

陆地棉主要产量相关性状的 SSR 标记关联分析

王娟, 董承光, 刘丽, 孔宪辉, 王旭文, 余渝

(新疆农垦科学院棉花研究所/农业部西北内陆区棉花生物学与遗传育种重点实验室, 石河子 832000)

摘要: 高产优质育种是我国棉花育种的主要目标。寻找与目标性状关联的分子标记, 可克服常规育种的盲目性, 提高分子标记辅助选择育种的准确性。本研究对 118 份陆地棉种质资源的衣分、单铃重、单株铃数及子指等 4 个产量相关性状进行 2 年 2 点的表型鉴定, 并利用覆盖全基因组的、有多态性的 214 对 SSR 标记进行标记与性状的关联分析。结果表明: 118 份材料的 4 个产量相关性状表型变异丰富, 平均变异系数的变幅在 6.1% ~ 19.1% 之间, 且在各环境中表现较为稳定; 基因型分析表明, 214 对标记共检测到 460 个等位变异, 基因多样性指数平均为 0.5151, PIC 值平均为 0.4587, 表明该批标记具有较多的等位变异数和较高的基因多样性; 群体结构分析表明该批材料可分为 4 个亚群, 且各类群中材料与地理来源无对应关系; 关联分析结果显示, 在显著条件下 ($-\log_{10}P > 1.3, P < 0.05$), 共有 39 个标记位点能够在 2 个及 2 个以上的环境中同时检测到, 其中有 4 个标记位点同时与 2 个以上性状相关联, 进一步比较发现, 有 7 个位点与前人研究结果一致, 其余 32 个位点为新发现的位点。研究结果可为陆地棉产量性状遗传改良的分子标记辅助选择提供理论依据。

关键词: 陆地棉; 产量性状; 分子标记; 关联分析

Association Analysis of Yield-related Traits with SSR Markers in Upland Cotton (*Gossypium hirsutum*)

WANG Juan, DONG Cheng-guang, LIU Li, KONG Xian-hui, WANG Xu-wen, YU Yu
(Cotton Research Institute, XinJiang Academy of Agricultural and Reclamation Science/Key Laboratory of China Northwestern Inland Region, Ministry of Agriculture, Shihezi 832000)

Abstract: Improving cotton yield and fiber quality is a major breeding goal for Chinese upland cotton. Discovering molecular markers associated with the target traits can overcome the blindness of conventional breeding and improve the accuracy of molecular marker assisted selection breeding. In this study, a diversity panel consisting of 118 upland cotton accessions was grown in four different environments, and the yield-related traits including lint percentage (LP), boll weight (BW), bolls per plant (BN) and seed index (SI) were measured, and 214 polymorphic SSR markers of covering the whole genome were performed in association mapping. Molecular marker data and four phenotypic traits were analyzed by the method of MLM (mixed linear model) in Tassel 5.0 on the basis of population structure, analysis loci with elite allelic variation and typical materials carrying elite alleles were identified based on phenotypic effect values. The results were showed: 118 upland cotton accessions were rich in phenotypic variation, the mean coefficients of variance (CV) ranged from 6.1% - 19.1%, and were more stable in different environments. The average value of skewness coefficient were between 0.8-0.9 and average values of kurtosis coefficient were between 1.0-2.6. The results showed that phenotypic variation was in normal distribution. Genotyping was performed with genome-wide Simple Sequence Repeats (SSRs), a total of 214 markers were polymorphic and generated 460 allele loci, the average genetic diversity index were 0.5151 with range from 0.0333 to 0.8725, and polymor-

收稿日期: 2016-12-29 修回日期: 2017-02-15 网络出版日期: 2017-06-13

URL: <http://kns.cnki.net/kcms/detail/11.4996.S.20170613.0834.008.html>

基金项目: 新疆生产建设兵团博士资金(2013BB001); 国家自然科学基金(31260340)

第一作者主要从事棉花分子育种研究。E-mail: cottonwj@126.com; 董承光为共同第一作者, 主要从事棉花遗传育种研究。E-mail: dcg318@163.com

