

• 药材与资源 •

药用芍药栽培品种的 SNP 标记开发及品种鉴定

刘宇辰^{1,2}, 张振宇¹, 尹旻臻², 储姗姗¹, 彭华胜^{1,2*}, 黄璐琦^{1,2*}

1. 安徽中医药大学药学院, 安徽 合肥 230012

2. 中国中医科学院中药资源中心, 道地药材品质保障与资源持续利用全国重点实验室, 北京 100700

摘要: 目的 基于简化基因组, 建立一种准确鉴定药用芍药 *Paeonia lactiflora* 栽培品种的单核苷酸多态性 (single nucleotide polymorphism, SNP) 分子标记方法。方法 利用特异位点扩增片段测序 (specific-locus simplified fragment sequencing, SLAF-seq) 技术对浙江、安徽、四川和湖南的 13 个药用芍药栽培品种进行 SNP 标记开发。通过多个指标如杂合率、缺失率和多态性来筛选候选标记, 并通过遗传多样性、群体结构和主成分分析等方法进一步评估候选标记。最后采用能够区分药用芍药栽培品种的最简 SNP 组合绘制指纹图谱。结果 通过 SLAF-seq 测序共获得 3 703 329 个 SLAF 标签, 其中多态性标签有 60 749 个, 共获得 412 873 个群体 SNP 标记, 最后筛选的 47 个核心 SNP 标记位点表现出高多态性, 可以很好地区分药用芍药栽培品种。结论 基于简化基因组, 建立了主要药用芍药栽培品种的 SNP 标记开发和指纹图谱绘制的方法, 为药用芍药的品种鉴定和白芍药材的临床应用提供借鉴。

关键词: 药用芍药; 栽培品种; SLAF-seq; 单核苷酸多样性; 指纹图谱

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2025)01-0239-09

DOI: 10.7501/j.issn.0253-2670.2025.01.023

Development of SNP markers and identification of medicinal cultivars of *Paeonia lactiflora*LIU Yuchen^{1,2}, ZHANG Zhenyu¹, YIN Minzhen², CHU Shanshan¹, PENG Huasheng^{1,2}, HUANG Luqi^{1,2}

1. School of Pharmacy, Anhui University of Chinese Medicine, Hefei 230012, China

2. State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China

Abstract: Objective Based on reduced-representation genome, we developed a single nucleotide polymorphism (SNP) molecular marker method for the accurate identification of medicinal cultivars of *Paeonia lactiflora*. **Methods** SNP markers were developed for 13 medicinal cultivars of *P. lactiflora* from Zhejiang, Anhui, Sichuan and Hunan provinces based on specific-locus simplified fragment sequencing (SLAF-seq) technology. Candidate markers were screened by multiple indicators such as heterozygosity rate, deletion rate and polymorphism, and further evaluated by genetic diversity, population structure and principal component analysis. Finally, fingerprinting was performed using the simplest SNP combination that could distinguish the medicinal cultivars of *P. lactiflora*. **Results** A total of 3 703 329 SLAF labels were obtained by SLAF-seq sequencing, including 60 749 polymorphic labels, and 412 873 population SNP markers were obtained. Finally, a total of 47 core SNP marker sites selected showed high polymorphism, and the SNP fingerprint drawn by this method could accurately distinguish medicinal cultivars of *P. lactiflora*. **Conclusion** A method of SNP marker development and fingerprint mapping for medicinal cultivars of *P. lactiflora* based on reduced-representation genome was established, which could provide reference for variety identification of *P. lactiflora* and clinical application of *Cynanchum otophyllum*.

Key words: *Paeonia lactiflora* Pall.; cultivars; SLAF-seq; single nucleotide polymorphism; fingerprint

白芍是我国的传统中药, 具有平肝止痛、养血调经和敛阴止汗的功效。《中国药典》2020 年版^[1]规

定白芍来源于毛茛科植物芍药 *Paeonia lactiflora* Pall. 的干燥根。白芍经过长期栽培, 在安徽亳州、浙

收稿日期: 2024-08-09

基金项目: 国家重点研发计划项目 (2022YFC3500903); 中国中医科学院科技创新工程项目 (CI2024E001); 安徽省高校自然科学研究重点项目 (2023AH050854)

作者简介: 刘宇辰, 硕士研究生, 研究方向为中药资源学。E-mail: liuyuchen823@163.com

*通信作者: 黄璐琦, 中国工程院院士, 研究方向为分子生药学。E-mail: huangluqi01@126.com

彭华胜, 教授, 研究方向为本草学与中药资源学。E-mail: hspeng@126.com

江磐安、四川中江和湖南常宁等地形成了多个产区，前 3 个产区为道地产区，所产白芍分别习称“亳白芍”“杭白芍”“川白芍”^[2-3]。有学者前期^[4]对各地药用芍药栽培品种进行了调查，报道了 6 个芍药栽培品种，他们在花型、花色、心皮等形态方面有明显差异。课题组前期对全国 5 个白芍主产区开展了进一步调查，发现药用芍药栽培品种多达 12 个^[5]。李秋晗等^[6]研究表明不同产地的白芍药材如亳白芍、川白芍等在化学成分上存在差异；李霓冰等^[7]发现浙江生产的白芍药材质量最佳，其次为山东、安徽。因此，开展药用芍药栽培品种的分子标记研究，也可对白芍药材的临床应用提供依据。

