

DOI:10.13275/j.cnki.lykxyj.2022.004.008

栗杂交 F_1 代群体遗传结构及其农艺性状关联分析

江锡兵¹, 章平生¹, 张东北², 吴仁超², 吴剑²,
吴聪连², 赖俊声², 龚榜初^{1*}

(1. 中国林业科学研究院亚热带林业研究所, 浙江 杭州 311400; 2. 浙江省庆元县自然资源和规划局, 浙江 庆元 323800)

摘要: [目的] 解析控制栗属农艺性状的自然等位变异并获得与其相关联的 SSR 位点, 为分子标记辅助选择和高效育种奠定基础。[方法] 以涵盖 9 个杂交组合的 235 份栗杂交子代混合群体为材料, 采用 SSR 标记对其进行群体遗传结构和连锁不平衡 (LD) 分析, 并将 32 个高多态性 SSR 标记与 25 个农艺性状进行关联分析。[结果] 统计概率 $P < 0.05$ 时, 32 对 SSR 标记组成 496 个位点中存在一定的连锁不平衡, 其中, 74 个位点的连锁不平衡水平较高, 占总标记对的 14.92%; 而 $P < 0.01$ 时, 32 个 SSR 位点间 LD 总体水平较低。群体遗传结构分析结果显示, 当 $K=4$ 时, ΔK 值接近最大。据此, F_1 代混合群体被划分为 4 个亚群, 且每个亚群中子代分布呈现出一定的遗传分化, 平均混合度为 0.120。GLM 模型中, 高达 28 个 SSR 标记位点与树高等 24 个农艺性状呈极显著关联, 每个性状可检测到关联的 SSR 位点数 1~22 个不等, 解释率范围为 4.65%~24.02%; MLM 模型中, 得出 26 个 SSR 位点与树高等 23 个性状呈极显著关联, 解释率范围为 5.04%~24.02%。综合两种模型关联分析结果, 发现 15 个 SSR 标记与树高等 3 个生长性状关联, 14 个 SSR 标记与 1 年生枝条长度等 3 个枝条性状关联, 26 个 SSR 标记与叶片长度等 18 个叶片表型及其光合生理性状关联。[结论] ICMA017s 等 15 个 SSR 标记分别与树高等 13 个农艺性状高度关联, 标记对性状的解释率均在 10.00% 以上, 且存在同一标记与多个性状、同一性状与多个标记高度关联的现象。

关键词: 栗属植物; 杂交子代; SSR 标记; 群体遗传结构; 关联分析

中图分类号: S718.46

文献标志码: A

文章编号: 1001-1498(2022)04-0072-12

壳斗科 (Fagaceae) 栗属 (*Castanea* Mill.) 植物共有 7 个种, 自然分布于欧亚及北美大陆的温带广阔地域。其中, 板栗 (*C. mollissima* Blume)、锥栗 (*C. henryi* (Skan) Rehd. et Wils.) 和茅栗 (*C. seguinii* Dode) 为我国特有种。栗属中国特有种在栗属植物研究和利用中占有重要地位, 其品质优良, 适应性和抗逆性强, 尤其对栗疫病和墨水病具有较强的抵抗力, 是进行食用栗品种改良的重要基因来源, 对世界栗属资源保护和可持续利用具

有重要作用^[1]。我国栗属育种研究起始于上世纪 50—60 年代, 通过常规育种方法选育了一大批应用于生产的优良品种等^[2-3], 目前传统的选择育种和杂交育种仍是栗属品种选育和性状遗传改良的主要方法。

栗属为多年生木本植物, 其童期长, 选育一个新品种或对某一性状进行遗传改良周期漫长, 至少需要 10~15 a, 且花费巨大。同时, 栗属农艺性状大多是受微效多基因控制的数量性状, 遗传基础

收稿日期: 2021-09-13 修回日期: 2021-10-25

基金项目: 中央级公益性科研院所基本科研业务费专项资金重点项目 (CAFYBB2017ZA004-6); 浙江省农业新品种选育重大科技专项 (2021C02070-4)

* 通讯作者: 龚榜初, 研究员, 主要研究方向: 经济林育种和栽培。E-mail: gongbc@126.com

复杂,且易受环境影响,导致传统的育种方法效率不高。确定控制农艺性状的自然等位变异,不仅有助于解析其遗传基础,且可以为分子标记辅助育种乃至定向育种提供有效的基因资源和分子标记,具有重要的理论意义和应用价值^[4]。近年来,关联分析(Association analysis)成为解析复杂数量性状遗传基础的一个重要手段。关联分析又称连锁不平衡作图(Linkage disequilibrium mapping)或关联作图(Association mapping),是一种以连锁不平衡为基础,检测群体内处于连锁不平衡状态的标记或候选基因的遗传变异与目标性状显著关联频率的方法,目前已广泛应用于农林作物。截至目前,关联分析在林木中应用主要以自然群体为主,而以杂交群体为材料进行关联分析鲜有报道。

简单重复序列(SSR)广泛分布于各类真核生物基因组中,为共显性标记,具有多态性高、重复性好、便于检测等特点^[5]。SSR标记现已成为生物遗传特性研究的一种重要分子标记,被广泛应用于栗属植物遗传多样性分析、品种鉴定、遗传图谱构建等研究^[5-11]。本课题组以锥栗种内、板栗和锥栗种间共9个杂交组合 F₁ 代群体(235个单株)为材料,前期采用32对高多态性栗属 SSR 标记对其遗传多样性、遗传效应等进行了分析^[12]。在此基础上,本研究进一步对其进行群体遗传结构和连锁不平衡分析,并将 SSR 标记与 F₁ 代生长、枝条、叶片表型及其光合生理等农艺性状表型数据进行关联分析,以期获得与农艺性状相关的分子标记,为栗属植物分子标记辅助选择和高效育种奠定基础。