通信作者: 余渝, 主要从事棉花遗传育种研究。E-mail: xjyuyu021@sohu.com

phism information content (PIC) were 0.4587 with range from 0.0328 to 0.8594. Then 99 markers, accounting for 46.3% of the total, their genetic diversity index and PIC were both more than average together. The results indicated that the alleles and the genetic diversity were larger in upland cotton accessions. 118 cotton varieties were divided into 4 subgroups by analysis of population structure, there was no corresponding relation between each kind of group of accessions and the geographical regions. Association analysis results were showed that 39 marker loci were detected at least two of the four environments ($-\log_{10} P > 1.3, P < 0.05$). 14 marker loci were associated in LP with explained variance rang from 3.29% to 9.1%, with an average of 5.37%. 8 marker loci were associated in BW with explained variance range from 3.68% to 9.56%, with an average of 5.88%. 5 marker loci were associated in BN with explained variance range from 3.74% to 5.77%, with an average of 4.44% and 12 marker loci were associated in SI with explained variance range from 3.49% to 11.14%, with an average of 5.91%, respectively. The four marker loci of 39 associated markers were associated with more than two traits simultaneously. For example, JESPR065a was simultaneously associated with BW and SI, and NAU1169c was simultaneously associated with LP and SI. The seven marker loci (CIR0183, JESPR0153, NAU3308, NAU6094, NAU3201, NAU1167, NAU3995) were coincident with previous studies, and 32 were identified as new association markers. The results were excellent genetic resources for future cotton breeding and could provide a theoretical basis for molecular assisted selection of upland cotton about yield traits.

Key words: upland cotton; yield traits; molecular marker; association analysis

棉花是我国重要的经济作物之一,在国民经济中占据着重要的地位^[1]。棉花产业对我国经济发展具有十分重要的作用。近年来,由于农业劳动力成本较高及气候等原因,棉农植棉比较效益有所下降,植棉面积也有逐年下降趋势,长江流域棉区和黄河流域棉区表现尤为严重。在这种形势下,提高棉花单位面积产量进而提高植棉效益成为当务之急。棉花衣分、单铃重、单株铃数及子指等性状是棉花的主要产量指标性状,这些性状是由较为复杂的数量性状基因(QTL)控制的。在棉花育种中,开展以上性状的 QTL 定位研究,挖掘出稳定表达的 QTL 和优异等位基因,对提高棉花产量具有重要的理论与应用价值。

经典的研究数量性状遗传分析方法是基于双亲的连锁分析,该方法通过两个纯合亲本杂交构建不同类型的分离群体,利用分子标记技术结合表型性状数据,将控制数量性状的基因座定位到染色体上。利用该方法研究人员通过构建不同双亲的分离群体定位了多个控制棉花产量及纤维品质性状的 QTL^[2-8],为研究棉花数量性状基因的遗传方式提供了重要基础,但由于陆地棉群体标记的数量有限和环境影响,对控制性状的 QTL 定位精度较差,从而限制了后续的图位克隆及分子标记辅助选择等研究。关联分析方法是近年来研究复杂数量性状的一种有效方法,它可快速对目标基因进行精细定位,并针对特定候选基因提供大量信息,验证候选基因的

功能。关联分析与传统的连锁分析相比,具有三大特点:(1)不需要专门构建作图群体,自然群体或种质资源均可作为研究材料,且不需知道其系谱背景,只要存在遗传变异即可;(2)广泛的遗传材料使检测的等位基因数目更多,提高了作图的分辨率;(3)可同时考察多个性状大多数 QTL 的关联位点及其等位变异,使得作图效率更高。因此,该方法目前在玉米^[9]、小麦^[10]、大豆^[11]等作物中得到广泛应用。在棉花上,关联分析研究开展较晚,但发展较快,国内外研究学者已利用该方法对棉花主要农艺性状^[12-13]、产量^[14]、纤维品质^[15-16]、抗逆^[17-18]等进行关联定位,极大地促进了关联分析在棉花遗传和育种研究中的应用,为基因组辅助选择在棉花资源创制、新品种选育中的进一步应用奠定了重要的理论基础。