分子标记不依赖表型性状，稳定性强，是探索植物遗传变异、开展品种鉴别研究的重要工具^[8-9]。与以限制性片段长度多态性（restriction fragment length polymorphisms, RFLP）为首的第 1 代分子标记和以微卫星（simple sequence repeats, SSR）为首的第 2 代分子标记相比，第 3 代分子标记单核苷酸多态性（single nucleotide polymorphism, SNP）标记更适合大量标记的高通量基因分型、数据库整合和数据共享，并且具有数量多、变异来源丰富、经济高效的优势^[10]，在遗传图谱构建、种质资源的 DNA 指纹分析和生物多样性检测等领域发挥着重要作用^[11]。特异位点扩增片段测序（specific-locus simplified fragment sequencing, SLAF-seq）技术是简化基因组测序的一种，能通过高通量测序在全基因组范围内快速鉴定高准确性的 SNP 标记，近年来，在物种的群体遗传学研究中获得广泛应用。Du 等^[12]利用 SLAF-seq 技术从中国、欧洲和北美的山楂属植物中获得 933 450 个群体 SNP 标记并用于中国山楂属物种的起源进化研究。段义忠等^[13]利用 SLAF-seq 技术开发获得 102 025 个高度一致性的沙冬青 *Ammopiptanthus mongolicus* (Maxim. ex Kom.) Cheng f. 群体 SNP，并进行 10 个地点的沙冬青的遗传分析。Li 等^[14]利用 SLAF-seq 技术开发了基因组水平的 SNP，对野生和栽培黍的遗传多样性和种群结构进行了研究。可见，SLAF-seq 逐渐成为物种群体遗传学研究中的一项可靠工具。

芍药作为著名的药用植物与观赏植物，品种鉴定的分子标记受到学者的高等关注。郭先锋等^[15]对 5 种野生芍药和 26 个栽培芍药品种的亲缘关系进行了随机扩增多态性 DNA 标记（random amplified polymorphic DNA, RAPD）分析；Yoshie 等^[16]利用

特定序列扩增（sequence-characterized amplified regions, SCAR）标记成功鉴定 2 个药用芍药栽培品种；万映伶等^[17]利用 33 对 SSR 引物，从遗传变异和遗传距离角度对我国 4 个主要栽培区的 386 个观赏芍药品种的遗传多样性进行评价。这些研究主要集中于观赏芍药的品种，对于药用芍药的品种鉴定报道较少。

本实验利用 SLAF-seq 技术测定了 13 个药用芍药栽培品种的基因序列，获得大量的 SLAF 标签，进而开发一批稳定性高、特异性强的 SNP 标记。通过多个指标如杂合率、缺失率和多态性来筛选候选标记，并通过遗传多样性、群体结构和主成分分析等方法进一步评估候选标记。最后采用能够区分药用芍药栽培品种的最简 SNP 组合绘制指纹图谱。旨在为药用芍药的品种鉴定以及白芍药材的临床应用提供借鉴。

1 材料

课题组 2022 年 5 月—2024 年 5 月通过产地调查、种质收集和移栽观察等方式，对浙江省磐安县、安徽省亳州市、四川省中江县和湖南省常宁市当地长期栽培的白芍品种进行记录，采集并制作植物标本。标本经中国中医科学院中药资源中心彭华胜教授鉴定为芍药 *P. lactiflora* Pall.。本实验共收集 13 个药用芍药栽培品种，28 份芍药样本（表 1），另外收集 3 份草芍药 *P. obovata* Maxim. 样本（编号 CSY1、CSY2、CSY3）作为系统发育分析的外类群。

2 方法

2.1 基因组 DNA 的提取

采用 QIAGEN 植物基因组 DNA 小提试剂盒 DNeasy Plant Mini Kit 提取芍药叶片样本的基因组 DNA，用 1% 琼脂糖凝胶电泳法检测提取基因组 DNA 的完整性，并用 Nanodrop 微量紫外分光光度计检测 DNA 的浓度及质量，确保提取的 DNA 质量满足建库测序要求。

2.2 酶切建库

选取芍药近缘种牡丹 *Paeonia ostii* T. Hong & J. X. Zhang 基因组（<https://ftp.cngb.org/pub/CNSA/data5/CNP0003098/CNS0560369/CNA0050666/Paeoniaostii.Fengdan.genome.fa.gz>）作为参考基因组进行酶切预测^[18]，牡丹基因组大小为 12.28 Gb。利用百迈克生物技术有限公司自主研发的酶切预测软件对参考基因组进行酶切预测。对得到的酶切片段（SLAF 标签）进行 3'端加 A 处理、连接 Dual-index^[19]