1 材料与方法

1.1 试验材料

供试材料为锥栗种内、锥栗和板栗种间共9个杂交组合235份杂交 F₁ 代单株及其亲本(表1)。其中,4个亲本 YLZ 1、YLZ 2、YLZ 24 和 YLZ 26 为浙江省审(认)定锥栗良种, YLZ 24 早熟、大果高产, YLZ 26 特别高产稳产, YLZ 1 和 YLZ 2 耐贮藏、优质且适宜加工; YLZ 14 和 YLZ 15 为锥栗优株无性系,均是晚熟、优质品系;‘魁栗’为浙江省板栗良种,果型特大,平均单果质量 20~25 g。2011 年通过人工控制授粉获得杂交种实,

2012 年播种育苗,2013 年春季建立杂交子代测定林,同时栽植所有亲本 1 年生嫁接苗若干株,株行距 4 m × 4 m,每年进行精细抚育管理,235 份子代单株及其亲本生长发育状况良好。

表 1 栗杂交组合概况
Table 1 Cross combination of chestnut

组合编号 Combination No.	杂交组合 Combination	子代数量 Number of hybrids
C1	YLZ 26 × YLZ 14	14
C2	YLZ 26 × YLZ 15	30
C3	YLZ 24 × YLZ 1	30
C4	YLZ 24 × YLZ 15	30
C5	YLZ 1 × YLZ 24	30
C6	YLZ 1 × YLZ 2	17
C7	YLZ 14 × YLZ 1	24
C8	魁栗 Kuili × YLZ 15	30
C9	魁栗 Kuili × YLZ 1	30

1.2 样品采集与 DNA 提取

于 2019 年 4 月份采集 235 个子代单株及其亲本幼嫩、无病虫害的新鲜叶片,每个单株采集嫩叶 3~5 片,装于事先做好标记的自封袋中,并将自封袋置于装有干冰的泡沫箱中带回实验室。

采用北京艾德莱生物科技公司生产的 DNA 试剂盒提取 235 份子代及其亲本基因组 DNA,所提取的 DNA 经 1% 的琼脂糖凝胶电泳检测质量和紫外分光光度计测定浓度后, -20 °C 低温下保存备用。

1.3 SSR 标记引物筛选与毛细管电泳分析

从 120 对引物中筛选出条带清晰、多态性高、重复性好的 32 对 SSR 引物,采用毛细管电泳仪对所有样本的 SSR 扩增产物进行分析(图 1)。32 对 SSR 引物具体信息、SSR 扩增程序、毛细管电泳数据获得等内容详见本课题组前期研究^[12]。

1.4 表型性状测定

分别于 2018—2019 年(连续 2 a) 1、7、11 月份对 235 份子代单株的枝条性状、叶片表型及其光合生理指标、生长性状进行调查测定,具体方法参照本课题组章平生等^[13-14]前期研究。

1.5 数据分析

利用 TASSEL 5.0 软件和 R 语言完成标记间的

连锁不平衡 (LD) 分析及关联分析, STRUCTURE 2.3.4 软件进行群体遗传结构分析。



注: 图中横坐标为扩增片段大小, 纵坐标为扩增片段强度。

Note: the abscissa is the size of the amplified fragment and the ordinate is the intensity of the amplified fragment.

图 1 SSR 标记毛细管电泳

Fig. 1 Capillary electrophoresis of SSR markers

根据 R^2 确定标记间的关联程度, 即假设有两个连锁的座位 A 和 B, 其等位基因分别为 A、a 和 B、b, 用 π_A 、 π_a 、 π_B 、 π_b 分别表示相应的等位基因频率, 用 π_{AB} 、 π_{aB} 、 π_{Ab} 和 π_{ab} 分别表示 4 种单倍型 AB、aB、Ab 和 ab 的频率, 则实际观测到的单倍型频率与期望单倍型频率间的差异 D 及 R^2 的计算公式为:

$$D_{ab} = \pi_{AB} - \pi_A \pi_B$$

$$R^2 = (D_{ab})^2 / \pi_A \pi_a \pi_B \pi_b$$

应用 STRUCTURE 2.3.4 软件估测栗杂交子代的群体遗传结构。依据 Evanno 等^[15] 的统计模型, 以最大似然法和 Delta K 值对 235 份杂交子代进行亚群划分。数据录入后, 初始阶段 MCMC 的不作数迭代默认设置为 10000 次, 再将不作数迭代后的 MCMC 调为 100000 次, 群体的数目 (K) 设为 2~13, 以此估计可能的群体数目。计算 Q

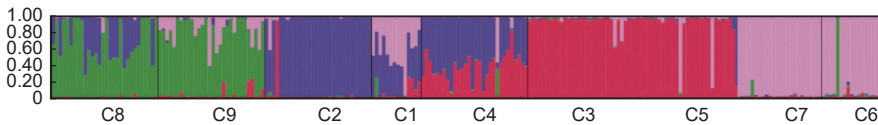


图 3 杂交 F_1 代群体遗传结构

Fig. 3 Genetic structure of F_1 hybrid population

由图 3 可知, 235 份杂交子代群体被划分为 4 个亚群, 而并未按照其杂交组合分为 9 个亚群, 但每个亚群中子代分布呈现出一定的遗传分化。其中, 第 1 亚群主要包括组合 C8 (‘魁栗’ × YLZ 15 号) 和 C9 (‘魁栗’ × ‘YLZ 1 号’), 且有少部分组合 C2、C6 和 C7 的子代混入其中; 第 2 亚群主

参数, 作为关联分析的协变量。

采用 TASSEL 5.0 软件中的一般线性模型 (General linear model, GLM) 和混合线性模型 (Mixed linear model, MLM) 对 SSR 标记和农艺性状进行关联分析。GLM 以 Q 作为协变量进行回归分析; MLM 采用 $Q + K$ 方法分析。

2 结果与分析

2.1 杂交 F_1 代群体遗传结构分析

群体遗传结构是影响连锁不平衡 (LD) 水平的一个重要因素, 亚群数量的增加会使连锁不平衡的程度增强, 可能导致多态性基因位点与目标性状间产生伪关联, 造成假阳性结果的出现, 因此, 在进行关联分析时对群体遗传结构的分析是不可或缺的一步。本研究结果显示, 当 $K=4$ 时, 图像出现明显的拐点 (图 2), Delta K 值也接近最大。据此, 235 个杂交 F_1 代最终被划分为 4 个亚群 (图 3), 平均混合度为 0.120。