本研究以我国不同棉区的 118 份陆地棉品种资源为主要研究对象,通过对影响棉花产量的衣分、单铃重、单株铃数及子指等 4 个重要性状的多年多点性状调查鉴定,并结合 SSR 标记的全基因组扫描关联分析研究进一步发掘以上性状的关联位点及优异等位变异,研究结果可为棉花高产育种中分子标记的辅助选择和基因组选择育种提供理论依据。

1 材料与与方法

1.1 试验材料与表型鉴定

试验材料为来自我国不同生态棉区及部分国外

材料,共计 118 份(表 1),其中西北内陆棉区 67 份、北部特早熟棉区 21 份,其他棉区以及国外材料 30 份。试验于 2013-2014 年分别在新疆农垦科学院棉花研究所北疆试验基地(石河子)与新疆农垦科学院棉花研究所南疆试验基地(库尔勒)2 个不同环境种植上述 118 份品种资源材料,各试验点田间试验采用完全随机区组设计,双行区,行长 5 m,株距 10 cm,平均行距 35 cm,3 次重复,以当地常规方法进行田间管理。按照杜雄明等^[19]的棉花种质资源性状描述规范,在棉花成铃后期分别在各重复中每小区选中间连续 10 株调查单株结铃数,收获时选取各材料中部 20 铃室内扎花考种获得衣分、单铃重和子指等性状的表型数据。

表 1 118 份材料来源及分布

Table 1 The origin and distribution of 118 materials

棉区 Cotton growing region	材料数 Accession number	来源 Origin
西北内陆棉区 North-western inland cotton growing region	67	新疆北疆(48);新疆南疆(15);甘肃敦煌(4)
北部特早熟棉区 North China specific early maturation region	21	辽宁(12);山西晋中(9)
其他棉区及国外材料 Other cotton growing region and abroad	30	黄河流域棉区(17);长江流域棉区(8);国外材料(5)

1.2 SSR 标记及基因型鉴定

以 L. Zhao 等^[20]发表的四倍体棉花遗传图谱为参考,平均每隔 8~10 cM 选取一个 SSR 标记,共计 557 对进行多态性筛选,最终获得多态性引物 214 对。SSR 序列来自 CMD(Cottonmarker database)数据库(<http://www.cottonmarker.org>),由生工生物工程股份有限公司(上海)合成。分子标记名称中的大写字母,如 NAU、BNL 等代表引物的来源,“a”、“b”、“c”、“d”代表一个标记在各材料中不同的多态性位点,位点顺序按分子量由大到小顺序排列。各材料的基因组 DNA 提取、SSR 实验操作及 PCR 扩增等参照前人的研究方法进行^[21-22],PCR 产物分析采用毛细管电泳分析仪,操作方法参照说明书进行,数据结果可自动保存到系统自带软件 ProSize2.0 中,根据胶视图中 DNA Ladder 以及单个样品不同位点片段大小,输入预设范围,系统可自动读取 0,1 数据,缺失由“?”代替,并以 Excel 形式输出。

1.3 数据分析

各表型性状的一般性描述统计及各环境性状的趋势分析采用 SPSS19.0^[23]软件进行;分子标记的基因型遗传多样性分析采用 PowerMarkerV3.25 软件^[24];基于基因型的群体结构分析采用 Structure 2.3.1 软件^[25],并得到该 K 值对应的 Q 矩阵;采用 Tassel5.0^[26]软件结合基因型数据生成 K 矩阵,并结合上述各基因型数据、各环境的表型性状值、Q 矩阵,利用混合线性模型 MLM(mixed linear model)进行性状和标记之间的关联分析,并计算标记位点在 $-\log_{10}P > 1.3$ 即 $P < 0.05$ 时对表型变异的贡献率(R^2)。

