

# 中国主要花生品种品质性状关联分析

严 玫, 韩锁义, 董文召, 黄冰艳, 刘 华, 张忠信, 徐 静, 孙子琪, 杜 培, 汤丰收, 张新友

(河南省农业科学院经济作物研究所/国家油料作物改良中心河南花生分中心/农业部黄淮海油料作物重点实验室/  
河南省油料作物遗传改良重点实验室, 郑州 450002)

**摘要:** 以来自全国的 136 个花生主要育成品种及育种骨干亲本作为供试群体, 测定其蛋白质、脂肪及油酸含量等性状。选用 64 个均匀分布在复合遗传图谱不同遗传连锁群上多态性较好的 SSR 标记进行多位点扫描。通过标记的基因型值, 利用 Structure 软件对群体进行结构划分并得到矫正 Q 值, 采用 Tassel 软件中 GLM(Q) 方法将供试花生品种连续 3 年的品质性状与 SSR 标记进行关联分析。结果表明: (1) 依据基因型数据对材料群的结构划分, 供试群体最终可被划分为 5 个亚群, 群体特点和结构分析表明供试的花生材料适合关联分析; (2) 通过关联分析, 发掘与 2010 年、2011 年和 2012 年品质性状显著关联的 SSR 位点分别有 18 个、31 个和 26 个; (3) 通过综合分析, 能够连续 3 年重复检测出与品质性状关联的 SSR 位点 4 个, 共计等位变异位点 40 个。

**关键词:** 花生; 品质性状; 关联分析

## Association Analysis between Quality Traits and SSR Markers in Major Peanut Cultivars Developed in China

YAN Mei, HAN Suo-yi, DONG Wen-zhao, HUANG Bing-yan, LIU Hua,

ZHANG Zhong-xin, XU Jing, SUN Zi-qi, DU Pei, TANG Feng-shou, ZHANG Xin-you

(Industrial Crops Research Institute, Henan Academy of Agricultural Sciences/Peanut Subcenter (Henan), National Oil Crops Improvement Center/Key Laboratory of Oil Crops in Huanghuaihai Plains, Ministry of Agriculture/Henan Provincial Key Laboratory for Oil Crops Improvement, Zhengzhou 450002)

**Abstract:** One hundred and thirty six peanut cultivars including main breeding parents originated from different provinces in China were selected as research populations and the quality traits including protein content, oil content, and oleic acid content were tested. Whole genome of the cultivars were scanned with 64 SSRs of a high level of diversity and located dispersedly on a composite genetic linkage map constructed at HAAS. By using Structure 2.3.1, the genotypes of peanut were used to detect the cultivars and assign individuals to subpopulations by Q value. Then by using Tassel 2.1, GLM (Q) model was successfully established by joining 136 genotypes of germplasm and quality traits obtained in 3 years with Q value of population structure as the covariate. The main results showed that 136 cultivars according to individual genotype were assigned to 5 subpopulations. Because the population structure was consistent with the characters and geographic origins of cultivars respectively, these selected germplasm was proved to be suitable for association analysis. By constructing GLM (Q) with 64 genome-wide SSR loci and quality traits from the year 2010 to 2012, 18, 31, and 26 SSR markers significantly associated with quality traits were detected respectively, 4 of which were identified to be significantly associated with quality traits in 3 years tests and 40 alleles corresponding to these markers were totally identified.

**Key words:** Peanut; quality trait; association analysis

收稿日期: 2013-02-28 修回日期: 2013-03-21 网络出版日期: 2013-10-23

URL: <http://www.cnki.net/kcms/detail/11.4996.S.20131023.1509.004.html>

**基金项目:** 国家重点基础研究发展计划 (“973”计划) 项目 (2011CB109304); 国家花生产业技术体系建设项目 (CARS-14); 国家“十二五”农村领域国家科技计划课题 (“863”计划) 项目 (2013AA102602)

第一作者主要从事花生分子遗传育种研究。E-mail: hnxpzhym06@126.com

通信作者: 张新友, 研究方向为花生遗传育种。E-mail: haasz@126.com

花生是人类重要的油料作物之一,目前全球花生年平均种植面积达 3600 万  $\text{hm}^2$ ,年总产 3800 万 t 左右<sup>[1]</sup>。全世界种植花生面积最大的 6 个国家依次是印度、中国、尼日利亚、塞内加尔、苏丹和缅甸。我国是世界花生主要生产和消费国,自 1993 年以来,生产和消费量一直居世界第 1 位<sup>[2]</sup>。花生作为我国重要的食品工业原料和主要食用植物油来源之一,其品质改良越来越受到育种者的高度重视<sup>[3]</sup>。

关联分析是一种基于连锁不平衡方法(LD, linkage disequilibrium)检测自然群体中基因位点及其等位变异,并将等位基因变异与目标性状联系起来分析其基因作用效应的方法<sup>[4]</sup>。与连锁分析相比,关联分析具有不需要构建作图群体,定位精度高、广度大、多位点的优点<sup>[5-6]</sup>。