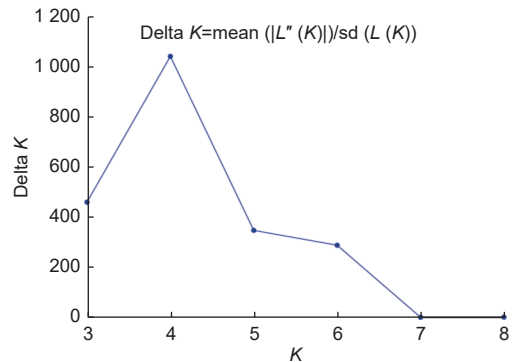


图 2 不同 K 值下 Delta K 值的变化
Fig. 2 Change of Delta K value under different K values

要为组合 C2 (‘YLZ 26 号’ × YLZ 15 号)、C4 (‘YLZ 24 号’ × YLZ 15 号) 和 C1 (‘YLZ 26 号’ × YLZ 14 号), 同样有少部分组合 C3、C5 和 C7 的子代混入其中; 第 3 亚群主要包括组合 C3 (‘YLZ 24 号’ × ‘YLZ 1 号’) 和 C5 (‘YLZ 1 号’ × ‘YLZ 24 号’), 少部分组合 C6 和 C7 的子代混入

其中; 而第4亚群则主要为组合 C6 ('YLZ 1号' × 'YLZ 2号') 和 C7 (YLZ 14号 × 'YLZ 1号'), 组合 C9 的个别子代混入其中。此结果与前期研究中 UPGMA 和 PCA 分类结果基本一致^[12]。

2.2 SSR 标记间连锁不平衡分析

利用 TASSEL 5.0 软件对 32 对 SSR 标记进行连锁不平衡 (LD) 分析, 根据标记间 R^2 来确定两两标记间的关联程度, 利用 R 语言 ggplot 和 pheatmap 软件包对所得数据进行可视化分析

(图4)。可以看出, 统计概率 $P < 0.05$ 时, 32 对 SSR 标记组成 496 个位点中存在一定的连锁不平衡 (图4 中对角线上面淡红及红色位点), 其中, 74 个位点的连锁不平衡水平较高, 占总标记对的比例为 14.92%, 其余位点的连锁不平衡水平相对较低; 而 $P < 0.01$ 时, 大多数位点间连锁不平衡水平较低, 且有 37 个标记对的 R^2 值为 0, 表现出完全的连锁平衡现象。LD 分析结果表明 32 个 SSR 位点间连锁不平衡总体水平较低。

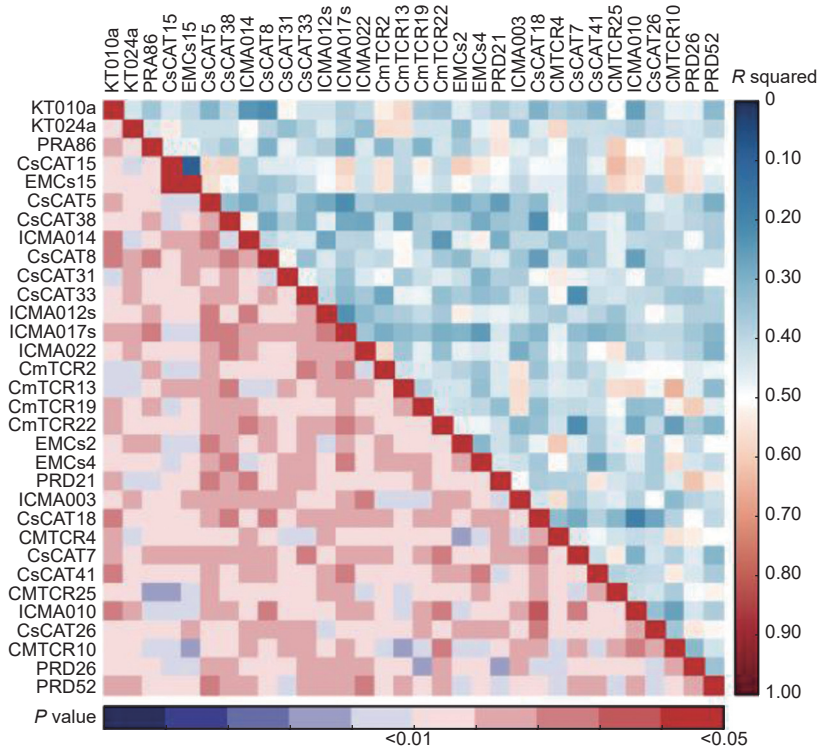


图4 32个SSR标记间LD的分布情况

Fig. 4 The distribution of LD among 32 SSR markers

2.3 农艺性状与SSR标记的关联分析

利用 TASSEL 5.0 软件一般线性模型 (GLM) 和混合线性模型 (MLM) 对杂交子代涵盖生长、枝条、叶片表型及其光合生理的 25 个农艺性状 (连续 2 a 数据平均值) 与 32 个 SSR 标记进行关联分析, 并结合比较包法隆尼 (Bonferroni threshold, $P < 1/1048 = 9.5 \times 10^{-4}$) 阈值, 对标记进行筛选, 结果见表 2。

GLM 模型中, 高达 28 个 SSR 标记位点与树高等 24 个农艺性状呈极显著关联, 每个性状可检测到关联的 SSR 位点数 1~22 个不等, 解释率范围为 4.65%~24.02%。不同性状关联的位点数不

同, 3 个生长性状中, 与树高相关联的位点数最多, 达 16 个, 其中位点 ICMA017s、PRD21 关联解释率相对较高, 分别能解释 12.06%、11.60% 的树高变异, 而其它位点的解释率均不足 10.00%。3 个枝条性状中, 关联位点数最多的性状为节间距, 但其位点解释率偏低, 均在 8.00% 以下。在叶片表型及其光合生理性状中, 叶片宽度、叶面积及叶片干质量关联的位点数最多, 均达 22 个。其中, 叶片宽度、叶面积均与位点 CsCAT7 表现出较高的关联解释率, 分别为 16.70%、18.73%; 与叶片干质量关联解释率最高的位点是 ICMA017s, 能解释 15.50% 的叶片干质量变异; 部分叶片性状