2 结果与分析

2.1 表型性状数据分析

118 份种质资源材料的 4 个主要产量相关性状 2 年 2 点 4 种环境的表型结果如表 2 所示。结果显示,各性状的平均变异系数的变幅在 6.1%~19.1% 之间,其中单株铃数的变异系数最大为 19.1%,衣分的变异系数最小为 6.1%;各性状的平均偏度系数值在 -0.8~0.9 之间,平均峰度系数值在 1.0~2.6 之间,表现出较小的偏度系数和峰度系数,说明 118 份材料 4 个性状的平均表型值比较符合正态分布,为后续的关联分析提供了很好的表型数据基础;2 年 2 点 4 种环境的 4 个表型性状的 boxplot 图(图 1)较为直观地描述了多年多点性状变化趋势的稳定性,结果表明 4 个性状中变化趋势较稳定的为子指与单铃重,不同环境中波动幅度较大的为单株铃数,表明该性状受环境影响最大,可能与当年气候条件及栽培管理过程中的水肥调控和化学调控等有关。

2.2 基因型数据分析

利用选取的 557 对 SSR 标记对 118 份材料进行扩增分析,结果发现有 214 对标记在群体中产生明显的多态性,占所选标记总数的 38.4%;等位变异位点检测显示 214 个标记共检测到 460 个等位变异位点,平均每个标记检测到 2.15 个,变幅为 1~5 个,其中检测到 2 个以上等位变异位点的标记共 166 个,占多态性标记总数的 77.6%;214 个标记所揭示的基因多样性指数平均为 0.5151,变幅为 0.0333~0.8725(图 2a);PIC 值平均为 0.4587,变幅为 0.0328~0.8594(图 2b),基因多样性指数和 PIC 值均大于平均数的标记有 99 个,占总标记数的 46.3%。表明该批标记具有较多的等位变异数和较高的基因多样性。

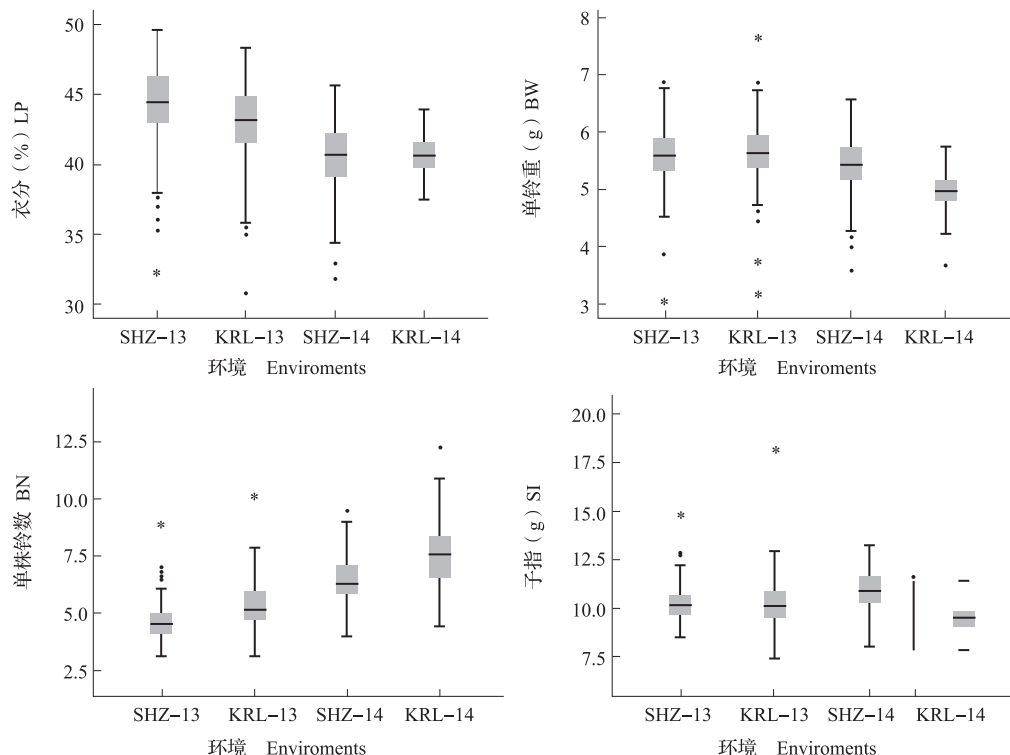
表 2 在 4 个环境下的 4 个产量相关性状的描述统计表

Table 2 Descriptive statistics for yield and its components traits in four different environments