表 1 药用芍药栽培品种信息

Table 1 Information on medicinal cultivars of *P. lactiflora*

采样点	品种名	拉丁名	药材名	编号	样品/份	形态特征
浙江省磐安县	银妆素裹	<i>P. lactiflora</i> 'Yin Zhuang Suguo'	杭白芍	YZSG	2	花重瓣，白色；雌蕊瓣化；茎和叶无毛
	大红袍	<i>P. lactiflora</i> 'Dahongpao'	杭白芍	DHP	2	花单瓣，红色；心皮密被毛；茎和叶无毛
	红云连天	<i>P. lactiflora</i> 'Hongyun Liantian'	杭白芍	HYLT	2	花单瓣，粉红色；心皮密被毛；茎和叶无毛；根各分支明显
	交叉根	<i>P. lactiflora</i>	杭白芍	JCG	2	花单瓣，粉红色；心皮密被毛；茎和叶无毛；根交叉呈盘状
	长柄红花	<i>P. lactiflora</i> 'Changbing Honghua'	杭白芍	CBHH	2	花单瓣，红色；心皮无毛；茎和叶有毛，叶背面仅沿叶脉有毛
	乌龙探海	<i>P. lactiflora</i> 'Wulongtanhai'	杭白芍	WLTH	2	花单瓣，紫红色；心皮无毛；茎和叶有毛，叶背面仅沿叶脉有毛
	胭脂点月	<i>P. lactiflora</i> 'Yanzhidian Yue'	杭白芍	YZDY	2	花单瓣，粉红色；心皮密被毛；茎和叶有毛，叶背面仅沿叶脉有毛
安徽省亳州市	亳白芍	<i>P. lactiflora</i> 'Bobaishao'	亳白芍	BBS	2	花单瓣，红色；心皮疏生柔毛；茎和叶无毛
四川省中江县	重瓣白花芍药	<i>P. lactiflora</i> 'Chongbanbaihua'	川白芍	ZJBH	4	花重瓣，白色；心皮密被毛；茎和叶背面有毛
	重瓣粉花芍药	<i>P. lactiflora</i> 'Chongbanfenhua'	川白芍	ZJFH	2	花重瓣，粉红色；心皮无毛；茎和叶无毛
湖南省常宁市	湖南粉花芍药	<i>P. lactiflora</i> 'Hunanfenhua'	\	CNFH	2	花单瓣，粉红色；心皮密被毛；茎和叶有毛，叶背面疏生短柔毛
	湖南红花芍药	<i>P. lactiflora</i>	\	CNHH	2	花单瓣，红色；心皮无毛；茎和叶有毛，叶背面仅沿叶脉有毛
	重瓣白花芍药	<i>P. lactiflora</i>	\	CNBH	2	花重瓣，白色；心皮密被毛；茎和叶背面有毛

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 Illumina 平台进行测序。

2.3 SLAF 标签和 SNP 标记的开发

经过 SLAF-seq 建库产生特定酶切片段，一个酶切片段就是一个 SLAF 标签。本研究测序产生的 reads 来源于同一限制性内切酶对不同样本作用产生的长度相同的酶切片段，根据序列相似性将各样本的 reads 进行聚类，聚类到一起的 reads 来源于一个 SLAF 标签。同一 SLAF 标签在不同样本间的序列相似度远高于不同 SLAF 标签间的相似度；一个 SLAF 标签在不同样本间序列有差异（即有多态性），即可定义为多态性 SLAF 标签。

SNP 标记的开发是以每个 SLAF 标签中深度最高的序列类型作为参考序列，利用 bwa (0.7.10-r789) [20] 将测序 reads 比对到参考序列上，并使用 GATK (v3.8) [21] 和 samtools (v1.9) [22] 2 种方法开发 SNP，以 2 种方法得到的 SNP 标记交集作为最终可靠的 SNP 标记数据集。

2.4 系统发育树构建

将开发的 SNP 标记根据完整度 > 0.5，次要等位基因比例 (minor allele frequency, MAF) > 0.05 的标准进行初步过滤。基于初筛的群体一致性 SNP，使用 MEGA X [23] 软件，基于邻接法 (neighbor-joining)，采用 Kimura 2-parameter 模型，bootstrap 重复 1 000 次，构建所有样本的系统发育树。

2.5 高质量核心 SNP 标记筛选和遗传多样性分析

基于非遗传群体材料构成的群体开发的 SNP，应当具有标记质量高、代表性强、标记的材料区分度高、在基因组上分布均匀、特异性强等特点。基于这些原则 [24] 进行筛选：(1) 标记在基因组上分布均匀；(2) 标记不存在缺失位点，即位点完整度 100%；(3) 丢弃次要等位基因比例低于 20% 的位点；(4) 丢弃多态信息含量 (polymorphism information content, PIC) 小于 0.35 的位点；(5) 哈迪-温伯格检验，保留 *P* 值大于 0.01 的位点；(6) 筛选的标记前后 100 bp 处，不存在其他位点突变。

采用自写的 perl 脚本对各位点进行遗传多样性分析，计算参数包括有效等位基因数 (effective number of allele, N_e)、观察杂合度 (H_o)、期望杂合度 (H_e)、Nei 基因多样性 (H)，Shannon 多样性指数 (I) 与位点 PIC。群体内样本各位点指数求均值为群体值。

2.6 群体遗传结构和主成分分析

采用 Admixture [25] 软件进行群体结构分析。针对研究群体，预先设定亚群数目 (*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 [26]。使用 R 包 Pophelper 生成每个 *K* 值的 Q 矩阵堆叠图 (<http://royfrancis.github.io/pophelper>)。使用 Eigensoft [27] 软件包中的 smartPCA 程序基于 SNP 数据进行主成分分析，得到样本的聚类情况。

2.7 SNP 指纹图谱构建

基于筛选的高质量核心 SNP 标记获得所有样本的分型结果, 挑选出重复样本基因型一致的 SNP 位点并转化为二元编码数据, 通过 Excel 软件绘制 SNP 指纹图谱。

3 结果与分析

3.1 建库测序结果

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 *HinCII*+*EcoRV*, 酶切片段长度在 364~414 bp 的序列定义为 SLAF 标签。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^[28], 本实验的酶切效率为 87.59%, 表明酶切反应和 SLAF 建库正常。