表 2 GML 和 MLM 模型关联标记及其对农艺性状的解释率

Table 2 Explanation ratio of associated markers with agronomic traits in GML and MLM models %

性状 Trait	树高 Height/ m		地径 Ground diameter/cm		冠径 Crown diameter/m		1年生枝条粗度 Annual branch thickness/mm		1年生枝条长度 Annual branch length/cm		节间距 Knot spacing/cm	
	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM
均值 Mean	6.23 ± 0.80		15.44 ± 3.23		5.08 ± 0.70		6.32 ± 0.61		42.38 ± 5.08		2.63 ± 0.30	
标记 Marker	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM
CmTCR4	5.50	6.21					8.24	8.51	11.02	11.08		
CmTCR10			5.24	5.35							5.65	5.82
CmTCR13	6.15	6.15					6.75	6.97				
CmTCR19												
CmTCR22	8.85	8.85	5.56	5.59			5.25	5.19	5.01	5.06	5.76	5.73
CmTCR25												
CsCAT15	7.09	7.09									5.55	5.69
CsCAT18											7.03	6.96
CsCAT26	4.81											
CsCAT38												
CsCAT41												
CsCAT5	7.02	7.02	8.07	8.10								
CsCAT7	8.81	8.81	5.59	5.62			5.01	5.19				
CsCAT8	8.15	8.51										
EMCs15	8.60	8.60	4.88		7.50	8.19			6.68	6.79		
EMCs2												
EMCs4	5.06				5.11							
ICMA003											6.47	6.40
ICMA010	7.55	7.55	9.48	9.52			4.67	5.10			7.93	7.80
ICMA014												
ICMA017s	12.06	12.06	13.79	13.88	8.59	12.78					6.19	6.23
ICMA022	6.57	6.57										
KT010a	5.84	5.84										
KT024a							8.66	8.78				
PRA86	7.75	7.74	5.17									
PRD21	11.60	11.60	7.49	7.50	5.67							
PRD26							11.75	12.21			7.10	7.22
PRD52							5.45	5.59			6.64	8.58
性状 Trait	叶柄长度 Petiole length/mm		叶片长度 Leaf length/cm		叶片宽度 Leaf width/cm		叶面积 Leaf area/cm ²		叶片厚度 Leaf thickness/mm		叶形指数 Leaf index	
均值 Mean	15.68 ± 2.74		16.87 ± 1.47		5.84 ± 0.61		71.29 ± 11.96		0.178 ± 0.030		2.93 ± 0.28	
标记 Marker	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM
CmTCR4			6.09	6.18	6.68	6.68	8.02	8.23	10.75	10.75		
CmTCR10	6.08	5.97	4.81		6.08	6.72	6.65	6.70			6.19	
CmTCR13			7.09	7.10	6.82	6.82	9.86	9.84	7.07	7.07		
CmTCR19					4.82							
CmTCR22			6.63	6.61	8.85	8.85	8.28	8.41	7.02	9.03	5.79	5.81
CmTCR25												
CsCAT15			8.25	8.29	10.83	10.83	12.47	12.64				

· 续表 2 ·

性状 Trait	叶柄长度 Petiole length/mm		叶片长度 Leaf length/cm		叶片宽度 Leaf width/cm		叶面积 Leaf area/cm ²		叶片厚度 Leaf thickness/mm		叶形指数 Leaf index	
	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM
均值 Mean	15.68 ± 2.74		16.87 ± 1.47		5.84 ± 0.61		71.29 ± 11.96		0.178 ± 0.030		2.93 ± 0.28	
标记 Marker	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM
CsCAT18			7.92	7.93	12.24	12.24	9.43	9.46				
CsCAT26					9.26	9.26	8.98	9.03	7.92	7.92	6.70	6.73
CsCAT38							4.66		6.62			
CsCAT41												
CsCAT5			7.20	7.22	8.63	9.31			5.93			
CsCAT7			8.07	8.09	16.70	16.70	18.73	18.51	8.48	14.69	5.15	5.21
CsCAT8					6.48	8.31	6.98	7.07				
EMCs15							4.90		24.02	24.02		
EMCs2			7.23	7.36	6.65	6.65	7.91	8.19				
EMCs4			8.53	8.53	8.04	8.04	9.86	9.85				
ICMA003			7.83	7.80	5.84	6.73	6.20	6.20				
ICMA010			5.70	6.59	11.91	11.91	11.49	11.59			6.26	6.25
ICMA014	5.86	6.48							5.98			
ICMA017s			9.72	9.81	16.69	16.69	14.17	17.46	6.85	9.07	5.34	5.36
ICMA022												
KT010a			6.40	5.84	7.10	7.10	5.62	5.90				
KT024a	5.16		4.93		16.16	16.16	12.59	12.50			8.35	8.34
PRA86			10.43	10.41	14.08	14.08	13.61	13.61	10.90	15.87	11.23	11.28
PRD21			10.45	10.52	11.13	11.13	10.18	10.12	5.56		10.02	10.08
PRD26			7.62	7.62	12.43	12.43	12.18	12.22	6.58	7.19	5.15	5.60
PRD52			11.92	11.93	8.89	8.89	10.47	10.43				
性状 Trait	叶片鲜质量 Fresh leaf weight/g		叶片干质量 Dry leaf weight/g		比叶质量 Specific leaf weight/(g·cm ⁻²)		叶绿素a Chlorophyll a/ (mg·g ⁻¹)		叶绿素b Chlorophyll b/ (mg·g ⁻¹)		叶绿素总量 Chlorophyll a + b/ (mg·g ⁻¹)	
均值 Mean	1.43 ± 0.31		0.68 ± 0.14		0.009 5 ± 0.001 0		0.42 ± 0.08		0.23 ± 0.03		0.65 ± 0.11	
标记 Marker	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM
CmTCR4	9.96	10.25	11.98	12.53	11.62	11.39						
CmTCR10	7.30	6.22										
CmTCR13	9.65	9.67	11.00	11.15								
CmTCR19									4.68			
CmTCR22	6.16	6.24	6.91	7.06	8.27	7.23	5.05	5.24	7.05	7.69	5.64	5.94
CmTCR25			5.86	6.40	6.63	7.11						
CsCAT15	7.41	7.56	8.37	8.65								
CsCAT18	9.12	9.20	13.17	13.35			5.02				4.65	
CsCAT26	5.72	6.20	5.78	6.25	7.77							
CsCAT38	4.97	5.21	6.60	7.11								
CsCAT41												
CsCAT5	8.77	8.78	10.13	10.20	6.16	5.79			5.89		4.78	
CsCAT7	11.36	11.13	12.46	11.99	9.27	8.16			4.91	5.86		
CsCAT8	6.24	6.30	7.11	7.23	5.15							
EMCs15	12.07	11.89	14.04	13.62	16.20	15.25						