性状 Traits	环境 Environments	平均数 Mean	最小值 Min.	最大值 Max.	标准差 SD	变异系数 (%) CV	偏度 Skewness	峰度 Kurtosis
衣分 (%) LP	SHZ-13	44.1	32.3	49.6	3.1	6.9	-1.2	1.8
	KRL-13	42.7	30.8	48.4	3.0	7.1	-1.1	1.7
	SHZ-14	40.4	31.8	45.6	2.7	6.7	-0.8	0.8
	KRL-14	40.7	37.5	43.9	1.5	3.7	-0.1	-0.3
	平均 Mean	42.0	33.1	46.9	2.6	6.1	-0.8	1.0
单铃重 (g) BW	SHZ-13	5.6	3.1	6.8	0.5	9.5	-1.0	4.3
	KRL-13	5.6	3.2	7.6	0.6	10.2	-0.6	3.9
	SHZ-14	5.4	3.6	6.6	0.5	10.1	-0.4	0.5
	KRL-14	5.0	3.7	5.8	0.3	6.5	-0.4	1.7
	平均 Mean	5.4	3.4	6.7	0.5	9.1	-0.6	2.6
单株铃数 BN	SHZ-13	4.6	3.1	9.0	0.8	18.2	1.8	6.3
	KRL-13	5.3	3.1	10.2	1.1	20.9	1.0	2.4
	SHZ-14	6.5	4.0	9.5	1.2	18.8	0.4	-0.3
	KRL-14	7.6	4.5	12.3	1.4	18.5	0.3	0.3
	平均 Mean	6.0	3.7	10.3	1.1	19.1	0.9	2.2
子指 (g) SI	SHZ-13	10.2	8.5	14.8	0.9	9.0	1.4	4.9
	KRL-13	10.3	7.4	18.1	1.2	12.0	2.2	3.2
	SHZ-14	10.9	8.0	13.2	0.9	8.2	-0.2	0.0
	KRL-14	9.4	7.8	11.6	0.7	7.6	0.3	0.6
	平均 Mean	10.2	7.9	14.4	0.9	9.2	0.9	2.2

SHZ-13、KRL-13、SHZ-14、KRL-14 分别代表 2013 年石河子环境、2013 年库尔勒环境、2014 年石河子环境、2014 年库尔勒环境。下同
SHZ-13, KRL-13, SHZ-14, KRL-14 represent different planting environments (Shihezi, SHZ; Korla, KRL) in 2013 and 2014.

LP: Lint percentage, BW: Boll weight, BN: Bolls per plant, SI: Seed index. The same as below



箱图两端表示其性状的极值范围;点圈表示个别极值;中间直线表示性状中位数

Both ends of boxplot indicate the extreme range of traits, The circles of both ends of boxplot indicate single the maximum and minimum value, The line in the middle of boxplot indicate median

图 1 2013-2014 年在石河子和库尔勒 4 个表型性状的 boxplot 图

Fig. 1 The boxplot of four phenotypic traits in Shihezi and Korla from 2013 to 2014

