [8-9]、

收稿日期: 2023-11-02

基金项目: 国家中药材产业技术体系建设专项资金资助; 国家重点研发计划专项“乡村产业共性关键技术研发与集成应用”课题 (2022YFD1602105)

作者简介: 刘 东 (1994—), 男, 陕西凤县人, 博士研究生, 研究方向为药用植物栽培和育种研究。E-mail: liudong804214@163.com

*通信作者: 杜 弢, 博士生导师, 教授, 从事药用植物栽培和育种研究。E-mail: gslzdt@163.com

简单重复序列标记 (simple sequence repeats, SSR) [10-12]、ISSR 标记 (inter-simple sequence repeat, ISSR) [13-15] 和表达序列标签 SSR (expressed sequence tags-SSR, EST-SSR) [16] 等分子标记技术进行遗传多样性、亲缘关系及遗传演化等方面的研究, 并取得很大进展, 但这些分子标记技术都有一定的缺陷, RAPD 技术易受到模板的质量和浓度、引物序列短、PCR 循环次数、基因组 DNA 的复杂性等方面的影响, 造成 RAPD 技术的重复性差; AFLP 技术操作复杂、成本较高, 对基因组纯度和反应条件要求较高, ISSR 标记对体细胞遗传变异的检测效率低, 因此, 由不同分子生物技术所得结果也存在些许差异。

特异位点扩增片段测序 (specific-locus amplified fragment sequencing, SLAF-seq) 是基于高通量测序技术发展起来的一种简化基因组测序技术, 具有通量高、准确性高、成本低、周期短的技术特点 [17]。SLAF-seq 技术是通过生物信息学方法设计最佳酶切方案, 构建 SLAF 文库, 筛选特异性长度的酶切片段 (SLAF 标签), 利用高通量测序技术获得大量序列, 通过软件比对分析, 开发大量的全基因组范围的特异性 SNP 标记, 利用大量的 SNP 标记进行亲缘关系及遗传变异分析 [18-19]、性状关联分析 [20] 和构建遗传

连锁图谱 [21] 等工作, 同时为分子育种、系统进化和种质资源鉴定等提供技术保障 [22-23]。李贝贝等 [24] 利用 SLAF-seq 技术进行了 304 份葡萄种质遗传多样性分析, 实验结果表明, 我国地方品种亲缘关系较近, 证实了 ‘小黑葡萄’ 是野生种的一个过渡类型的推测。Kozich 等 [25] 利用 SLAF-seq 简化基因组测序技术对甘薯栽培种及不同倍性的近缘野生种的遗传多样性和进化关系进行分析, 实验结果表明, 栽培种和六倍体近缘野生种的遗传关系最近, 进一步验证了栽培甘薯可能来源于 *Ipomoea trifida* (6X)。目前, 采用 SLAF-seq 技术在菰蓝上开发大量 SNP 位点的相关研究鲜见报道。因此, 本实验利用 SLAF-seq 技术测定了 25 个不同生态类型菰蓝材料的基因序列, 获得大量的 SLAF 标签, 进而开发一批稳定性高、特异性强的 SNP 位点。基于开发的 SNP 标记对不同生态类型菰蓝间遗传多样性及群体结构进行分析, 旨在从基因水平揭示不同生态类型之间的遗传关系。

1 材料

1.1 材料信息

25 个不同生态类型菰蓝的种子均是从全国各地收集而来, 均为人工栽培品, 引种地信息见表 1, 所有种子经甘肃中医药大学杜弢教授鉴定为十字花

表 1 菰蓝种子编号及叶片特征信息

Table 1 Table of seed number and leaf characteristic information of *Isatis tinctoria*