性状 Trait	叶片鲜质量 Fresh leaf weight/g		叶片干质量 Dry leaf weight/g		比叶质量 Specific leaf weight/(g·cm ⁻²)		叶绿素a Chlorophyll a/(mg·g ⁻¹)		叶绿素b Chlorophyll b/(mg·g ⁻¹)		叶绿素总量 Chlorophyll a + b/(mg·g ⁻¹)	
均值 Mean	1.43 ± 0.31		0.68 ± 0.14		0.0095 ± 0.0010		0.42 ± 0.08		0.23 ± 0.03		0.65 ± 0.11	
标记 Marker	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM
EMCs2	7.96	8.36	10.68	11.48	4.65							
EMCs4	7.45	10.04	9.10	8.95								
ICMA003	5.16	5.09	5.84	6.30					4.77			
ICMA010	14.63	14.75	14.99	15.22					5.93	5.66	4.90	
ICMA014												
ICMA017s	13.73	13.86	15.50	15.76	7.52	7.16			6.09		4.97	
ICMA022												
KT010a			5.04	5.16								
KT024a	6.92	6.78	6.44	6.18								
PRA86	10.27	11.90	12.88	12.69	14.45	13.30						
PRD21	7.54	7.47	8.38	8.31								
PRD26	9.66	9.67	9.89	9.93	8.23	7.74						
PRD52												
性状 Trait	类胡萝卜素 Carotenoids/ (mg·g ⁻¹)		叶脉分枝角 Branch angle of vein/(°)		叶缘齿数 Number of leaf edge teeth		叶尖角 Leaf tip angle/(°)		叶基角 Leaf basal angle/(°)		叶脉数 Leaf vein number	
均值 Mean	0.47 ± 0.13		58.37 ± 3.84		30.88 ± 2.89		25.81 ± 7.16		112.98 ± 11.68		34.93 ± 2.70	
标记 Marker	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM
CmTCR4							5.46	5.60				
CmTCR10							5.87	5.96			5.16	
CmTCR13							8.89	9.93				
CmTCR19											5.90	
CmTCR22									6.36	7.66		
CmTCR25											5.07	
CsCAT15												
CsCAT18	5.21						7.26	7.30	6.40	6.40		
CsCAT26			4.95									
CsCAT38												
CsCAT41							4.67					
CsCAT5							6.15	6.25				
CsCAT7									5.96	5.96		
CsCAT8							4.97					
EMCs15											5.47	
EMCs2												
EMCs4											7.95	7.95
ICMA003					5.58	5.68	8.80	8.88			5.33	
ICMA010			8.44	8.54								
ICMA014												
ICMA017s												
ICMA022												
KT010a									5.06			
KT024a												
PRD86											9.47	9.47
PRD21											5.76	
PRD26									7.20	7.20		
PRD52							5.07	5.04				

关联的位点较少, 如类胡萝卜素仅与位点 CsCAT18 相关联, 叶缘齿数仅与 ICMA003 相关联, 叶绿素 a 仅与 CmTCR22 和 CsCAT18 相关联, 叶脉分枝角仅与 CsCAT26 和 ICMA010 相关联。此外, 不同位点关联的性状数也不相同, 其中位点 CsCAT41 仅能解释 4.67% 的叶尖角变异, 位点 ICMA022 仅能解释 6.57% 的树高变异, 而位点 CmTCR22 与树高、地径等 17 个农艺性状关联, 关联的性状最多, 但解释率均在 10.00% 以下。

MLM 模型中, 得出 26 个 SSR 位点与树高等 23 个农艺性状呈极显著关联, 解释率范围为 5.04%~24.02%。与 GLM 模型关联分析的结果相比, 部分性状关联的位点数以及部分位点关联的性状数均有所减少, 但解释率差异不明显, 其中位点 CsCAT18 不再与类胡萝卜素和叶绿素 a 关联, 位点 CsCAT26 不再与叶脉分枝角关联。

综合两种模型关联分析结果, 发现 15 个 SSR 标记与树高、地径、冠径 3 个生长性状关联, 14 个 SSR 标记与 1 年生枝条长度、1 年生枝条粗度、节间距 3 个枝条性状关联, 26 个 SSR 标记与叶片长度等 18 个叶片表型及其光合生理性状关联。进一步对各性状高度关联的 SSR 标记进行筛

选分析, 最终得出 ICMA017s 等 15 个标记分别与树高等 13 个农艺性状高度关联, 标记对性状的解释率均在 10.00% 以上 (表 3)。其中, ICMA017s 与 3 个生长性状高度关联, PRD21 与树高性状高度关联, 表明这 2 个 SSR 标记与栗树体生长高度相关; PRD26、CmTCR4 分别与 1 年生枝条粗度、1 年生枝条长度高度关联, 表明这 2 个标记与栗枝条生长发育高度相关; PRA86 等 15 个标记分别与代表叶片大小、形状、厚度以及质量的 8 个表型性状高度关联, 表明这些标记与栗叶片生长发育高度相关, 同时, 叶片各性状高度关联的标记数量差异较大, 其中叶片干质量高度关联的标记数量达 10 个, 其次为叶片宽度和叶面积, 均为 9 个, 而叶形指数关联的标记数量最少, 仅 2 个。此外, 还存在同一标记同时与多个性状高度关联的现象, 如 ICMA017s 与树高、地径、冠径、叶片宽度、叶面积、叶片鲜质量及叶片干质量均存在高度关联, PRD21 与树高、叶片长度、叶片宽度、叶形指数均高度关联, CmTCR4 与 1 年生枝条长度、叶片厚度、叶片鲜质量、叶片干质量及比叶质量均高度关联, 表明某些标记位点可能同时控制多个性