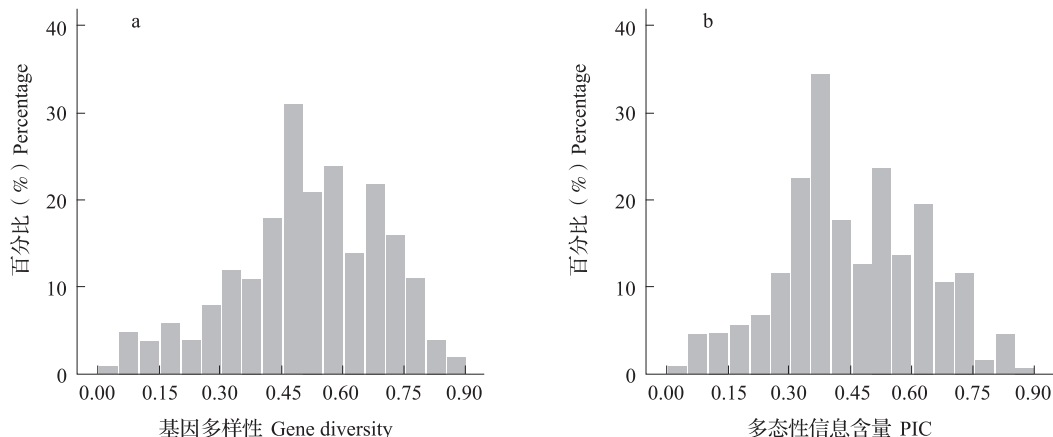


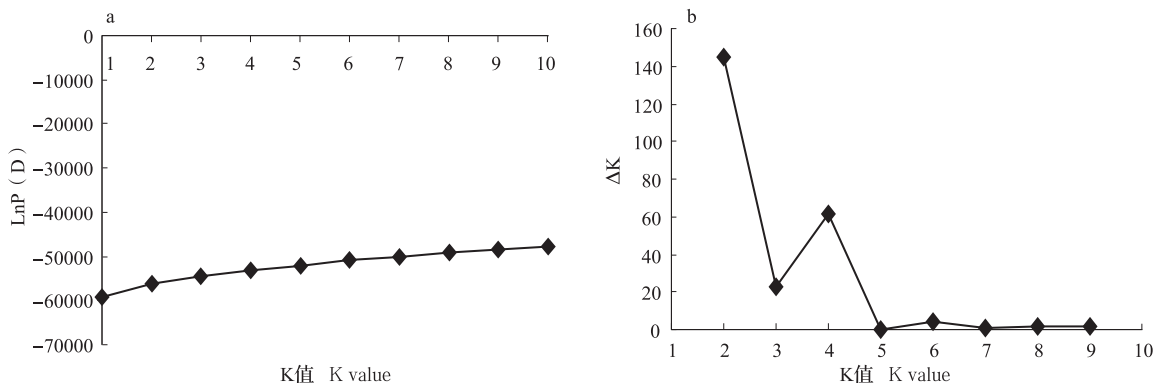
图2 214个标记位点的基因多样性(a)和多态性信息含量(b)频率分布图

Fig. 2 Distribution of gene diversity (a) and PIC (b) of 214 polymorphic marker loci

2.3 群体结构分析

利用STRUCTURE 2.3.1软件进行群体结构分析发现, $\ln P(D)$ 值随假定亚群数 K 值的增大而持续增大(图3a), 无法判断 K 的取值。因此, 参照 G. Evanno 等^[25]的方法通过 ΔK 来确定群体结构的 K 值。在 $K=4$ 时, ΔK 出现峰值拐点(图3b), 由此推测该群体材料被分为4个亚群。进一步以材料在某群体中 Q 值大于或等于0.6划分为一个亚群为原

则^[27], 分析各供试材料对应的 Q 值。结果表明118份材料的4个亚群分别包含34、19、23和12份材料, 各占比例为28.8%、16.1%、19.5%和10.2%; 另有30份材料在任意一个亚群内 Q 值均小于0.6而无法明确其归类, 单独划为一个混合群, 占比例为25.4%。群体结构分析结果显示各类群中材料与地理来源无对应关系。



a: K 值与 $\ln P(D)$ 值的变化图; b: K 值与 ΔK 值的变化图

A: Magnitude of $\ln P(D)$ as a function of K ; b: Magnitude of ΔK as a function of K

图3 118份材料基于群体结构分析的 K 值、 $\ln P(D)$ 值和 ΔK 值变化图

Fig. 3 K , $\ln P(D)$ and ΔK based on population structure analysis for 118 upland cotton cultivars

2.4 各性状与 SSR 标记的关联分析

采用 Tassel 软件中的 MLM ($Q + K$) 方法, 在4种环境中, 对118份材料的4个产量相关性状进行关联分析, 为了降低假阳性的存在, 本研究只对 $-\log_{10} P > 1.3$ ($P < 0.05$) 的显著条件下且能够在2个及以上环境中同时检测到的位点进行分析。表3结果显示共检测到39个SSR标记位点与4个性状相关联。其中检测到14个位点与衣分相关联, 在不同