序号	编号	引种地	叶片表现特征
1	2016-1	陕西咸阳	叶片倒披针型; 叶顶端尖; 叶面平坦, 中度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 无波状
2	2016-2	河北石家庄	叶片倒卵圆型; 叶顶端钝尖; 叶面皱缩, 轻度蜡质; 叶柄基部浅绿色; 叶缘轻度缺刻, 中度波状
3	2016-3	内蒙古	叶片倒卵圆型; 叶顶端尖; 叶面平坦, 中度蜡质; 叶柄基部绿色; 叶缘无缺刻, 轻度波状
4	2016-4	河南洛阳	叶片倒卵圆型; 叶顶端尖; 叶面平坦, 轻度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 无波状
5	2016-5	河南三门峡	叶片倒卵圆型; 叶顶端尖; 叶面平坦, 中度蜡质; 叶柄基部绿色; 叶缘无缺刻, 无波状
6	2016-6	山东聊城	叶片倒卵圆型; 叶顶端钝尖; 叶面平坦, 轻度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 无波状
7	2016-7	甘肃张掖	叶片倒披针型; 叶顶端钝尖; 叶面平坦, 中度蜡质; 叶柄基部绿色; 叶缘无缺刻, 轻度波状
8	2016-8	山西浑源	叶片倒披针型; 叶顶端钝尖; 叶面平坦, 轻度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 无波状
9	2016-9	宁夏固原隆德县	叶片倒卵圆型; 叶顶端尖; 叶面平坦, 中度蜡质; 叶柄基部浅绿色; 叶缘轻度缺刻, 轻度波状
10	2016-13	甘肃定西	叶片倒披针型; 叶顶端钝尖; 叶面平坦, 中度蜡质; 叶柄基部浅绿色; 叶缘轻度缺刻, 无波状
11	2016-14	河北安国	叶片倒卵圆型; 叶顶端钝尖; 叶面平坦, 中度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 无波状
12	2016-15	山西万荣	叶片倒卵圆型; 叶顶端钝尖; 叶面平坦, 中度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 轻度波状
13	2016-16	山东莒县	叶片倒卵圆型; 叶顶端钝尖; 叶面平坦, 中度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 无波状
14	2016-17	甘肃渭源	叶片倒卵圆型; 叶顶端钝尖; 叶面平坦, 中度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 轻度波状
15	2016-18	云南昆明	叶片倒卵圆型; 叶顶端尖; 叶面平坦, 中度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 无波状
16	2016-19	安徽亳州	叶片倒卵圆型; 叶顶端钝尖; 叶面平坦, 轻度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 无波状
17	2016-20	河北涉县	叶片倒卵圆型; 叶顶端尖; 叶面平坦, 中度蜡质; 叶柄基部绿色; 叶缘无缺刻, 无波状
18	2016-21	山西晋中市榆社县	叶片倒卵圆型; 叶顶端钝尖; 叶面平坦, 中度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 轻度波状
19	2016-22	甘肃陇西	叶片倒卵圆型; 叶顶端钝尖; 叶面平坦, 中度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 无波状
20	2016-23	甘肃渭源	叶片倒披针型; 叶顶端尖; 叶面平坦, 中度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 无波状
21	2016-24	甘肃定西安定区	叶片倒披针型; 叶顶端钝尖; 叶面平坦, 少度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 轻度波状
22	2016-25	山东东平县	叶片倒披针型; 叶顶端钝尖; 叶面平坦, 中度蜡质; 叶柄基部绿色; 叶缘无缺刻, 中度波状
23	2016-26	山东平邑县	叶片倒卵圆型; 叶顶端钝尖; 叶面平坦, 中度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 无波状
24	2016-27	甘肃成县	叶片倒卵圆型; 叶顶端钝尖; 叶面皱缩, 中度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 轻度波状
25	2018-1	河北石家庄	叶片倒卵圆型; 叶顶端钝尖; 叶面皱缩, 轻度蜡质; 叶柄基部浅绿色; 叶缘无缺刻, 轻度波状

科菘蓝属植物菘蓝 *Isatis tinctoria* L. 的种子。试验于甘肃中医药大学和政药用植物园实施, 园内海拔 2 430 m, 年均降水量 660 mm, 年均气温 5.1 °C, 无霜期 110 d。土壤为肥力均一的黑垆土。

1.2 试验方法

将 25 个不同生态类型菘蓝种子条播于甘肃中医药大学和政药用植物园, 各试验小区随机排列, 3 次重复, 每个试验小区的面积为 19.58 m²。条播用种量为 1.2 g/m², 沟深 5 cm, 沟宽 10 cm, 行距 30 cm。6 月 17 日出苗, 6 月 27 日一次性完成间苗, 株距 10 cm; 10 月 25 日观测各个试验小区叶片表观特征。在每个试验小区内随机挑选 5 株, 从每株相同部位采取幼嫩叶片并将其混装, 1 个试验小区形成 1 份混样, 并对其编号, 迅速投入液氮中速冻保存。