表 3 性状高度关联 SSR 标记统计

Table 3 The traits with high associated SSR markers

性状 Trait	显著关联标记 Significant associated markers	解释率 Explanation ratio
树高 Height	ICMA017s、PRD21	11.60%~12.06%
地径 Ground diameter	ICMA017s	13.79%~13.88%
冠径 Crown diameter	ICMA017s	12.78%~12.78%
1年生枝条粗度 Annual branch thickness	PRD26	11.75%~12.21%
1年生枝条长度 Annual branch length	CmTCR4	11.02%~11.08%
叶片长度 Leaf length	PRA86、PRD21、PRD52	10.41%~11.93%
叶片宽度 Leaf width	CsCAT7、CsCAT15、CsCAT18、ICMA010、ICMA017s、KT024a、PRA86、PRD21、PRD26	10.83%~16.70%
叶面积 Leaf area	CsCAT7、CsCAT15、ICMA010、ICMA017s、KT024a、PRA86、PRD21、PRD26、PRD52	10.12%~18.73%
叶片厚度 Leaf thickness	CmTCR4、EMCs15、PRA86	10.75%~24.02%
叶形指数 Leaf index	PRA86、PRD21	10.02%~11.28%
叶片鲜质量 Fresh leaf weight	CmTCR4、CsCAT7、EMCs15、ICMA010、ICMA017s、PRA86	10.25%~14.75%
叶片干质量 Dry leaf weight	CmTCR4、CmTCR13、CsCAT5、CsCAT7、CsCAT18、EMCs15、EMCs2、ICMA010、ICMA017s、PRA86	10.13%~15.76%
比叶质量 Specific leaf weight	CmTCR4、EMCs15、PRA86	11.39%~16.20%

状,且某一性状同时受多个标记位点控制。

3 讨论

农艺性状大多是受多基因控制的数量性状,且各性状间存在一定的关联,通过常规育种技术对农艺性状进行遗传改良往往效果不显著,且周期长、效率低。在 DNA 水平利用分子标记对农艺性状进行早期选择和辅助育种前景广阔。本研究以涵盖 9 个杂交组合的 235 份栗杂交子代混合群体为材料,在进行群体遗传结构和连锁不平衡(LD)分析的基础上,利用 32 个高多态性 SSR 标记与 25 个农艺性状进行关联分析,以获得与农艺性状相关联的 SSR 标记位点,从而为栗属植物分子标记辅助育种特别是杂交子代的选择和高效育种奠定基础。

连锁不平衡也称为配子相不平衡(Gametic phase disequilibrium)、配子不平衡(Gametic disequilibrium)或等位基因关联(Allelic association),是指群体内不同座位等位基因(可以是标记,也可是基因/QTL 间与标记)间的非随机关联^[16]。LD 在染色体上的分布一般用 LD 衰减散点图和 LD 配对检测的矩阵图来描述,前者可以观测 LD 随遗传或物理距离的衰减速率,后者可以直接观测同一染色体的基因座或基因的多态性位点间 LD 的线性排列^[17]。本研究 LD 分析采用了后者,结果较为直观,统计概率 $P < 0.05$ 时,32 对 SSR 标记组成 496 个位点中存在一定的连锁不平衡,其中,74 个位点的连锁不平衡水平较高,占总标记对的 14.92%,其余位点的连锁不平衡水平相对较低;而 $P < 0.01$ 时,32 个 SSR 位点间连锁不平衡总体水平较低。LD 水平决定了关联分析的解析精度,是开展关联分析研究的理论基础。LD 的一个明显特性是群体依赖性,选择的群体不同,其 LD 水平显著不同。当所用群体来源有限时,其 LD 将维持在一个较高的水平。而选用多样性较高的群体则包括更多不同来源的研究个体,其 LD 水平一般较低;同时,群体混合可通过引进不同祖先来源和等位基因频率的染色体而影响群体的 LD 水平^[18]。本研究所使用的材料为涵盖 9 个杂交组合 235 份杂交子代的混合群体,其亲本涉及锥栗和板栗 2 个物种的 7 个优良品种或优株无性系,并通过杂交即基因的交换和重组产生了

大量的遗传变异类型,来源较为复杂多样,且前期研究表明不同组合子代群体具有丰富的遗传多样性(Shannon's 多样性指数范围 0.8816~1.1317)^[12],其 LD 水平较低。

群体结构的存在和亚群内等位基因频率的不均等分布将导致多态性位点和表型的假阳性关联,分析群体遗传结构对表型性状的影响,对于防止假阳性关联是非常必要的^[19]。因此,群体结构分析是关联分析的前提。本研究所用的栗不同杂交组合群体并非自然群体,在一定程度上不同组合间存在一定的亲缘关系,而对供试群体进行结构分析可将群体分为适宜的几个亚群,将群体结构信息作为协变量来进行关联分析,可以提高结果的准确性,在一定程度上使得假阳性得到控制。群体遗传结构分析表明,当 $K=4$ 时,Delta K 值接近最大。据此,235 个杂交 F_1 代群体最终被划分为 4 个亚群,且每个亚群中子代分布也呈现出一定的遗传分化,平均混合度为 0.120。此结果与前期研究中 UPGMA 和 PCA 分类结果基本一致。

关联分析又称连锁不平衡作图,是一种以基因间连锁不平衡为基础,鉴定某一群体内目标性状与标记间关系的分析方法。与连锁作图分析相比,关联分析所用群体具有更为广泛的遗传基础,可同时对同一基因座的多个等位基因进行分析,而绝大部分常规 QTL 作图所用群体通常为 2 亲本杂交重组后代,其基因座一般只涉及 2 个等位基因;同时,关联分析作图定位更为准确和具有更高的分辨率,可实现对 QTL 的精细定位,甚至直接定位到基因本身^[20]。因此,关联分析与 QTL 作图分析既互为补充,又比连锁作图分析有着更为广泛的应用。曹珂等^[21]以分布于桃(*Prunus persica* L.) 8 个连锁群上的 53 个 SSR 标记与 104 份桃地方品种单果质量及物候期性状表型数据进行关联分析,得到 27 个与桃单果质量及 6 个物候期性状关联的 QTLs。作者等^[11]以来自山东等 10 个省份 95 个板栗地方品种为试材,将 17 对 SSR 标记与板栗地方品种农艺性状进行关联分析,经假阳性检验发现叶柄长度和淀粉含量分别与标记 CsCAT 5 和 CsCAT 22 呈显著关联。张琳等^[22]以筛选出的 16 对高多态性 SSR 标记对 36 份产油量较好、观赏性较佳的牡丹(*Paeonia suffruticosa* Andr.) 材料进行关联分析,发现 3 个 SSR 位点与单株种子产量、聚