环境中对表型变异解释率范围为3.29%~9.17%, 平均为5.37%; 检测到8个位点与单铃重相关联, 在不同环境中对表型变异解释率范围为3.68%~9.56%, 平均为5.88%; 检测到5个位点与单株铃数相关联, 在不同环境中对表型变异解释率范围为3.74%~5.77%, 平均为4.44%; 检测到12个位点与子指相关联, 在不同环境中对表型变异解释率范围为3.49%~11.14%, 平均为5.91%。

表 3 表型性状与标记的关联分析结果

Table 3 Association analysis results of phenotypic traits

性状 Traits	标记位点 Marker loci	SHZ-13		KRL-13		SHZ-14		KRL-14	
		$-\log_{10}P$	R^2 (%)	$-\log_{10}P$	R^2 (%)	$-\log_{10}P$	R^2 (%)	$-\log_{10}P$	R^2 (%)
衣分	BNL3089b	1.44	3.75	1.74	4.95	1.47	4.02		
LP	CIR0183a	2.05 **	5.94	1.94	5.70				
	DPL0641b					1.38	5.83	1.34	5.40
	JESPR065b	1.42	3.70	1.57	4.34				
	JESPR153a	1.59	4.28	1.76	5.04				
	NAU0437a			1.41	6.20			1.50	6.37
	NAU0923a	2.56 **	7.86	1.84	5.32			1.30	3.29
	NAU1169c	1.39	3.58	1.58	4.39				
	NAU3031a	1.98	5.70	2.17 **	6.58				
	NAU3308b	2.70 **	8.41	2.83 **	9.17	1.63	4.63		
	NAU3404c	2.66 **	8.23	2.52 **	7.95				
	NAU3881a	2.04 **	5.90	1.91	5.60				
	NAU3995e	1.67	4.58			1.38	3.72		
	NAU5463a	1.39	3.59	1.52	4.17	1.38	3.72		
	单铃重	HAU3101a	1.58	6.39	1.31	5.24	1.41	5.83	
BW	JESPR065a	1.61	4.37	3.01 **	9.56				
	NAU1302	1.41	3.68	2.09 **	6.06				
	NAU2967a	2.50 **	7.68			2.17 **	6.62		
	NAU2984b	2.78 **	8.74	2.14 **	6.23				
	NAU2984c	2.22 **	6.60	1.86	5.24				
	NAU3308b	1.76	4.90	1.64	4.42				
	NAU3522d					1.63	4.59	1.41	3.74
铃数	BNL3875c	1.45	3.76					1.65	4.52
BN	NAU0874d	1.54	4.07	2.08 **	5.77				
	NAU2954b			1.55	3.93	1.47	3.82		
	NAU5262a	1.90	5.35					1.55	4.17
	NAU6094b					1.88	5.23	1.43	3.74
子指	DPL0238a			1.50	4.20	1.46	3.98		
SI	JESPR065a	1.35	3.57	2.51 **	8.06				
	NAU1167d	1.76	5.05	1.32	3.52				
	NAU1169c	1.83	5.30			2.18 **	6.68		
	NAU3201b	1.34	3.54	2.19 **	6.81				
	NAU3800b	1.73	9.03					1.84	9.45
	NAU3995d	1.70	4.84	1.79	5.24	1.42	3.83		
	NAU5262a	1.64	4.60					1.45	3.88
	NAU5335a					1.40	3.77	1.33	3.49
	NAU5335b	2.69 **	8.64	3.25 **	11.14				
	NAU5335c	1.69	4.79	2.66 **	8.68				
	NAU7182b	1.83	7.61	1.89	8.12				

** 表示 $-\log_{10}P > 2 (P < 0.01)$ ** indicate $-\log_{10}P > 2 (P < 0.01)$

本研究进一步与前人研究相比较(表 4)结果发现:有 7 个产量性状关联的标记位点,在前人研究的衣分、子指和铃重等产量性状中也有对应的 QTL 位点。其中,与衣分相对应的标记位点 CIR0183、

JESPR0153、NAU3995 与前人研究结果^[3,8,28]完全一致;另外 4 个 NAU3308、NAU6094、NAU3201、NAU1167 标记位点,在前人研究中^[4-7,29]所对应的性状不完全一致,但也均指向产量性状。