自 2001 年 J. M. Thornsberry 等<sup>[7]</sup>首次将关联分析方法引入植物研究领域的 12 年间,该方法已被成功应用于 20 余种植物<sup>[8-17]</sup>。在这些研究中,由于标记数量和基因型分型技术的限制,大部分物种距成功实现全基因组关联分析还有一定的距离,因此绝大部分还是主要利用候选基因法进行关联分析用以寻找效应较大的同源基因。但随着一些模式植物基因组测序的完成和一些重要农作物全基因组测序的即将完成,全基因组关联分析方法已逐步在水稻<sup>[18-19]</sup>、拟南芥<sup>[20-21]</sup>、玉米<sup>[22]</sup>、小麦<sup>[23]</sup>及大豆<sup>[24]</sup>上得以成功应用。

本研究采用全基因组关联分析方法,利用随机均匀分布于花生 17 条拟染色体上的标记信息来估计群体内部个体间的遗传关系(即 Q 值)。由于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现型的假阳性关联<sup>[25]</sup>,为了有效地减少并消除由于群体自然分层对关联分析结果的影响,经过基于 LD 测算后群体基因型的结构分析,将花生种质群体各个亚群对应的 Q 值作为协变量,基于 GLM (general liner model) 模型对 2010 年、2011 年和 2012 年 3 年供试材料的蛋白质含量、脂肪含量和油酸含量等品质性状的表型变异进行回归分析,最终得到一系列与品质性状关联的 SSR 位点及对应标记的表型变异解释率。

## 1 材料与方法

### 1.1 试验材料

**1.1.1 田间试验材料** 本研究所用的 136 份花生品种均由中国农业科学院油料作物研究所于 2010 年提供,材料来自 15 个省份,其中 124 份为建国后

不同时期育成的花生优良品种及育种骨干亲本,另外 12 份为传统育种中被选为骨干亲本的农家品种<sup>[26]</sup>。

试验材料于 2011 年、2012 年在河南现代农业研究开发基地(河南原阳)种植。田间试验为完全随机设计,设 2 次重复,行长 1.5 m,行距 0.4 m,株距 15 cm。田间管理采用常规管理方法。

**1.1.2 分子标记** 选用 64 个 SSR 分子标记,其中 42 个 SSR 标记均匀分布在以郑 8903 × 豫花 4 号遗传图谱为基础、结合其他 3 个遗传群体构建的复合遗传连锁图谱的 17 个连锁群上,另外 22 个标记均匀分布在另外构建的 3 个遗传图谱的连锁群上<sup>[27]</sup>。引物均由上海英俊(invitrogen)公司合成。

### 1.2 试验方法

**1.2.1 品质性状检测** 采用 FOSS Infratec-1255 近红外食品(饲料)品质分析仪对花生子仁品质性状进行测定,检测性状包括蛋白含量(换算系数 5.46)、脂肪含量和油酸含量共 3 个性状<sup>[28]</sup>,其中关于蛋白质含量和油酸含量的近红外检测模型是基于连续多年化学分析测试数据矫正建立起来的,多年检测验证结果表明数据准确可靠。2010 年对由中国农科院油料所繁殖、提供的种子样品进行检测,2011 年和 2012 年田间试验材料收获后,每个花生品种选取 5 株混合样品,从中选取 15 粒子仁进行品质性状检测,2 次重复。

**1.2.2 基因组 DNA 提取及 SSR 分析** 每份花生种质取 2 粒种子,净水中浸泡 24 h,在装有干净河沙的容器中置 26 ~ 28 °C 条件下培养,出苗后采用 CTAB 法抽提叶片 DNA<sup>[29]</sup>,然后用 1% 的琼脂糖凝胶水平电泳检测 DNA 的完整性和纯净度,利用 NanoProp 2000 Thermo scientific 检测 DNA 的质量和浓度。最后将 DNA 浓度稀释至工作浓度 20  $\text{ng}/\mu\text{L}$  用于 PCR 扩增。

SSR 扩增程序及电泳、显色等试验流程参照文献<sup>[30]</sup>。将各 SSR 扩增的每一条带视为一个等位变异位点,参照 Marker DL2000(TAKARA 公司),详细记录每一带带的分子量,依据 SSR 重复单元的大小对条带进行人工矫正,确定等位变异数目和分子量大小,最终形成位点等位变异矩阵<sup>[31]</sup>。

**1.2.3 群体结构(Q 矩阵)分析** 群体结构评估采用 Structure 2.3.1 软件<sup>[32]</sup>进行基于数学模型的类群划分,将群体中各材料划分至对应的亚群。首先设定亚群数  $K$  值范围为 1 ~ 20,进行 5 次迭代运算,将每次运行的 burnin period 和 MCMC burnin period 重

复数分别设为 100000 和 1000000,模型设定中将亲缘关系设为混合,等位基因频率设为相关。

**1.2.4 关联分析** 使用 Tassel 2.1 软件(<http://www.maizegenetics.net/bioinformatics/tassel/>) 的 GLM 程序,将各个品种的 Q 值作为协变量,将 3 个品质性状连续 3 年的数据分别对标记变异进行逐一回归分析<sup>[33]</sup>。

## 2 结果与分析

### 2.1 群体结构

种群结构划分主要由 4 类因素决定:(1)划分种群结构的不同可能性——划分类型数( $K$  值变化)。(2)多次迭代运算后的结果一致性与稳定性。(3)种质相关的基本材料信息作为辅助参考。(4)种群结构划分的不同可能性间  $\Delta K$  的变化<sup>[34]</sup>。