31 份样本测序后共获得 931.00 Mb reads 数据, 药用芍药栽培品种及外类群测序结果如表 2 所示。各样本得到的 reads 数量各不相同, 最多有 109 109 244 条 reads (CNBH2), 最少有 5 146 388 条 reads (ZJBH4)。测序的 GC 碱基比例的平均值为 36.63%, 最小的 GC 碱基比例为 35.47% (CNFH2), 最大的 GC 碱基比例为 38.85% (DHP2), 此次测序所有样本 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。此次实验测序质量值 *Q*₃₀ 均大于 85%, 最小的 *Q*₃₀ 为 85.07% (CNFH2), 最大的 *Q*₃₀ 为 97.23% (JCG2、ZJBH1), 平均 *Q*₃₀ 为 95.02%。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。

3.2 SLAF 标签开发及 SNP 信息统计

每个样本标签平均测序深度为 26.04×, 共获得 3 703 329 个 SLAF 标签, 具体的芍药 SLAF 标签数统计结果如表 3 所示, 其中共获得 60 749 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 412 873 个群体 SNP 标记。SNP 信息统计结果见表 3。

3.3 系统发育树分析

基于筛选的 51 436 个高一致性群体 SNP 标记, 通过 MEGAX 软件, 基于邻接法 (neighbor-joining, NJ), 采用 Kimura 2-parameter 模型, bootstrap 重复

表 2 药用芍药栽培品种及外类群测序结果

Table 2 Sequencing results of medicinal cultivars of *P. lactiflora* and outgroups

样本	reads 数量	GC 碱基比例/%	<i>Q</i> ₃₀ 比例/%
BBS1	28 912 296	35.95	97.04
BBS2	29 191 964	36.18	97.12
CBHH1	18 815 464	37.18	90.65
CBHH2	28 103 758	38.42	96.01
CNBH1	12 219 786	36.26	96.03
CNBH2	109 109 244	35.97	96.53
CNFH1	27 163 394	35.64	97.10
CNFH2	11 974 736	35.47	85.07
CNHH1	23 445 454	36.72	89.01
CNHH2	77 697 638	36.30	96.75
DHP1	19 261 284	36.01	89.92
DHP2	13 030 046	38.85	95.65
HYLT1	34 995 856	35.97	96.31
HYLT2	24 064 802	35.85	96.30
JCG1	39 266 546	35.91	96.78
JCG2	25 246 654	35.96	97.23
WLTH1	28 385 556	35.78	96.73
WLTH2	23 001 638	36.95	97.14
YZDY1	34 321 960	38.75	96.29
YZDY2	35 888 906	38.30	96.10
YZSG1	35 258 160	35.73	92.66
YZSG2	17 166 050	35.99	88.13
ZJBH1	29 742 924	35.89	97.23
ZJBH2	37 534 952	35.81	96.88
ZJBH3	23 752 718	37.34	96.12
ZJBH4	5 146 388	37.59	94.18
ZJFH1	40 809 254	35.92	95.90
ZJFH2	28 353 362	36.59	95.89
CSY1	24 087 870	37.26	96.48
CSY2	20 607 404	37.42	96.69
CSY3	24 467 876	37.69	95.82
平均值		36.63	95.02

1 000 次, 构建各样本的系统发育树, 如图 1 所示。28 份芍药样本可分为 3 个亚群。其中, 亚群 I 中浙江的“交叉根”(JCG)和“红云连天”(HYLT)亲缘关系很近, 2 个品种除地下部分根形态有区别, 地上部分形态特征一致, 推测这 2 者可能为同一品种, 此外“红云连天”的 2 个样本并未聚在一支, 可能是 HYLT2 取样时鉴定不准确; 浙江栽培的“长柄红花”(CBHH)与湖南栽培的“湖南红花芍药”(CNHH)亲缘关系很近难以区分, 并且 2 个品种在

表 3 药用芍药栽培品种的 SLAF 标签和 SNP 统计结果

Table 3 SLAF label and SNP statistics of medicinal cultivars of *P. lactiflora*

样本编号	SLAF 标签数	测序总深度	平均深度	SNP 总数	SNP 数量	杂合率/%
BBS1	531 013	12 984 522	24.45	412 873	147 604	12.40
BBS2	548 872	13 087 208	23.84	412 873	158 601	12.93
CBHH1	516 064	6 994 371	13.55	412 873	157 519	12.97
CBHH2	711 655	11 897 526	16.72	412 873	104 104	10.59
CNBH1	239 438	5 439 227	22.72	412 873	66 178	9.21
CNBH2	740 024	49 355 772	66.69	412 873	212 595	14.73
CNFH1	482 965	12 211 378	25.28	412 873	131 515	12.45
CNFH2	217 027	4 069 820	18.75	412 873	69 659	10.16
CNHH1	406 019	8 611 665	21.21	412 873	151 227	12.83
CNHH2	745 970	34 706 631	46.53	412 873	204 318	14.54
DHP1	387 393	7 311 697	18.87	412 873	131 490	14.19
DHP2	375 937	5 467 732	14.54	412 873	118 141	11.24
HYLT1	379 915	15 626 336	41.13	412 873	119 577	10.30
HYLT2	349 595	10 720 857	30.67	412 873	108 047	10.44
JCG1	587 699	17 381 571	29.58	412 873	162 854	12.76
JCG2	509 373	11 199 532	21.99	412 873	146 103	11.99
WLTH1	519 706	12 779 788	24.59	412 873	136 926	12.50
WLTH2	574 961	10 162 553	17.68	412 873	145 709	12.23
YZDY1	712 724	14 961 440	20.99	412 873	106 724	11.25
YZDY2	687 858	15 651 351	22.75	412 873	111 513	10.88
YZSG1	345 693	14 307 905	41.39	412 873	140 198	11.27
YZSG2	320 069	6 288 602	19.65	412 873	105 712	9.90
ZJBH1	563 546	13 314 746	23.63	412 873	156 774	12.33
ZJBH2	521 658	16 971 311	32.53	412 873	143 017	11.51
ZJBH3	392 553	10 343 746	26.35	412 873	129 653	10.61
ZJBH4	158 430	2 058 095	12.99	412 873	64 940	11.22
ZJFH1	433 546	18 217 414	42.02	412 873	133 686	10.61
ZJFH2	387 572	12 533 648	32.34	412 873	120 952	10.23