2 方法

2.1 基因组 DNA 提取和检测

用 CTAB 法从预先保存的叶片中提取基因组 DNA, 用传统的 1% 琼脂糖凝胶电泳法检测提取基因的完整性, 并用 NanoDrop 1000C 微量紫外分光光度计检测 DNA 的浓度及纯度, 确保提取的 DNA 质量满足测序要求。采用超纯水将质量浓度统一调至 50 ng/μL, -20 °C 保存备用。

2.2 酶切建库

根据最优酶切方案的选择原则, 应用酶切预测软件进行酶切预测, 确定最优的酶切方案, 当最适酶切方案确定之后, 对已检测合格的样品基因组 DNA 进行酶切建库。对得到的酶切片段 (SLAF 标签) 进行 3' 端加 A 处理、连接 Dual-index 测序接头、PCR 扩增、DNA 纯化、混样、切胶选取目的片段。

2.3 基因组测序及数据质量评估

经文库质检合格后, 利用 Illumina 平台进行测序^[26]。为了评估酶切实验方案和酶切结果的准确性, 选用水稻日本晴 *Oryza sativa indica* L. 作酶切方法和数据质量的评估对照。Dual-index 识别测序所得的原始数据, 进行过滤优选得到高质量双端 Clean reads, 再进行测序质量和数据量的质量评估。

2.4 SLAF 标签和 SNP 分析

根据生物信息学分析, 在群体中开发全基因组范围的 SLAF 标签, 以每个 SLAF 标签中深度最高的序列类型作为参考序列, 利用 BWA^[27] 将测序 reads 比对到参考序列上, 并使用 GATK^[28] 和 SAMtools^[29] 2 种软件开发 SNP, 以 2 种软件开发获

得的 SNP 交集作为最终可靠的 SNP 标记数据。

2.5 群体遗传结构和亲缘关系分析

将开发的 SNP 标记根据完整度 > 0.8, 次要等位基因频率 (MAF) > 0.05 的标准进行过滤^[30], 基于筛选出的高一致性的群体 SNP, 使用统计软件 MEGA X^[31] 和 Admixture^[32] 对群体进行系统进化树和群体结构分析。

2.6 主成分分析 (principal component analysis, PCA)

PCA 是一种统计方法。在群体遗传学上, 主要基于不同材料的 SNP 差异程度 (或分歧程度) 聚类成不同亚群, 其结果可以用于与其他聚类方法的结果相互验证。基于 SNP, 通过 EIGENSOFT 软件进行 PCA。

3 结果与分析

3.1 叶片表观特征分析

形态学特征是植物遗传差异的最直观表现, 田间栽培验证也是最简单的遗传变异的鉴别方法。田间植株叶片表观特征调查结果如表 1 所示, 对比分析发现 25 个不同生态类型菘蓝叶片表观特征可以概述如下: 叶形分为倒卵圆形和倒披针形 2 种, 叶尖分为尖和钝尖 2 种, 叶片表面都有蜡粉层, 均是蜡质叶, 叶柄基部根据颜色分为浅绿色和绿色 2 种, 叶缘根据缺刻程度分为无缺刻 (全缘叶) 和轻度缺刻 2 种, 观察叶片整体形态特征, 分为无波状叶、轻度波状叶和中度波状叶。

3.2 建库评估

酶切效率是评价简化基因组试验是否成功的一个关键指标, 测序 reads 插入片段中残留酶切位点的比例越高, 酶切效率越好^[33]。本实验经过 SLAF 电子酶切预测软件分析, 确定限制性内切酶切组合为 RsaI-HaeII, 酶切片段长度在 314~364 bp 的序列定义为 SLAF 标签。为进一步评估酶切方案的有效性与酶切效率, 以水稻日本晴测序数据为酶切方案和数据质量评估对照, 利用 SOAP^[34] 软件将对照的测序 reads 与其参考序列进行比对, 试验双端比对效率的结果为 96.09%, 正常酶切比例为 92.39%, 说明酶切反应处于正常, 此次试验选定的酶切方案正确可行, SLAF 建库正常。