经过 LD 测算,剔除遗传距离近的,最终确定均匀分布于复合遗传图谱连锁群上的 36 个 SSR 标记及对应的 355 个等位基因位点进行 LD 验证,再以基于 Bayesian 模型和遗传距离聚类的方法分析 136 份材料构成的群体的遗传结构<sup>[35]</sup>。通过 Structure 2.3.1 软件每次运算到不同的似然值,选择  $K = 2$  至  $K = 7$  的运算结果作散点图(图 1),结果在  $K = 5$  时其模型后验概率( $\ln P(D)$ )最大。根据选择似然值最大原则,可将整个群体分为 5 个亚群,依据各材料在不同亚群的 Q 值可将其归入对应的亚群(表 1)。

通过品种的基因型数据对群体结构划分的分析,可知当群体样品的等位变异频率特征类型值

$K = 5$  时服从 Hardy-Weinberg 平衡,即判定样本最适的亚群值为  $K = 5$ ,花生品种群体可分成 5 个亚群。根据  $K$  值绘制相应的品种群体结构图(图 2),分析各亚群数目的生物学意义和品种来源地情况发现:以 2010 年的品质性状为例,表现为亚群 I 的蛋白质含量普遍较低的同时脂肪酸含量普遍较高,而亚群 V 的蛋白质含量普遍较高但同时脂肪酸含量较低,这一现象又在 2012 年品质性状的表现中得到了较好的验证。

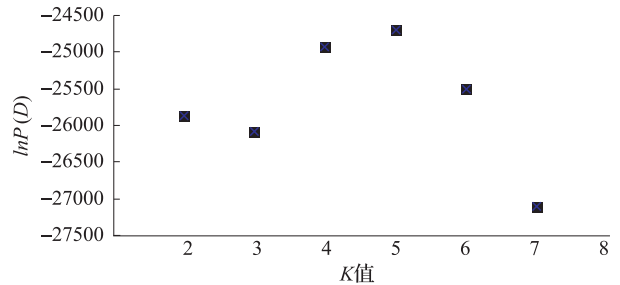


图 1 不同亚群值  $K$  的后验概率  $\ln P(D)$  变异趋势  
Fig. 1 Posterior probabilities ( $\ln P(D)$ ) for each hypothetical number of subpopulations ( $K$ ) in the range of  $K = 2$  to  $K = 7$

另外,花生品种群体亚群划分与分省亚群划分相关,表现为红色代表的亚群 I 基本上来源于黄淮海地区,绿色代表的亚群 II 则由源于长江流域和一些河南花生品种两部分构成,蓝色和黄色代表的亚群 III 和 IV 没有显著的地域特征;粉红色代表的亚群 V 中 70% 为两广和福建沿海地区的花生品种。以上结果说明,亚群 I 和亚群 V 的群体结构有强的独立性,与地理来源关系更为紧密。

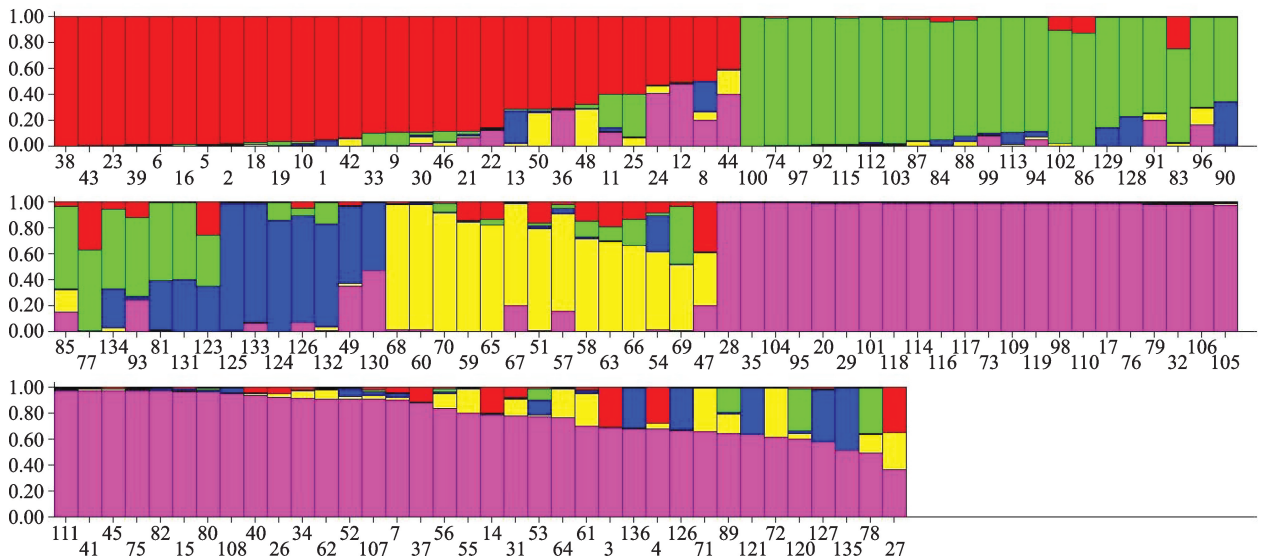


图 2 136 份花生品种依据 Q 值划分的亚群结构图 ( $K = 5$ )