合蒴果数等 9 个性状极显著关联 ($P < 0.01$)。关联分析所使用材料既可以是自然群体, 也可以是多亲本杂交混合群体。如对测交群体进行关联分析, 可以直接定位 F₁ 杂种当代控制目标性状的遗传位点或区域, 进而直接反映出 F₁ 杂种当代的遗传效应, 是对传统连锁分析和关联分析群体定位结果的有益补充^[23]。姚俊修从 16 个鹅掌楸 (*Liriodendron chinense* (Hemsl.) Sarg.) 种间杂交组合中抽取 430 份子代个体, 在测定其树高和胸径两个表型性状的基础上, 利用 101 个 SSR 标记检测每个个体的基因型; 然后利用 Tassel 3.0 软件分析不同标记间的连锁不平衡 (LD 值), Structure 2.3 软件对子代群体进行分群; 在群体结构矫正基础上, 将表型与分子标记进行关联分析, 发现 6 个标记与树高性状相关联, 7 个标记与胸径性状关联^[24]。本研究中, 综合两种模型关联分析结果, 发现 15 个 SSR 标记与树高等 3 个生长性状关联, 14 个 SSR 标记与 1 年生枝条长度等 3 个枝条性状关联, 26 个 SSR 标记与叶片长度等 18 个叶片表型及其光合生理性状关联。进一步对各性状高度关联的 SSR 标记进行筛选分析, 最终得出 ICMA017s 等 15 个 SSR 标记分别与树高等 13 个农艺性状高度关联, 标记对性状的解释率均在 10.00% 以上。

同时, 本研究发现存在同一标记同时与多个性状高度关联的现象, 如 ICMA017s 与树高、地径、冠径、叶片宽度、叶面积、叶片鲜质量及叶片干质量 7 个性状均存在高度关联, CmTCR4 与 1 年生枝条长度、叶片厚度、叶片鲜质量、叶片干质量及比叶质量等 5 个性状均高度关联, PRD21 与树高、叶片长度、叶片宽度、叶形指数 4 个性状均高度关联; 还存在同一性状与多个标记高度关联的现象, 如叶片干质量高度关联的 SSR 标记数量多达 10 个, 其次为叶片宽度和叶面积, 均为 9 个, 叶片鲜质量、叶片长度、比叶质量和树高性状高度关联的标记分别为 5、3、3 和 2 个。表明某些标记位点可能同时控制多个性状, 且某一性状同时受多个标记位点控制。此现象在柳树 (*Salix babylonica* L.)^[25]、日本落叶松 (*Larix kaempferi* (Lamb.) Carr.)^[26]、豇豆 (*Vigna unguiculata* (L.) Walp.)^[27] 的研究中也有发现。说明植物中存在一个基因控制多个农艺性状或者影响这些农艺性状的

部分基因存在紧密连锁的现象, 从而导致标记位点变异具有多效性^[28]。据此推测生长、枝条、叶片表型及其光合生理等农艺性状的这种相关性可能是由于基因的一因多效表达或者是连锁不平衡产生的。

4 结论

供试杂交群体 32 个 SSR 位点间 LD 总体水平较低; 根据 Delta K 值计算结果, 235 个栗杂交 F₁ 代混合群体被划分为 4 个亚群, 且每个亚群中子代分布呈现出一定的遗传分化, 平均混合度为 0.120; 综合 GLM 和 MLM 两种模型关联分析结果, 得出 15 个 SSR 标记与树高等 3 个生长性状关联, 14 个 SSR 标记与 1 年生枝条长度等 3 个枝条性状关联, 26 个 SSR 标记与叶片长度等 18 个叶片表型及其光合生理性状关联。其中, ICMA017s 等 15 个 SSR 标记分别与树高等 13 个农艺性状高度关联, 标记对性状的解释率均在 10.00% 以上, 且存在同一标记与多个性状、同一性状与多个标记高度关联的现象。研究结果为栗属植物分子标记辅助育种特别是杂交子代的选择和高效育种奠定了基础。

参考文献:

- [1] 张宇和, 柳 懿, 梁维坚, 等. 中国果树志: 板栗榛子卷[M]. 北京: 中国林业出版社, 2005: 7-9.
- [2] 田寿乐, 明桂冬, 张美勇, 等. 中国板栗育种进展[C]//中国园艺学会干果分会第4届研讨会论文集, 新疆阿克苏, 2007: 79-84.
- [3] 赵永孝. 我国板栗育种工作的回顾与展望[J]. 落叶果树, 2000 (3): 16-17.
- [4] Zhang X, Zhang H, Li L J, et al. Characterizing the population structure and genetic diversity of maize breeding germplasm in Southwest China using genome-wide SNP markers[J]. BMC Genomics, 2016, 17: 697.
- [5] Inoue E, Ning L, Hara H, et al. Development of simple sequence repeat markers in Chinese chestnut and their characterization in diverse chestnut cultivars[J]. Journal of the American Society for Horticultural Science, 2009, 134(6): 610-617.
- [6] 田 华, 康 明, 李 丽, 等. 中国板栗自然居群微卫星(SSR)遗传多样性[J]. 生物多样性, 2009, 17 (3): 296-302.
- [7] 朱 蕾, 康 明. 板栗和锥栗同域居群的空间遗传结构[J]. 热带亚热带植物学报, 2012, 20 (1): 1-7.
- [8] 艾呈祥, 余贤美, 张力思, 等. 中国部分板栗品种的SSR标记[J]. 农业生物技术学报, 2007, 15 (2): 283-289.
- [9] Nie X H, Wang Z H, Liu N W, et al. Fingerprinting 146 Chinese chestnut (*Castanea mollissima* Blume) accessions and select-