3.3 测序质量评估

本实验 25 个样品共获得 63.32 兆数据量, 菘蓝及对照测序结果如表 2 所示。各样品所得到的读长总量各不相同, 最大的读长总量为 6 086 419

表 2 菰蓝及对照测序结果

Table 2 <i>I. tinctoria</i> and control sequencing results			
样品	读长总量	GC 碱基比例/%	Q_{30} 比例/%
2016-1	2 460 866	40.25	96.60
2016-2	1 854 755	40.83	96.50
2016-3	1 938 556	40.39	95.66
2016-4	3 344 111	39.50	95.92
2016-5	2 356 307	39.58	96.41
2016-6	2 492 201	38.90	95.85
2016-7	2 074 529	40.93	96.53
2016-8	6 086 419	40.80	96.59
2016-9	2 972 700	39.95	96.57
2016-13	3 533 779	41.02	95.93
2016-14	2 821 422	39.47	95.75
2016-15	2 375 466	41.37	96.55
2016-16	2 361 488	39.76	96.49
2016-17	2 293 398	40.56	96.44
2016-18	1 789 757	40.87	96.18
2016-19	2 948 053	39.46	96.20
2016-20	3 355 413	39.64	95.74
2016-21	1 779 629	40.42	96.46
2016-22	1 498 688	40.91	96.44
2016-23	1 867 088	40.11	96.47
2016-24	2 635 901	38.76	96.33
2016-25	2 369 324	38.94	96.76
2016-26	1 739 391	40.40	96.51
2016-27	1 830 400	40.44	96.48
2018-1	2 549 081	39.89	96.45
水稻日本晴	327 810	38.13	96.38
平均值	—	40.13	96.31

(2016-8), 最小的读长总量为 1 498 688 (2016-22)。测序的 GC 碱基比例的平均值为 40.13%, 最小的 GC 碱基比例为 38.76% (2016-24), 最大的 GC 碱基比例为 41.37% (2016-15), 此次测序所有样品 GC 碱基比例均比较高, 说明测序达到要求。测序质量值 (Q) 是对测序数据质检的关键指标, 是测序下机后单碱基错误率大小的具体评判标准, 一般测序质量值用 Q_{30} 表示。此次试验测序质量值 Q_{30} 均大于 95%, 平均 Q_{30} 为 96.31%, 最小的 Q_{30} 为 95.66% (2016-3), 最大的 Q_{30} 为 96.60% (2016-1), 说明此次测序碱基错误率较低, 测序质量较好。结合读长总量、GC 碱基比例及 Q_{30} 比例等各个结果的数据, 可以得出此次测序质量较高, 所得数据结果可靠。

3.4 SLAF 标签开发及 SNP 信息统计

经过生物信息学等各方面分析, 此次试验的每个样品标签平均测序深度为 23.52X, 共获得 535 105 个菰蓝 SLAF 标签, 具体的菰蓝 SLAF 标签数统计结果如表 4 所示, 其中共获得 34 412 个多态性 SLAF 标签。以每个 SLAF 标签中深度最高的序列类型作为参考序列, 进行 SNP 位点开发, 此次实验共得到 63 002 个群体 SNP 标记。SNP 信息统计结果见表 3。

表 3 菰蓝 SLAF 标签和 SNP 统计结果

Table 3 Results of <i>I. tinctoria</i> SLAF label and SNP statistics						
样品编号	SLAF 标签数	测序总深度	平均深度	SNP 总数	SNP 数量	杂合率/%
2016-1	82 812	2 289 009	27.64	63 002	26 725	13.29
2016-2	87 732	1 715 487	19.55	63 002	22 468	9.38
2016-3	107 977	1 731 078	16.03	63 002	32 560	12.06
2016-4	125 480	3 065 923	24.43	63 002	33 306	14.79
2016-5	82 931	2 183 700	26.33	63 002	26 579	14.11
2016-6	93 083	2 279 349	24.49	63 002	27 990	13.54
2016-7	105 535	1 925 838	18.25	63 002	29 285	13.37
2016-8	114 585	5 688 530	49.64	63 002	30 731	23.02
2016-9	100 078	2 754 495	27.52	63 002	29 727	12.50
2016-13	101 242	3 232 809	31.93	63 002	29 359	12.93
2016-14	97 098	2 581 796	26.59	63 002	30 196	12.63
2016-15	106 372	2 206 597	20.74	63 002	28 824	13.85
2016-16	99 711	2 192 906	21.99	63 002	28 058	11.57
2016-17	98 315	2 124 543	21.61	63 002	29 911	16.39
2016-18	99 449	1 642 442	16.52	63 002	29 187	18.93
2016-19	128 620	2 694 791	20.95	63 002	33 294	15.17
2016-20	100 798	3 091 517	30.67	63 002	30 434	21.91
2016-21	84 260	1 654 899	19.64	63 002	27 275	18.25
2016-22	98 246	1 379 069	14.04	63 002	27 660	17.24
2016-23	80 463	1 729 548	21.49	63 002	25 489	10.64
2016-24	114 687	2 412 839	21.04	63 002	31 912	15.75
2016-25	81 977	2 208 708	26.94	63 002	25 827	11.16
2016-26	79 575	1 612 755	20.27	63 002	26 383	11.21
2016-27	103 993	1 683 476	16.19	63 002	30 357	18.59
2018-1	100 137	2 361 838	23.59	63 002	29 556	17.91