Fig. 2 The bar blot of 136 peanut germplasm classified into 5 clusters by Q factor ( $K = 5$ )

表 1 136 份花生品种结构亚群分类表

Table 1 136 peanut cultivars was structured into 5 subpopulations

品种 Cultivar	来源 Origin	品种 Cultivar	来源 Origin	品种 Cultivar	来源 Origin	品种 Cultivar	来源 Origin	品种 Cultivar	来源 Origin
亚群 I Subpopulation I									
豫花 4 号(7888)	河南	豫花 3 号	河南	开农 30	河南	皖花 5 号	安徽	淮花 8 号	山东
豫花 5 号	河南	豫花 3 号	河南	开农 37	河南	皖花 7 号	安徽	冀油 6 号	河北
豫花 6 号	河南	濮科花 1 号	河南	开农 49	河南	潍花 6 号	山东	冀油 5 号	河北
豫花 8 号	河南	豫花 9326	河南	晋华 7 号	山西	山花 7 号	山东	唐油 4 号	河北
豫花 10 号	河南	豫花 9840	河南	泰花 3 号	江苏	抗青 10 号	山东	鄂花 3 号	湖北
豫花 16 号	河南	远杂 9307	河南	泰花 6 号	江苏	丰花 5 号	山东		
亚群 II Subpopulation II									
徐花 4 号	江苏	皖花 6 号	安徽	伏花生	山东	豫花 9719	河南	开农 36	河南
如东碗儿青	江苏	天府 7 号	四川	山花 10 号	山东	豫花 9620	河南	开农 H053	河南
泰花 4 号	江苏	天府 9 号	四川	花 37	山东	开农 41	河南	冀花 6 号	河北
徐彩花(紫)1 号	江苏	天府 14 号	四川	山花 9 号	山东	开农 8 号	河南	粤花 8 号	广东
徐州 402	江苏	天府 15 号	四川	豫花 1 号	河南	开农 53	河南		
徐州 68-4	江苏	天府 18 号	四川	豫花 9331	河南	开农 60	河南		
亚群 III Subpopulation III									
潍花 9 号	山东	丰花 6 号	山东	山花 8 号	山东	桂花 95	广西	黔花生 4 号	贵州
黔花生 2 号	贵州	嫩花 3 号	黑龙江						
亚群 IV Subpopulation VI									
豫花 15 号	河南	开农 8598	河南	中花 10 号	湖北	天府 8 号	四川	白沙 1016	广东
豫花 9327	河南	中花 5 号	湖北	中花 16	湖北	皖花 8 号	安徽	贺油 11 号	广西
开 51-2	河南	中花 8 号	湖北	鄂花 2 号	湖北	晋花 3 号	山西		
亚群 V Subpopulation V									
湛油 75	广东	粤油 52	广东	66 抗黄 1 号	福建	中琉球	福建	湘花生 1 号	湖南
狮南 3 号	广东	桂花 836	广西	福花 3 号	福建	小琉球	福建	皖花 4 号	安徽
粤油 256-2	广东	桂花 17	广西	福花 4 号	福建	粉红	福建	豫花 14 号	河南
湛油 62	广东	桂花 166	广西	福花 5 号	福建	三莱公	福建	远杂 9102	河南
百日仔	广东	贺油 10 号	广西	福花 8 号	福建	大屁股	福建	豫花 12 号	河南
粤油 20	广东	贺油 12 号	广西	泉花 8 号	福建	浦油 2 号	福建	唐花 10 号	河北
台山珍珠	广东	桂花 22	广西	泉花 327-9	福建	赣花 2881	江西	冀花 7 号	河北
湛油 62	广东	桂花红 35	广西	泉花 646	福建	赣花 2867	江西	黔花生 1 号	贵州
粤油 551	广东	梧油 1 号	广西	泉花 6 号	福建	红梅早	湖北	鲁花 12 号	山东
粤油 92	广东	梧油 4 号	广西	泉花 10 号	福建	鄂花 4 号	湖北	海花 1 号	山东
09-狮头企	广东	梧油 7 号	广西	福花 6 号	福建	芙蓉花生	湖南	花育 20 号	山东
汕油 199	广东	贺粤 1 号	广西	莆花 1 号	福建				

## 2.2 关联分析结果

将 2010 年、2011 年和 2012 年的品质性状分别对 SSR 标记进行关联分析,每年检测出与蛋白质、脂肪和油酸含量等性状呈显著关联的 SSR 标记累计分别是 18 个、31 个和 26 个,关联标记对蛋白质、脂肪和油酸的变异解释率分别为 0.0273 ~ 0.2464、0.0429 ~ 0.2596 和 0.0496 ~ 0.3547(表 2)。从整体上看,与油酸含量关联的 SSR 位点,其关联显著程度和表型变异解释率明显高于与蛋白质含量和脂肪酸含量的关联 SSR 位点。