植物形态上高度一致，由此推测两者为“异名同物”。另外，亚群 I 中“银妆素裹”（YZSG）这个品种的 2 个样本距离较远，其中 YZSG1 与四川的“重瓣粉花芍药”聚为一支，且 2 个品种均为重瓣花，而 YZSG2 与“单瓣”类型的芍药聚为一支，可能是由于该样本收集时鉴定不准确导致。整体来看，聚类结果与花型、花色等形态特征有一定相关性，与地理来源没有明确一致性。

3.4 核心 SNP 标记的遗传多样性分析

通过对杂合率、缺失率和多态性等指标进行严格筛选，最终获得了 48 个高质量核心 SNP 位点，用于鉴定药用芍药栽培品种的基因分型。对筛选的 48 个 SNP 位点进行遗传多样性分析，48 个 SNP 位点的 N_e 为 96， H_e 分布在 0.258~0.710，

H_e 分布在 0.458~0.500， H 分布在 0.465~0.508， I 分布在 0.650~0.693，PIC 分布在 0.353~0.375，MAF 均值为 0.413。以上指标说明筛选的 48 个核心 SNP 质量较高，适于对 28 份芍药样本进行基因分型和构建指纹图谱。

3.5 群体结构与主成分分析

根据筛选的 48 个高质量核心 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 所示。交叉验证聚类结果显示， $K=2$ 时交叉验证错误率最低，即最优分群数为 2（图 2-A、B）。结果表明 13 个品种之间遗传差异比较小，并没有形成明显的种群分化。

通过 PCA 聚类图（图 3）可知，28 份芍药样本

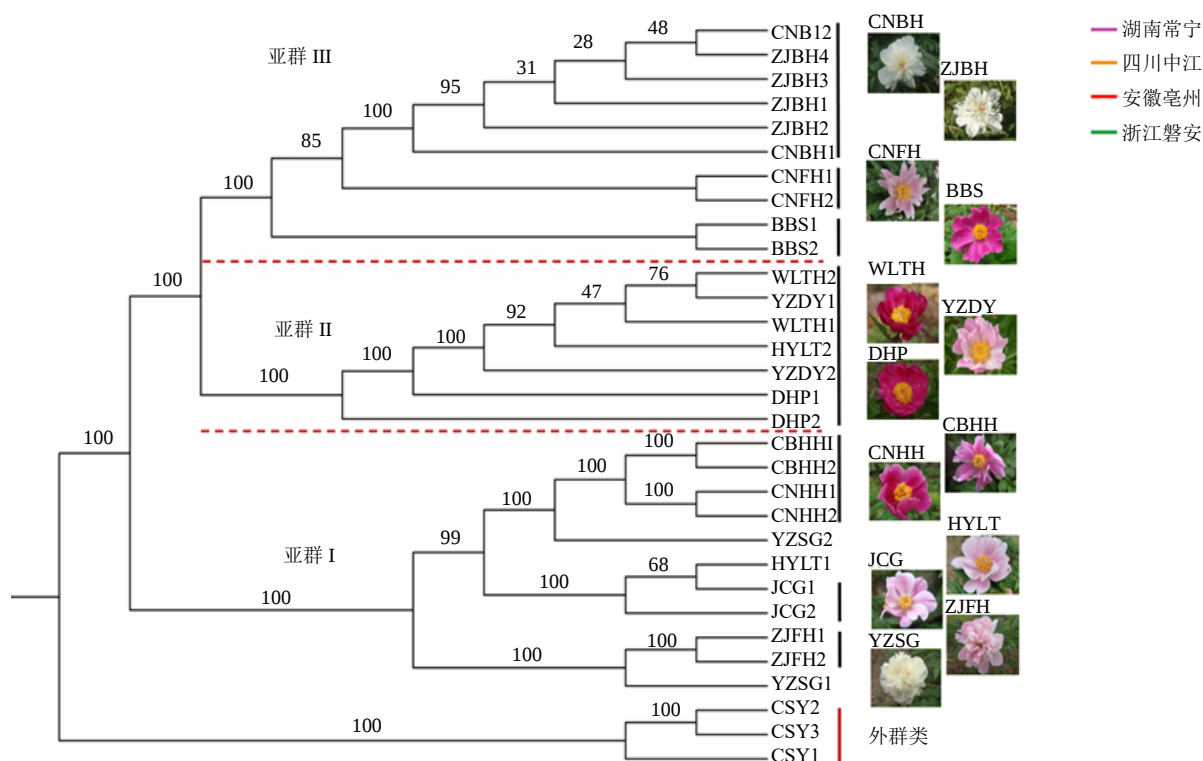
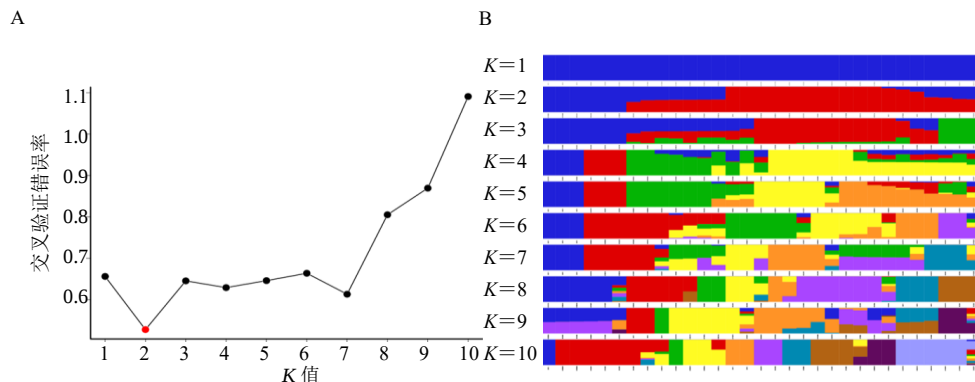