3.5 系统发育树分析

基于筛选出的高一致性 SNP 标记,通过 MEGA X 软件,基于邻接法 (neighbor-joining, NJ),采用 Kimura 2-parameter 模型,bootstrap 重复 1 000 次,构建各样品的系统发育树,如图 1 所示。25 份不同生态类型菘蓝聚成 3 组,第 1 组为 2016-22 和 2016-26,第 2 组为 2016-3、2016-4、2016-5、2016-16、2016-20、2016-21,第 3 组为 2016-8、2016-14、2016-15、2016-1、2016-2、2016-6、2016-18、2016-19、2016-27、2018-1、2016-7、2016-9、2016-13、2016-17、2016-23、2016-24、2016-25。

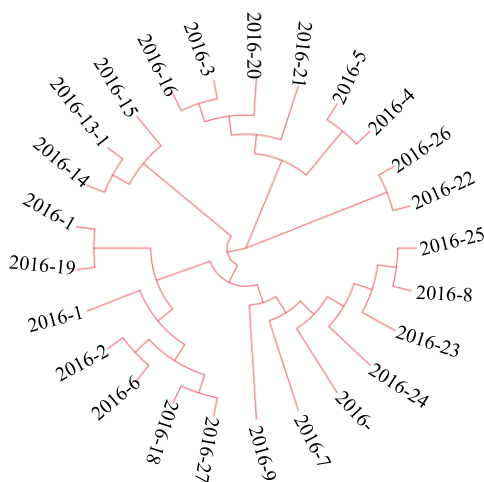


图 1 不同生态类型菘蓝系统发育树

Fig. 1 Phylogenetic trees of *I. tinctoria* of different ecological types

3.6 群体结构分析

根据试验筛选出的高一致性 SNP 标记,利用 Admixture 软件进行群体结构分析,经过交叉验证, 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结果如图 2、图 3 所示。当 $K=1$ 时,25 个不同生态类型菘蓝基因来源为一类;当 $K=10$ 时,25 个不同生态类型菘蓝可以分为 10 类。由于每个类群都存在交叉现象,说明每个类群间的亲缘关系相对较近。类群 1 主要是 2016-3 和 2016-9,类群 2 主要是 2016-1 和 2016-23,类群 3 主要是 2016-4、2016-5 和 2016-26,类群 4 主要是 2016-24 和 2016-25,类群 5 主要是 2016-13,类群 6 主要是 2016-6 和 2016-16,类群 7 主要是 2016-14 和 2016-15,类群 8 主要是 2016-19 和 2016-22,类群 9 主要是 2016-2 和 2016-7,类群 10 主要是 2016-18、2016-27 和 2018-1。而 2016-8、2016-17、2016-20 和 2016-21 这 4 个生态类型的游离在各个类群之间,2016-8 相比较而言,较为接近类群 10 (遗传特征值为 0.373 1),2016-17 相比较而言,较为接近类群 1 (遗传特征值为 0.261 3) 和类群 4 (遗传特征值为 0.245 9),2016-20 相比较而言,较为接近类群 1 (遗传特征值为 0.330 7) 和类群 3 (遗传特征值为 0.297 8),2016-21 相比较而言,较为接近类群 1 (遗传特征值为 0.405 3)。但是,根据 Admixture 各个 K 值的交叉验证错误率折线图走势及结果可以看出, $K=1$ 是应为最优 K 值。因此,25 个不同生态类型的菘蓝样品群体关系为均来自同一个祖先。