与蛋白质 2 年及 2 年以上关联到的 SSR 标记共

计 4 个,分别为 pPGPseq3E10、pPGSseq15D3、TC1G04 和 TC2E05,其中 TC2E05 在 3 年环境中均与蛋白质含量相关,且最大变异解释率为 21.99%;2 年及 2 年以上与脂肪显著关联的 SSR 标记共计 4 个,分别为 pPGSseq18G1、TC2E05、TC4G02 和 ARS251;与油酸 2 年及 2 年以上关联到的 SSR 标记共计 10 个,分别为 PM81、pPGPseq1B9、pPGSseq14H6、pPGPseq2D12B、RN26G09、ARS98、ARS124、ARS203、ARS376 和 ARS545,其中 pPGPseq2D12B、ARS376 和 ARS545 在 3 年环境中均与油酸含量相关,最大变异解释率为 35.47%。



表 2 2010-2012 年与品质性状呈显著关联的 SSR 位点及表型变异解释率

Table 2 SSR marker loci associated with quality traits from 2010 to 2012 and their explained phenotypic variation

位点 Locus	蛋白质含量 Protein content			脂肪酸含量 Oil content			油酸含量 Oleic acid content		
	2010	2011	2012	2010	2011	2012	2010	2011	2012
PM81							<b>0.1583</b>		0.0741
PM436		<b>0.0798</b>			<b>0.1604</b>				
PMc588								0.0642	
pPGPseq3E10		0.0483	<b>0.0439</b>			<b>0.1140</b>		0.0775	
pPGSseq17E3					<b>0.1389</b>				
pPGPseq1B9						0.0456		<b>0.1191</b>	0.1371
pPGSseq15C12		0.0803				0.1065			
pPGPseq4H11			<b>0.0490</b>						
pPGSseq14H6								0.083	0.0884
pPGSseq15D3	0.0861	0.0716			<b>0.3307</b>				
pPGSseq18G1				0.0795	0.1378				
pPGPseq2E6						0.1180	<b>0.1872</b>		
pPGPseq5D5				0.1562					
pPGPseq2D12B							0.1007	<b>0.1161</b>	<b>0.1479</b>
PM377					0.1330				0.1369
GC-57						<b>0.2015</b>		<b>0.0797</b>	
GC-94					<b>0.1625</b>				
EM-87									0.0970
TC1A02		<b>0.2464</b>		0.1515					
TC1G04		0.0540	<b>0.0442</b>				0.0789		
TC2E05	0.1102	<b>0.2199</b>	0.0680		0.1605	0.1134			
TC4G02					<b>0.1421</b>	0.1027			
TC6E01					<b>0.2596</b>				
RN26G09							<b>0.1462</b>		0.1215
RN34A10								<b>0.0797</b>	
ARS98					0.0845			<b>0.1375</b>	<b>0.1582</b>
ARS124						0.0478		0.0496	<b>0.0871</b>
ARS127					0.0429				
ARS166							0.0697		
ARS173					<b>0.1481</b>				
ARS180							<b>0.1369</b>		
ARS203	0.1056				0.1276			<b>0.1818</b>	<b>0.3203</b>
ARS251					0.1180	0.0935	<b>0.1626</b>		
ARS303			0.0544					0.1289	
ARS313			0.0296					0.0623	
ARS335			0.0273			0.0571			
ARS369									<b>0.1314</b>
ARS376							0.0965	<b>0.1012</b>	0.1056
ARS392							<b>0.1639</b>		0.0859
ARS415							0.0660		
ARS545			0.0391			<b>0.1043</b>	<b>0.2085</b>	<b>0.2634</b>	<b>0.3547</b>
ARS606		<b>0.0643</b>						0.0833	
ARS644		<b>0.0486</b>		0.0527					
PM50					<b>0.1188</b>				

以粗体格式显示的数值表示标记性状的关联显著程度为  $P < 0.01$

The bold statistics represent related significance of marked traits are at the 0.01 level

本研究全基因组关联分析选用的标记绝大部分是来自分布在已经构建的复合遗传图谱连锁群上的 SSR。因而,通过比对关联结果中的 SSR 位点及其对应的遗传图谱上的图位信息(表 3),可以发现与某单一性状关联的标记位点集中分布在某一个连锁群上。其中 2010 年与性状关联的标记中 PM81、ARS392、ARS251 和 ARS180 集中分布在第 2 连锁群上,共计相关等位变异位点有 32 个。

表 3 连续 2 年以上重复检出与品质性状关联的 SSR 在复合图谱上对应的图位信息

Table 3 The SSRs associated with quality traits in 2 years and their locus information in consensus mapping

连锁群 Linkage group	位点 Locus	复合 图位 (cM) Position	连锁群 Linkage group	位点 Locus	复合 图位 (cM) Position
LG1	ARS303	2	LG7	ARS644	62.6
LG1	ARS415	24.4	LG8	pPGPseq1B9	0
LG2	pPGSseq19G7	0	LG9	ARS545	33.9
LG2	PM81	10.6	LG12	PM436	41.7
LG2	ARS392	51.6	LG12	pPGSseq14H6	47.7
LG2	ARS251	60.3	LG15	ARS203	30.2
LG2	ARS180	73	LG15	TC1A02	52.2
LG7	ARS376	25.2	LG17	pPGPseq3E10	0
LG7	pPGSseq17E3	49.9	LG17	pPGPseq4H11	13.4