- ing a core collection using SSR markers[J]. *Journal of Integrative Agriculture*, 2021, 20(5): 1277-1286.
- [10] Nishio S, Terakami S, Matsumoto T, *et al.* Identification of QTLs for agronomic traits in the Japanese chestnut (*Castanea crenata* Sieb. et Zucc.) breeding[J]. *The Horticulture Journal*, 2018, 87(1): 43-54.
- [11] 江锡兵, 汤 丹, 龚榜初, 等. 基于SSR标记的板栗地方品种遗传多样性与关联分析[J]. *园艺学报*, 2015, 42 (12): 2478-2488.
- [12] 江锡兵, 章平生, 徐 阳, 等. 栗杂交F₁代SSR标记遗传多样性分析[J]. *园艺学报*, 2021, 48 (5): 897-907.
- [13] 章平生, 江锡兵, 徐 阳, 等. 栗属杂交F₁代生长与枝条性状遗传变异及杂种优势分析[J]. *西北植物学报*, 2020, 40 (9): 1582-1594.
- [14] 章平生, 江锡兵, 龚榜初, 等. 板栗与锥栗杂交F₁代叶片表型变异及杂种优势研究[J]. *植物研究*, 2021, 41 (2): 281-293.
- [15] Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study[J]. *Molecular Ecology*, 2005, 14(8): 2611-2620.
- [16] Gaut B S, Long A D. The lowdown on linkage disequilibrium[J]. *Plant Cell*, 2003, 15(7): 1502-1506.
- [17] Flint-Garcia S A, Thornsberry J M, Buckler E S. Structure of linkage disequilibrium in plants[J]. *Annual Review of Plant Biology*, 2003, 54: 357-374.
- [18] 王荣焕, 王天宇, 黎 裕. 植物基因组中的连锁不平衡[J]. *遗传*, 2017, 29 (11): 1317-1323.
- [19] 杨小红, 严建兵, 郑艳萍, 等. 植物数量性状关联分析研究进展[J]. *作物学报*, 2007, 33 (4): 523-530.
- [20] 谭贤杰, 吴子恺, 程伟东, 等. 关联分析及其在植物遗传学研究中的应用[J]. *植物学报*, 2011, 46 (1): 108-118.
- [21] 曹 珂, 王力荣, 朱更瑞, 等. 桃单果重与6个物候期性状的遗传关联分析[J]. *中国农业科学*, 2012, 45 (2): 311-319.
- [22] 张 琳, 王佩佩, 韩雅祯, 等. 油用牡丹主要相关性状与SSR分子标记的关联分析[J]. *东北林业大学学报*, 2019, 47 (3): 31-37.
- [23] 郭晋杰, 刘文斯, 郑云霄, 等. 基于4个测交群体玉米籽粒品质相关性状关联分析[J]. *农业生物技术学报*, 2019, 27 (5): 809-824.
- [24] 姚俊修. 鹅掌楸杂种优势分子机理研究[D]. 南京: 南京林业大学, 2013.
- [25] 蒋冬月. 柳树优良无性系遗传多样性及其重要性状的关联分析[D]. 北京: 中国林业科学研究院, 2015.
- [26] 刘昌勇. 日本落叶松生长及材性相关基因关联遗传研究[D]. 北京: 中国林业科学研究院, 2016.
- [27] 熊海铮. 豇豆遗传多样性及若干农艺性状关联分析[D]. 杭州: 浙江大学, 2016.
- [28] 张 军, 赵团结, 盖钧镒. 大豆育成品种农艺性状QTL与SSR标记的关联分析[J]. *作物学报*, 2008, 34 (12): 2059-2069.

Genetic Structure and Association Analysis with Agronomic Traits of F₁ Hybrids of Chestnut

JIANG Xi-bing¹, ZHANG Ping-sheng¹, ZHANG Dong-bei², WU Ren-chao², WU Jian², WU Cong-lian², LAI Jun-sheng², GONG Bang-chu¹

(1. Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou 311400, Zhejiang, China;

2. Qingyuan Bureau of Natural Resources and Planning, Zhejiang Province, Qingyuan 323800, Zhejiang, China)

Abstract: [Objective] To analyze the natural allelic variation which controls agronomic characters and obtain the associated SSR loci of chestnut for supporting molecular marker assisted selection and efficient breeding. [Method] A total of 235 chestnut hybrid progeny populations covering 9 cross combinations were used to analyze the genetic structure and linkage disequilibrium (LD) of the populations. And association analysis of 32 highly polymorphic SSR markers with 25 agronomic traits were conducted. [Result] There was a certain linkage disequilibrium in 496 loci of 32 SSR markers when $P < 0.05$, among which 74 loci had a higher level of linkage disequilibrium than the others, accounting for 14.92% of the total marker pairs. However, when $P < 0.01$, the overall level of LD among the 32 SSR loci was low. Genetic structure analysis results showed that when $K=4$, the Delta K value approximated the maximum. The F₁ mixed population was divided into four sub-populations, and the distribution of offspring in each sub-population showed a certain degree of genetic differentiation, with an average degree of mixing of 0.120. In the GLM model, up to 28 SSR loci were significantly associated with 24 agronomic traits, and the associated SSR loci ranged from 1 to 22 for each trait, with interpretation rates ranging from 4.65% to 24.02%. In the MLM model, 26 SSR loci were significantly associated with 23 traits, and the interpretation rates ranged from 5.04% to 24.02%. Based on the correlation analysis results of the two models, 15 SSR markers were associated with 3 growth traits, 14 SSR markers were associated with 3 branch traits, and 26 SSR markers were associated with 18 leaf phenotype and photosynthetic physiological traits. [Conclusion] 15 SSR markers including ICMA017s are highly correlated with 13 agronomic traits, and the interpretation rates of markers for traits are more than 10.00%. In addition, there is a phenomenon that the same marker is highly correlated with multiple traits and the same trait is highly correlated with multiple markers.

Keywords: *Castanea*; hybrids; SSR marker; population genetic structure; association analysis

(责任编辑: 金立新)