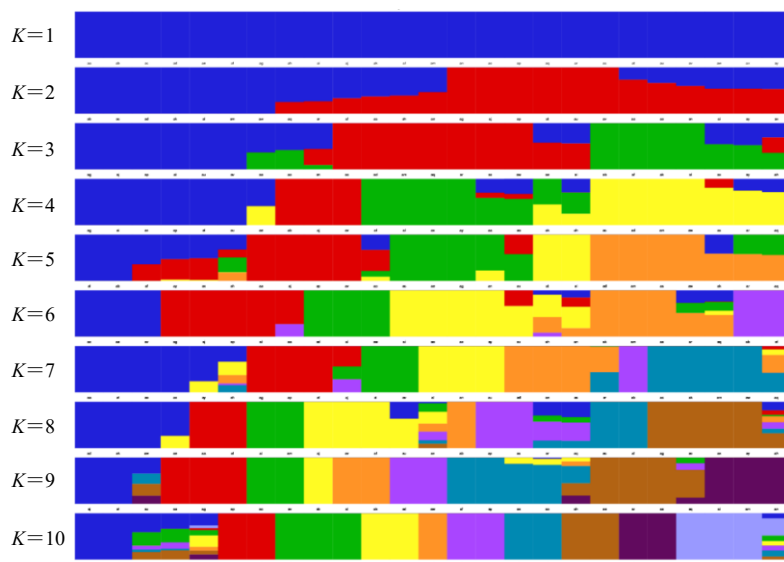
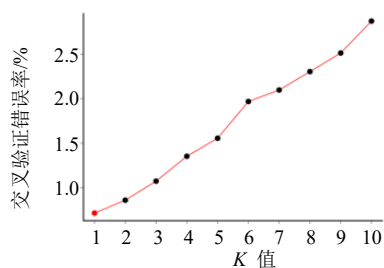


图 2 Admixture 各个 K 值对应的样品聚类结果

Fig. 2 Sample clustering results corresponding to K values of Admixture

图3 Admixture 各个 K 值的交叉验证错误率Fig. 3 Cross verification error rate of Admixture for each K value

3.7 PCA

基于 SNP, 通过 EIGENSOFT 软件, 进行 PCA, 得到样品的主成分聚类情况和具体数据如图 4 所示。通过 PCA 三维聚类图可知, 25 个样品的 SNP 结果聚为 1 个分组, 同时也佐证了“3.6”项群体结构分析的结果, 25 个不同生态类型菘蓝为同一个祖先, 拥有共同的血缘来源。

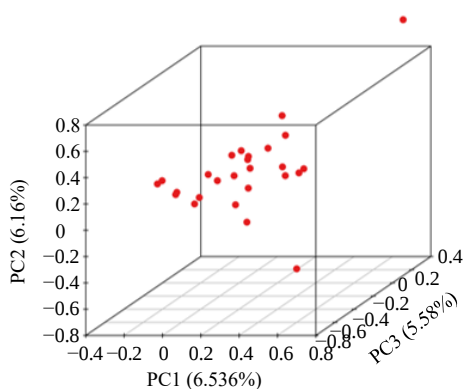


图4 不同生态类型菘蓝 PCA 三维聚类图

Fig. 4 PCA three-dimensional cluster map of *I. tinctoria* of different ecological types

4 讨论

十字花科菘蓝属有 7 种菘蓝, 分别是长圆果菘蓝 *I. oblongata* DC.、宽翅菘蓝 *I. violascens* Bunge、毛果菘蓝 *I. tinctoria* L. var. *praecox* (Kit.)、欧洲菘蓝 *I. tinctoria* L.、三助菘蓝 *I. costata* C. A. Mey.、菘蓝 *I. tinctoria* L.、小果菘蓝等 *I. minima* Bunge。根据《中国植物志》中记载的各种菘蓝形态特征描述, 比对表 1 中田间调查记录结果, 结合对 25 个不同生态类型菘蓝种植情况, 可以排除不是宽翅菘蓝、小果菘蓝、毛果菘蓝、长圆果菘蓝、欧洲菘蓝和三助菘蓝, 因为 25 个不同生态类型菘蓝均为二年生、茎及基生叶背面不带紫红色、植株被不具有白色柔毛、角果不着生柔毛和不具

有中脉 3 棱的特点, 因此, 可以确定此 25 个不同生态类型的菘蓝是中国特有种菘蓝, 与群体结构和 PCA 的结果一致。根据表 1 中叶片表观特征描述, 25 个不同生态类型菘蓝可以分为 5 钟类型, 第 1 种类型为倒卵圆形无缺刻皱缩叶 (2016-27、2018-1), 第 2 种类型为倒卵圆形缺刻皱缩叶 (2016-2), 第 3 种类型为倒卵圆形无缺刻平坦叶 (2016-3、2016-4、2016-5、2016-6、2016-14、2016-15、2016-16、2016-17、2016-18、2016-19、2016-21、2016-22、2016-26), 第 4 种类型为倒卵圆形缺刻平坦叶 (2016-9), 第 5 种类型为披针形无缺刻平坦叶 (2016-1、2016-7、2016-8、2016-13、2016-20、2016-23、2016-24、2016-25)。