连续 2 年与主要品质性状呈显著关联的位点中:与蛋白质含量呈显著关联的 pPGPseq3E10 ( $P < 0.01$ ) 位于 LG17 上;与脂肪酸呈显著关联的位点 ARS251 ( $P < 0.05$ ) 位于 LG2 上;与油酸含量显著关联的位点有 PM81、ARS392、ARS376、pPGPseq1B9、ARS545、pPGSseq14H6 和 ARS203,分别位于复合图谱的 LG2、LG2、LG7、LG8、LG9、LG12 和 LG15 上,其中 ARS376 和 ARS545 连续 3 年与油酸显著关联。

## 3 讨论

### 3.1 SSR 位点与关联分析方法选择

本试验采用全基因组关联分析方法,利用随机均匀分布于花生 17 条拟染色体上的标记信息来估计群体内部个体间的遗传关系,再以群体结构 Q 值作为协变量,基于 GLM(Q) 模型对 3 年的品质性状的表型变异进行回归分析,最终得到与性状显著关联的 SSR 位点。其中用于关联分析的 SSR 位点选

择依据有 2 点:一是分布在构建的复合遗传图的不同 LG 上的 SSR 且遗传多样性较好;二是尽可能选择并利用重组自交系群体郑 9803 × 豫花 4 号中获得的 QTL 位点连锁的 SSR,以便为关联分析与 QTL 分析结果的双向验证创造条件。

目前在植物全基因组关联分析中,仍倾向于选择核心种质或微核心种质作为自然群体,进行重要性状的关联分析<sup>[36]</sup>。已有很多研究证明核心种质是挖掘有利基因、进行关联作图的重要基础材料<sup>[11-13,37-38]</sup>。本研究选择的材料由中国农业科学院油料作物研究所提供,主要由建国后不同历史时期的花生骨干亲本和育成品种构成供试群体。该群体不同于严格意义上的核心种质或微核心种质群<sup>[39]</sup>,其代表的基因群并不能有效覆盖花生 (*Arachis hypogaea* L.) 的整个基因组。然而通过育种家的选择和努力,花生骨干亲本和花生育成品种往往是许多有益性状的聚合体,故该品种群体同微核心种质群体相比,包含了花生种质中的大部分有利基因。

在关联分析中,群体中的 LD 会受到遗传漂变、群体分层和自然选择等多种因素的影响,其中群体分层与目标性状基因效应协同干扰关联结果被认为是导致出现假关联和伪关联的最主要因素<sup>[8]</sup>。鉴于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现型的假阳性关联<sup>[25]</sup>,将花生种质群体各个亚群对应的 Q 值作为协变量,基于 GLM 模型对品质性状的表型变异进行回归分析,可有效地减少并消除由于群体自然分层对关联分析结果的影响。

花生是严格的自花授粉作物,其各类育成品种在自然进化和形成过程中均经历了瓶颈效应和强烈人工选择,因此其包含的等位基因较少,故采用全基因组扫描的方法进行关联分析是可行的,同时也能挖掘出较多的优异等位变异位点。

### 3.2 品质性状关联结果

本研究利用随机均匀分布于花生 17 条拟染色体上的 64 个 SSR 标记对 136 份花生骨干亲本进行全基因组多位点的扫描分析,在对供试材料进行基于混合模型的群体结构分析的基础上,将连续 3 年的花生主要品质性状对 SSR 标记进行关联分析。在矫正后群体中,共检测出 4 个 SSR 标记连续 3 年与品质性状显著关联:与蛋白质含量显著关联的位点为 TC2E05;与油酸性状显著关联的位点有 3 个,分别为 pPGPseq2D12B、ARS376 和 ARS545;没有发

现连续 3 年与脂肪性状显著关联的 SSR 标记。其原因可能在于,脂肪含量为综合数量性状,遗传较为复杂,受环境影响大,而且本研究所选用的标记数量有限,故有必要通过进一步增加多态标记的数量及重复表型鉴定来深入研究。

利用 2011 年、2012 年河南原阳地区试验数据重复检测出与脂肪含量显著关联 ( $P < 0.05$ ) 的位点是 pPGPseq2E6、TC2E05、TC4G02 和 ARS251,而利用姜慧芳提供的 2012 年同一批在武汉地区的试验数据分析,仅发现 3 个与性状相关的 SSR 位点 PMc588 ( $P < 0.05$ )、RN26G09 ( $P < 0.001$ ) 和 GM401 ( $P < 0.001$ ),表明两地与脂肪含量关联的位点差异显著。

综合以上结果不难看出,在 3 类主要品质性状中,油酸含量是受环境因素影响最小的性状,其次是蛋白质含量,而脂肪含量则较易受环境和气候变化的影响,致使其关联结果也随着环境和气候的变化而改变。X. Y. Zhang 等<sup>[40]</sup>利用重组自交系开展的遗传分析表明,脂肪遗传受 2 对连锁互补的主基因加多基因控制,主基因遗传率为 22.88%;油酸遗传受 2 对主基因控制,主基因遗传率为 75.11%;黄冰艳等<sup>[41]</sup>利用 4 个  $F_2$  群体所做的遗传分析表明,花生的油酸含量受 2 对主基因控制,同时存在着基因互作与多基因效应,主基因遗传力分别为 66% ~ 89%。这些研究结果虽然因所用材料不同而不完全一致,却共同说明一个问题:相对于脂肪含量的遗传,花生油酸含量的遗传比较简单。