图 1 基于 SNP 标记的药用芍药栽培品种系统发育树

Fig. 1 Phylogenetic tree of medicinal cultivars of *P. lactiflora* based on SNP markers图 2 不同 K 值所对应的交叉验证错误率及样本聚类图Fig. 2 Cross-validation error rate and sample clustering diagram corresponding to different K values

明显聚成 3 组。第 1 组中湖南和四川的“重瓣白花芍药”(CNBH, ZJBH)聚在一起;第 2 组中浙江的“胭脂点月”(YZDY)“乌龙探海”(WLTH)“红云连天”(HYLT)和“大红袍”(DHP)和安徽的“亳白芍”(BBS)以及湖南的“湖南粉花芍药”(ZJFH)聚在一起;第 3 组中浙江磐安的“交叉根”(JCG)“红云连天”(HYLT)“银妆素裹”(YZSG)“长柄红花”(CBHH)和四川的“重瓣粉花芍药”(ZJFH)以及湖南的“湖南红花芍药”

(CNHH)聚在一起。PCA 聚类结果与系统发育树和群体结构分析结果相验证,说明 48 个核心 SNP 标记具有可靠性。

3.6 构建药用芍药栽培品种的 SNP 指纹图谱

结合系统发育分析、群体结构与主成分分析,确定 13 个药用芍药栽培品种的 48 个 SNP 位点基因型以构建指纹图谱,碱基个体存在腺嘌呤(A)、胞嘧啶(C)、鸟嘌呤(G)和胸腺嘧啶(T)4 种类型,根据 SNP 的原理,每个 SNP 位点上可能都有

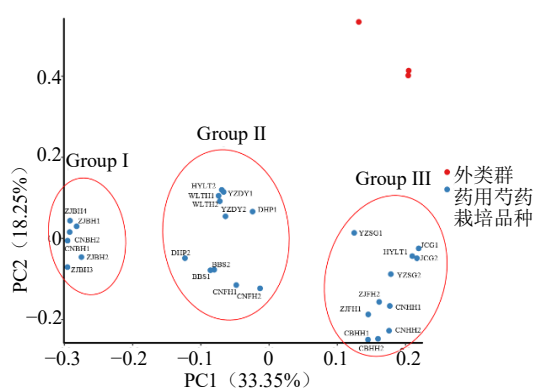


图 3 药用芍药栽培品种 PCA 聚类图

Fig. 3 PCA cluster map of medicinal cultivars of *P. lactiflora*

置换、颠换、缺失和插入 4 种变异类型。A/A、G/G、C/C、T/T 这 4 种类型属于纯合位点，而另外的 A/C、A/T、C/T、C/G 等均属于杂合位点，用不同颜色分别表示 4 种纯合位点，用白色表示所有的杂合位点。

13 个品种中，浙江的“大红袍”“红云连天”和“银妆素裹”这 3 个品种的不同样品基因型差别较大，故排除；另外，有 1 个 SNP 位点无效，也排除。最后，用 47 个 SNP 位点构建了 10 个药用芍药栽培品种的指纹图谱，构建的 SNP 指纹图谱（图 4）可以直观地表达出药用芍药栽培品种在基因组水平上的差异。