分子生物技术是现代生物学研究不可或缺的技术手段。本研究利用 SLAF-seq 技术在菘蓝全基因组范围内成功开发出了大量 SNP 位点, 利用得到的 63 002 个群体 SNP 标记分析了不同生态类型菘蓝种质的群体遗传多样性, 根据 SNP 群体遗传结构、PCA 及田间观测记录, 25 个不同生态类型的菘蓝为同一种属, 具有共同的祖先和血清来源。SNP 系统发育树将 25 个不同生态类型菘蓝聚为 3 类, 第 1、2 类对应田间观测调查的倒卵圆形无缺刻平坦叶, 而第 3 类涵盖了田间观测调查的 5 种类型。在 SNP 群体遗传结构的分析结果中, 2016-8、2016-17、2016-20 和 2016-21 这 4 个生态类型的游离在多个类群之间, 每一个类型都出现同属于几个类群的现象, 结合叶片表观特征调查的结果发现, 这 4 个类型的菘蓝均为倒披针形无缺刻平坦叶, 并且种子的引种地并不在同一个省份和地区, 说明 2016-8、2016-17、2016-20 和 2016-21 在市场流通比较频繁, 为药农和种植户所常选类型之一, 因此在频繁的引种交换过程中曾可能出现过与其他类群之间的基因交流, 或者经受各地不同环境气候的自然选择, 从而导致类群归类不具体的现象。

综上所述, 不同生态类型菘蓝形态上的差异可能来源于种内遗传差异, 差异的形成原因可能是由于菘蓝在国内各地都是采用种子直播的方式人工栽培, 并且种子大多都是自繁自用或区域性调种, 很少出现大范围的调种现象, 但是我国东西南北气候和环境差异较大, 因此菘蓝种群的种内基因大范围交流受阻, 外加各地气候和环境对其的自然选择进化, 形成了具有区域特点的不同生态类型, 表现了菘蓝的遗传