张新友等<sup>[42]</sup>利用重组自交系群体郑 9803 × 豫花 4 号进行的 QTL 分析表明,标记 ARS303 与蛋白质的 QTL 位点紧密连锁,本研究在 2012 年关联分析结果中也发现 ARS303 与蛋白质含量相关联 ( $P < 0.05, 0.0544$ ),说明标记 ARS303 很可能是与蛋白质含量性状相关的 QTL 位点,今后在高蛋白质材料的筛选中可加以应用。此外在关联分析中,2 年重复检出与蛋白质含量呈显著关联的位点 pPGPseq3E10 也同时与百果重和侧枝长的 QTL 位点紧密连锁,与脂肪含量呈显著关联的位点 ARS251 同时与结果枝数的 QTL 位点紧密连锁;与油酸含量呈显著关联的位点有 PM81、ARS376 和 ARS203,其中 PM81 也是侧枝长、结果枝和蛋白质含量共同的 QTL 位点,ARS376 在侧枝长 QTL 位点附近 7 cM 左右,ARS203 同时也是单株果数 QTL 位点。综合比较本研究中品质性状的关联位点与复合遗传连锁图谱上对应的 QTL 位点,结果表明在不同性状定位的方法

中,同一 SSR 位点对应的定位性状差异较大。推测出现这种现象的原因有 2 个:一是与所用试验材料的遗传基础有关,基于重组自交系为基础的遗传连锁图谱的 QTL 定位是通过高世代群体对 2 个亲本的遗传解析,而关联分析所用的材料数目和类型均多于 QTL 分析,为遗传解析提供了更为丰富多样的遗传背景资料;二是可能与遗传分析所采用的数学模型不同有关,2011 年国际半干旱所 (ICRISAT, international crops research institute for the semi-arid tropics) 的 C. Sarvamangala 等<sup>[43]</sup>曾尝试通过对 TG26 × GPBD4 高代 ( $F_9: F_{10}$ ) 的 146 个家系同时采用单标记分析 (SMA, single marker analysis) 和复合区间作图 (CIM, composite interval mapping) 2 种方法来定位与蛋白质含量、脂肪酸含量、油酸和亚油酸含量及油亚比相关的 QTL 位点,结果发现不同方法所得与各品质性状的 QTL 位点紧密连锁的 SSR 位点只有部分完全相同,这充分说明统计方法不同分析结果有明显差异。

通过比对与性状关联位点和其对应的遗传图谱上的图位信息可知,虽然本试验选用的均匀分布在 17 个花生拟染色体的众多标记中有 5 个位于 LG2 上<sup>[27]</sup>,但其中就有 4 个与油酸性状呈显著相关,可推测在该段染色体上含有控制油酸含量性状的多个基因位点,将与性状关联的 SSR 位点和其对应图位信息及连锁 QTL 比对,为下一步开展关联分析与数量遗传 QTL 在遗传上的双向验证和数量性状的物理图谱定位奠定了基础。此次关联分析研究结果所发掘的 SSR 位点可用于今后花生优异品质品种选育和分子标记辅助育种,同时对品质性状相关基因发掘有一定的参考价值。

#### 参考文献

- [1] FAO. Statistical database [R]. Rome: Food and Agriculture Organization (FAO) of the United Nations, 2010
- [2] 吴学君. 我国花生及花生产品出口结构与竞争力分析 [J]. 中国油料作物学报, 2010, 32(2): 309-314
- [3] 苏秋芹. 不同株型花生品质性状与主要数量性状的相关分析 [J]. 中国农学通报, 2006, 6(22): 192-194
- [4] Mackay I, Powell W. Methods for linkage disequilibrium mapping in crops [J]. Trends Plant Sci, 2007, 12: 57-63
- [5] 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
- [6] 陈秀辰, 杨旭, 成玉富, 等. 数量性状关联分析及其在蔬菜中的应用 [J]. 中国蔬菜, 2012(16): 1-9
- [7] Thornsberry J M, Goodman M M, Doebley J, et al. Dwarf 8 polymorphisms associated with variation in flowering time [J]. Nat Genet, 2001, 28: 286-289
- [8] 谭贤杰, 吴子恺. 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
- [9] 金亮, 包劲松. 植物性状 - 标记关联分析研究进展 [J]. 分子