4 讨论

芍药在我国已有悠久的栽培历史，可分为观赏

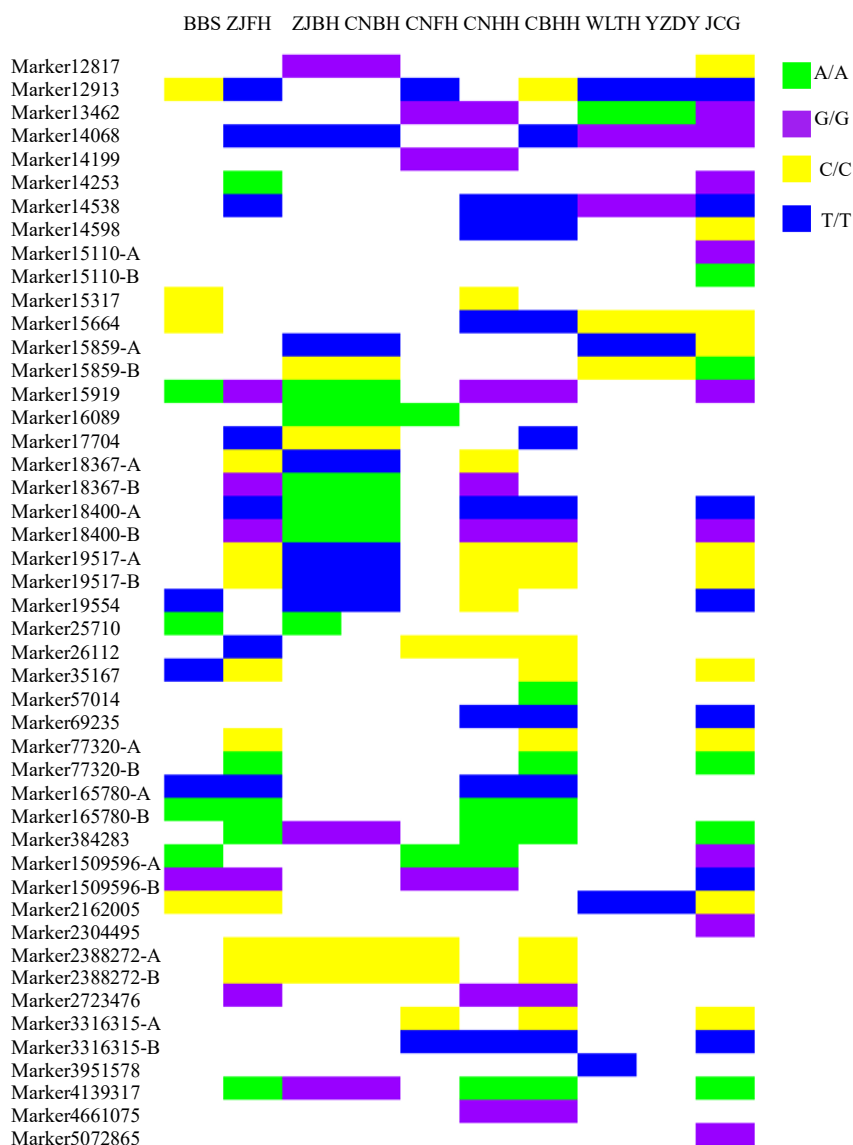


图 4 药用芍药栽培品种的 SNP 指纹图谱

Fig. 4 SNP fingerprint of medicinal cultivars of *P. lactiflora*

芍药和药用芍药,前人关于品种鉴定的研究主要集中于观赏芍药。进行药用植物的核心种质的构建要从源头上把关即首先要鉴定出优良的种质^[29],白芍不同品种之间亲缘关系远近及遗传多样性仍不明确,并且可能存在品种间的混杂现象,目前亟需挖掘高质量核心 SNP 标记,建立鉴定准确、检测高效的药用芍药栽培品种的分子鉴别技术。本实验通过 SLAF-seq 技术对安徽、浙江、四川和湖南的 13 个药用芍药栽培品种进行 SNP 标记开发,最后筛选的 47 个核心 SNP 标记位点表现出高多态性,能很好地区分药用芍药栽培品种。自近代,“杭白芍”“川白芍”“亳白芍”3 大商品格局中以“亳白芍”产量最大,以“杭白芍”质量最佳。本研究表明“亳白芍”这一品种与其他品种的遗传距离较大,结合植物形态和分子证据能很好地从其他产区的品种中区分开。而“杭白芍”在长期的引种和栽培过程中逐渐分化成 2 个分支,其中一个分支与四川和湖南的栽培品种聚集,也有可能是其他产区的栽培品种交流到浙江后,根据当地的土壤、气候等原因,发生了遗传分化^[30]。

简化基因组测序是在高通量测序基础上发展起来的,它可以通过寻找合适的限制性内切酶来降低基因组的复杂程度,用来研究物种的遗传变异。这种测序技术尤其适用于没有参考基因组的物种,能有效克服基因组复杂的问题^[13]。系统发育分析结果表明不同品种虽然地理来源不同,但亲缘关系较近,并且亲缘关系与花型、花色等形态特征相关,与地理来源没有明确一致性。其中,浙江的“红云连天”和“交叉根”这 2 个品种的植物形态相近,系统发育分析显示两者亲缘关系很近,SNP 指纹图谱也显示两个品种的基因型差异很小,因此浙江的“红云连天”和“交叉根”可能为同一品种。湖南的“湖南红花芍药”和浙江的“长柄红花”,以及湖南和四川的“重瓣白花芍药”之间,植物形态相近,系统发育分析显示亲缘关系很近,SNP 指纹图谱也显示基因型差异很小,因此“湖南红花芍药”和“长柄红花”可能为同一品种,湖南和四川的“重瓣白花芍药”也可能为同一品种,只是分布在不同区域。另外,SNP 指纹图谱相同的不同品种具有不同的形态学特征,比如“乌龙探海”和“胭脂点月”的基因型相似,但是花色明显不同,这是因为所开发的分子标记所在位点与表型性状之间不具备足够的关联性^[31]。其次,“大红袍”“红云连天”和“银妆

素裹”这 3 个品种的不同样本之间基因型差别较大,并且系统发育分析显示不同样本之间距离较远,说明这 3 个品种取样可能存在误差,需要进一步增加样本进行分析。另外值得注意的是目前芍药没有基因组,因此 SNP 标记是否在基因组中分布均匀尚不可知,具有一定局限性。