多样性。目前,根据现有结果和数据,可以将 25 个同一种属不同生态类型菰蓝分为倒卵圆形无缺刻皱缩叶、倒卵圆形缺刻皱缩叶、倒卵圆形无缺刻平坦叶、倒卵圆形缺刻平坦叶、倒披针形无缺刻平坦叶 5 个类群。此研究为后期进一步研究不同类型菰蓝的基因来源和遗传进化提供参考,为菰蓝种质分类、品种鉴定和分子辅助育种提供借鉴。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 唐首杰, 毕洋, 张飞明, 等. 连续三代减数分裂雌核发育团头鲂的遗传多样性分析和 RAPD 鉴别方法的建立 [J]. 浙江农业学报, 2019, 31(8): 1257-1271.
- [2] 徐雯, 瞿印权, 张玲玲, 等. 基于 RAPD 的福建产南方红豆杉遗传多样性研究 [J]. 中草药, 2017, 48(14): 2943-2949.
- [3] Shah U N, Mir J I, Ahmed N, *et al.* Genetic diversity analysis of walnut (*Juglans regia* L.) from Kashmir valley using RAPD and ISSR markers [J]. *Agrotechnology*, 2019, 8(1): 89-92.
- [4] Zhang X T, Gong P J, Shi Y, *et al.* Genetic inter-relationships among Chinese wild grapes based on SRAP marker analyses [J]. *Vitis J Grapevine Res*, 2018, 57: 151-157.
- [5] 吴红, 常永义, 郝燕, 等. 利用 RAPD 标记研究部分葡萄品种的亲缘关系 [J]. 果树学报, 2008, 25(6): 932-936.
- [6] 何庆元, 刘丽, 孙武, 等. 桔楼 SRAP 引物筛选及遗传多样性分析 [J]. 中草药, 2023, 54(20): 6813-6818.
- [7] Matsumoto S, Kiriwa Y, Takeda Y. Differentiation of Japanese green tea cultivars as revealed by RFLP analysis of phenylalanine ammonia-lyase DNA [J]. *Theor Appl Genet*, 2002, 104(6/7): 998-1002.
- [8] 董美超, 杨帆, 李进学, 等. 90 份鳄梨种质资源 AFLP 遗传多样性分析 [J]. 福建农业学报, 2020, 35(1): 13-19.
- [9] 杨成龙, 周明强, 班秀文, 等. 薏苡种质资源 AFLP 遗传多样性分析 [J]. 分子植物育种, 2020, 18(15): 5134-5142.
- [10] 杨培迪, 刘振, 赵洋, 等. 基于农艺性状和 SSR 标记亲缘关系分析的江华苦茶品系评价 [J]. 分子植物育种, 2021, 19(7): 2402-2409.
- [11] Sunseri F, Lupini A, Mauceri A, *et al.* Single nucleotide polymorphism profiles reveal an admixture genetic structure of grapevine germplasm from Calabria, Italy, uncovering its key role for the diversification of cultivars in the Mediterranean Basin [J]. *Aust J Grape Wine Res*, 2018, 24(3): 345-359.
- [12] 李贝贝, 姜建福, 张颖, 等. 葡萄品种 DNA 指纹数据库的构建及遗传多样性分析 [J]. 植物遗传资源学报, 2018, 19(2): 338-350.
- [13] 余孟娟, 赵博宇, 董诚明, 等. 不同产区艾叶 ISSR 遗传多样性分析 [J]. 中草药, 2023, 54(13): 4306-4311.
- [14] 陈苏丹. 菰蓝遗传多样性及其药材品质评价研究 [D]. 南京: 南京农业大学, 2014.
- [15] 张永辉, 刘崇怀, 樊秀彩, 等. ISSR 标记在中国野生葡萄分类中的应用 [J]. 果树学报, 2011, 28(3): 406-412.
- [16] 牛素贞, 刘玉倩, 杨锦标, 等. 贵州茶树地方品种的 EST-SSR 遗传多样性分析 [J]. 浙江农业学报, 2012, 24(5): 836-841.
- [17] 耿广东, 陈立杰, 张素勤. 基于 SLAF-seq 技术的古茶树 SNP 位点开发 [J]. 经济林研究, 2019, 37(2): 7-12.
- [18] 颜廷进, 蒲艳艳, 张文兰, 等. 基于 SNP 标记的菜豆品种真实性和纯度鉴定技术 [J]. 山东农业科学, 2019, 51(12): 111-119.
- [19] 俞超, 陈煜, 汪财生, 等. 基于 SLAF-seq 技术的红心火龙果 SNP 位点开发及遗传分析 [J]. 热带作物学报, 2017, 38(4): 591-596.
- [20] 张恒, 刘众杰, 樊秀彩, 等. 葡萄果粒形状简化基因组关联分析 [J]. 园艺学报, 2017, 44(10): 1959-1968.
- [21] Guo J Q, Li C J, Teng T, *et al.* Construction of the first high-density genetic linkage map of pikeperch (*Sander lucioperca*) using specific length amplified fragment (SLAF) sequencing and QTL analysis of growth-related traits [J]. *Aquaculture*, 2018, 497: 299-305.
- [22] Zhang Y X, Wang L H, Xin H G, *et al.* Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing [J]. *BMC Plant Biol*, 2013, 13: 141.
- [23] Sun X W, Liu D Y, Zhang X F, *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.
- [24] 李贝贝, 张恒, 姜建福, 等. 基于 SLAF-seq 技术的葡萄种质遗传多样性分析 [J]. 园艺学报, 2019, 46(11): 2109-2118.
- [25] Kozich J J, Westcott S L, Baxter N T, *et al.* Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform [J]. *Appl Environ Microbiol*, 2013, 79(17): 5112-5120.
- [26] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform [J]. *Bioinformatics*, 2009, 25(14): 1754-1760.

- [27] Li H, Handsaker B, Wysoker A, *et al.* The Sequence Alignment/Map format and SAMtools [J]. *Bioinformatics*, 2009, 25(16): 2078-2079.
- [28] McKenna A, Hanna M, Banks E, *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data [J]. *Genome Res*, 2010, 20(9): 1297-1303.
- [29] Tamura K, Peterson D, Peterson N, *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods [J]. *Mol Biol Evol*, 2011, 28(10): 2731-2739.
- [30] Alexander D H, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals [J]. *Genome Res*, 2009, 19(9): 1655-1664.
- [31] Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana* [J]. *Nature*, 2000, 408(6814): 796-815.
- [32] Li R Q, Yu C, Li Y R S, *et al.* SOAP2: An improved ultrafast tool for short read alignment [J]. *Bioinformatics*, 2009, 25(15): 1966-1967.

[责任编辑 时圣明]