- 植物育种,2009,7(6):1048-1063
- [10] Thumma B R, Nolan M F, Evans R, et al. Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. [J]. *Genetics*, 2005, 171:1257-1265
- [11] Hansen M, Kraft T, Ganestam S, et al. Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers [J]. *Genet Res*, 2001, 77(1):61-66
- [12] Wang M L, Zhu C S, Barkley N A, et al. Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection [J]. *Theor Appl Genet*, 2009, 120:13-23
- [13] Skøt L, Humphreys J, Humphreys M O, et al. Association of candidate genes with flowering time and water-soluble carbohydrate content in *Lolium perenne* (L.) [J]. *Genetics*, 2007, 177:535-547
- [14] 顾竟,李玲,宗绪晓,等. 豌豆种质表型性状 SSR 标记关联分析 [J]. *植物遗传资源学报*, 2011, 12(6):833-839
- [15] 李仁伟,王晨,戴思兰,等. 菊花品种表型性状与 SRAP 分子标记的关联分析 [J]. *中国农业科学*, 2012, 45(7):1355-1364
- [16] 危文亮,张艳欣,吕海霞,等. 芝麻资源群体结构及含油量关联分析 [J]. *中国农业科学*, 2012, 45(10):1895-1903
- [17] Cao K, Wang L R, Zhu G R, et al. Genetic diversity, linkage disequilibrium, and association mapping analyses of peach (*Prunus persica* L.) landraces in China [J]. *Tree Genet Genomes*, 2012, 8:975-990
- [18] Huang X H, Wei X H, Sang T, et al. Genome-wide association studies of 14 agronomic traits in rice landraces [J]. *Nat Genet*, 2010, 42(2):159-162
- [19] Jin L, Lu Y, Xiao P, et al. Genetic diversity and population structure of a diverse set of rice germplasm for association mapping [J]. *Theor Appl Genet*, 2010, 121:475-487
- [20] Atwell S, Huang Y S, Vilhjálmsson B J, et al. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines [J]. *Nature*, 2010, 465:627-631
- [21] Eva K, Chan F, Heather C, et al. Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping [J]. *Genet Soc Am*, 2010, 185:991-1007
- [22] Tian F, Bradbury P J, Brown P J, et al. Genome-wide association study of leaf architecture in the maize nested association mapping population [J]. *Nat Genet*, 2011, 43(2):159-162
- [23] Breseghello F, Sorrells M E. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars [J]. *Genetics*, 2006, 172:1165-1177
- [24] Li Y H, Smulders M J M, Chang R Z, et al. Genetic diversity and association mapping in a collection of selected Chinese soybean accessions based on SSR marker analysis [J]. *Conserv Genet*, 2011, 12:1145-1157
- [25] 王荣焕,王天宇,黎裕. 植物基因组中的连锁不平衡 [J]. *遗传*, 2007, 29(11):1317-1323
- [26] 禹山林. 中国花生及其系谱 [M]. 上海:上海科学技术出版社, 2008
- [27] 张新友. 栽培花生产量、品质和抗病性的遗传分析与 QTL 定位研究 [D]. 杭州:浙江大学, 2010
- [28] 黄冰艳,张新友,董文召,等. 河南省花生地方资源蛋白质和脂肪含量分析及育种利用策略 [J]. *植物遗传资源学报*, 2012, 13(3):414-417
- [29] 陈静,胡晓辉,苗华荣,等. CTAB 法抽提花生总 DNA 在 SSR 和 SRAP 中的扩增效果 [J]. *花生学报*, 2008, 37(1):29-31
- [30] Hong Y B, Liang X Q, Chen X P. Construction of genetic linkage map based on SSR markers in peanut (*Arachis hypogaea* L.) [J]. *Agr Sci China*, 2008, 7(8):915-921
- [31] 文自翔,赵团结,郑永战,等. 中国栽培和野生大豆农艺品质性状与 SSR 标记的关联分析 I. 群体结构及关联标记 [J]. *作物学报*, 2008, 34(7):1169-1178
- [32] Pritchard J K, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data [J]. *Genetics*, 2000, 155:945-59
- [33] Edwards Buckler Lab. Maizediversity research [EB/OL]. (2007-01-30) [2013-03-10] <http://www.maizegenetics.net/bioinformatics/tassel/>
- [34] Wang M L, Sukumaran S, Barkley N A, et al. Population structure and marker - trait association analysis of the US peanut (*Arachis hypogaea* L.) mini - core collection [J]. *Theor Appl Genet*, 2011, 123:1307-1317
- [35] Falush D, Stephens M, Pritchard J K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies [J]. *Genetics*, 2003, 164:1567-1587
- [36] 王荣焕,王天宇,黎裕. 关联分析在作物种质资源分子评价中的应用 [J]. *植物遗传资源学报*, 2007, 8(3):366-372
- [37] 黄莉,任小平,张晓杰,等. ICRISAT 花生微核心种质农艺性状和黄曲霉抗性关联分析 [J]. *作物学报*, 2012, 38(6):935-946
- [38] Belamkar V, Selvaraj M G, Ayers J L, et al. A first insight into population structure and linkage disequilibrium in the US peanut mini-core collection [J]. *Genetica*, 2011, 139:411-429
- [39] 董玉琛,曹永生,张学勇,等. 中国普通小麦初选核心种质的产生 [J]. *植物遗传资源学报*, 2003, 4(1):1-8
- [40] Zhang X Y, Zhu S J, Han S Y, et al. Inheritance of fat and fatty acid composition in peanut (*Arachis hypogaea* L.) [J]. *Agr Sci Technol*, 2011, 12(7):943-946, 974
- [41] 黄冰艳,张新友,苗丽娟,等. 花生油酸和亚油酸含量的遗传模式分析 [J]. *中国农业科学*, 2012, 45(4):617-624
- [42] 张新友,韩锁义,徐静,等. 花生主要品质性状的 QTLs 定位分析 [J]. *中国油料作物学报*, 2012, 34(3):311-315
- [43] Sarvamangala C, Gowda M V C, Varshney R K. Identification of quantitative trait loci for protein content, oil content and oil quality for groundnut (*Arachis hypogaea* L.) [J]. *Field Crop Res*, 2011, 122(1):49-59