综上所述,本研究结果表明简化基因组技术能从分子水平上检测不同产区的药用芍药栽培品种之间的基因差异,可以为药用芍药的品种鉴定和白芍药材的临床应用提供科学依据。芍药是一个野生和栽培,药用和观赏各种特点并存的物种,因此,对于品种的鉴定应从植物形态和分子方面综合考虑。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 中国药典 [S]. 一部. 2020: 108.
- [2] 王军, 彭华胜, 彭代银, 等. 亳州市及药材种植业发展沿革考 [J]. 中药材, 2017, 40(5): 1228-1233.
- [3] 梁伟龙, 林钦贤, 王斌, 等. 道地与非道地产区白芍药材质量的比较 [J]. 中国现代应用药学, 2020, 37(24): 3000-3004.
- [4] 查良平, 王德群, 彭华胜, 等. 中国药用芍药栽培品种 [J]. 安徽中医学院学报, 2011, 30(5): 70-73.
- [5] 张振宇, 储姗姗, 彭华胜. 中国药用芍药栽培历史与品种调查 [J]. 中国中药杂志, 2024, 49(22): 6241-6248.
- [6] 李秋晗, 张喜武, 闫广利, 等. 基于 UPLC-Q-TOF-MSE 结合多元统计方法的不同产地白芍药材质量评价 [J]. 中草药, 2023, 54(7): 2243-2253.
- [7] 李霓冰, 邓青, 谭苑芬. 不同产地白芍的成分的含量测定及质量评价 [J]. 实用中医内科杂志, 2023, 37(8): 133-135.
- [8] 蒋路园, 吴文丽, 张恺恺, 等. 南方红豆杉转录组 SSR 位点分析及其分子标记开发 [J]. 中草药, 2024, 55(3): 928-936.
- [9] 刘东, 孟瑾瑾, 王盼, 等. 不同生态类型菘蓝 SLAF-seq 分子标记 [J]. 中草药, 2024, 55(7): 2375-2382.
- [10] Nadeem M A, Nawaz M A, Shahid M Q, *et al.* DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing [J]. *Biotechnol Biotechnol Equip*, 2018, 32(2): 261-285.
- [11] Yang S, Gill R A, Zaman Q U, *et al.* Insights on SNP types, detection methods and their utilization in *Brassica* species: Recent progress and future perspectives [J]. *J Biotechnol*, 2020, 324: 11-20.
- [12] Du X, Zhang X, Bu H D, *et al.* Molecular analysis of evolution and origins of cultivated hawthorn (*Crataegus*

- spp.) and related species in China [J]. *Front Plant Sci*, 2019, 10: 443.
- [13] 段义忠, 王建武, 杜忠毓, 等. 基于 SLAF-seq 简化基因组技术的沙冬青 SNP 位点开发及遗传分析 [J]. 植物研究, 2018, 38(1): 141-147.
- [14] Li C X, Liu M X, Sun F J, *et al.* Genetic divergence and population structure in weedy and cultivated broomcorn millets (*Panicum miliaceum* L.) revealed by specific-locus amplified fragment sequencing (SLAF-seq) [J]. *Front Plant Sci*, 2021, 12: 688444.
- [15] 郭先锋, 王莲英. 部分芍药种质资源的 RAPD 分析 [J]. 园艺学报, 2007, 34(5): 1321-1326.
- [16] Yoshie Y, Ando H, Tamura T, *et al.* Development of SCAR markers to identify medicinal cultivars of *Paeonia lactiflora* [J]. *Biol Pharm Bull*, 2022, 45(3): 292-300.
- [17] 万映伶, 金亦佳, 刘爱青, 等. 不同栽培地芍药品种资源遗传多样性的 SSR 分析比较 [J]. 园艺学报, 2023, 50(4): 875-884.
- [18] Yuan J H, Jiang S J, Jian J B, *et al.* Genomic basis of the giga-chromosomes and giga-genome of tree peony *Paeonia ostii* [J]. *Nat Commun*, 2022, 13(1): 7328.
- [19] 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.
- [20] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform [J]. *Bioinformatics*, 2009, 25(14): 1754-1760.
- [21] 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.
- [22] Li H, Handsaker B, Wysoker A, *et al.* The Sequence Alignment/Map format and SAMtools [J]. *Bioinformatics*, 2009, 25(16): 2078-2079.
- [23] Kumar S, Stecher G, Li M, *et al.* MEGA X: Molecular evolutionary genetics analysis across computing platforms [J]. *Mol Biol Evol*, 2018, 35(6): 1547-1549.
- [24] Wang Y Y, Lv H K, Xiang X H, *et al.* Construction of a SNP fingerprinting database and population genetic analysis of cigar tobacco germplasm resources in China [J]. *Front Plant Sci*, 2021, 12: 618133.
- [25] Alexander D H, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals [J]. *Genome Res*, 2009, 19(9): 1655-1664.
- [26] Pritchard J K, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data [J]. *Genetics*, 2000, 155(2): 945-959.
- [27] Price A L, Patterson N J, Plenge R M, *et al.* Principal components analysis corrects for stratification in genome-wide association studies [J]. *Nat Genet*, 2006, 38(8): 904-909.
- [28] Li R Q, Yu C, Li Y R, *et al.* SOAP2: An improved ultrafast tool for short read alignment [J]. *Bioinformatics*, 2009, 25(15): 1966-1967.
- [29] 黄璐琦, 吕冬梅, 杨滨, 等. 药用植物种质资源研究的发展: 核心种质的构建 [J]. 中国中药杂志, 2005, 30(20): 1565-1568.
- [30] 徐攀, 梁卫青, 周洁, 等. 白芍种质资源的 DNA 分子标记研究 [J]. 中华中医药学刊, 2017, 35(4): 949-953.
- [31] Li Y, Yang K, Yang W, *et al.* Identification of QTL and qualitative trait loci for agronomic traits using SNP markers in the adzuki bean [J]. *Front Plant Sci*, 2017, 8: 840.

[责任编辑 时圣明]