表4 与前人研究的相关联标记位点的比较结果
Table 4 Comparison of the results of associated SSR loci with other research

标记位点 Marker loci	本研究 关联性状 Our research traits	前人研究的关联性状 及连锁的 QTL 位点 Reported research traits and linked with QTL loci
CIR0183	衣分 LP	衣分 LP(qLP-c22-1) ^[8]
JESPR0153	衣分 LP	衣分 LP(qLP-D13) ^[3]
NAU3308	衣分 LP、单铃重 BW	子指 SI(qSI-D2-1) ^[5]
NAU3995	衣分 LP、子指 SI	衣分 LP(qLP-A-1) ^[28]
NAU6094	铃数 BN	衣分 LP(qLP-A5-2) ^[7]
NAU3201	子指 SI	铃重 BW(qBS-D8-1) ^[29] 衣分 LP(qLP-D8-2) ^[29]
NAU1167	子指 SI	衣分 LP(qLP-A3-2) ^[4] 衣分 LP(qLP-2-1) ^[6]

3 讨论

关联分析是以标记位点间的连锁不平衡(LD)为基础的。因此,了解目标群体基因组 LD 状态,可以估计关联分析中需要的标记密度和数量。已有研究表明,棉花由于是常异花授粉作物,基因组位点间存在较高的重组率,因此具有较低的 LD 水平,LD 衰减分析表明,当 LD 系数 $r^2 \geq 0.1$ 和 $r^2 \geq 0.2$ 时,陆地棉品种群体内 LD 平均衰减距离分别为 8~25 cM 和 3~7 cM^[14-15,30]。因此可在陆地棉中利用关联分析方法进行有效的 QTL 定位,但要检测到大多数关联位点,标记间遗传距离需要达到 5 cM 以下。按照四倍体棉种 5200 cM 的遗传距离计算^[31],进行有效的全基因组关联分析,大约需要 1000 个以上多态性的标记。在本研究中,均匀选取的 557 对 SSR 标记仅有 214 对产生多态性,平均位点间遗传距离达到 10 cM 以上,因此,要想达到全基因组水平上的关联分析,本研究还需进一步增加标记密度从而检测到更多与目标性状相关联的标记。

群体结构是影响关联分析的主要因素之一。群体结构的存在会通过影响位点 LD 水平而影响关联分析的准确性,造成关联分析出现假阳性^[32],即检测到的多态性位点和表型性状的关联不是由功能位点造成的,而是由亚群的存在所导致。已有研究表明,同时使用群体结构信息(Q 值)和品种间亲缘关系(K 值)的 MLM(Q + K)模型要优于与单独使用 Q

值或 K 值的 GLM(Q) 和 MLM(K) 模型^[33]。本研究对多态性标记位点基因型数据采用 STRUCTURE 软件进行群体结构的分析和校正,计算每个材料归入各亚群的概率(Q 值),并将其作为协变量进行 MLM(Q + K)的关联分析,有效地矫正了亚群存在导致的伪关联。本研究所选 118 份材料的等位变异频率特征类型数 $K = 4$ 时, ΔK 达到峰值拐点,即可将所选材料分为 4 个亚群。但其所分亚群与所选材料的地理来源无直接关系。其原因可能是因为各生态棉区间棉花品种交流较广泛,造成遗传背景较为复杂,或者来源于同一系谱的材料在不同棉区进行选育,因此不能完全与生态区来源或系谱来源对应。

在本研究中,为最大程度地避免群体结构与亲缘关系对关联分析结果的影响,关联分析结果只对 $-\log_{10} P > 1.3$ ($P < 0.05$) 的显著条件下且能够在 2 个及以上环境中同时检测到的位点进行分析。结果共检测到 39 个 SSR 标记位点与 4 个性状相关联,其中,有多个标记位点同时与多个性状相关联, JESPR065a 同时与单铃重和子指相关联, NAU1169c 同时与衣分和子指相关联, NAU3308b 同时与衣分和单铃重相关联, NAU5262a 同时与单株铃数和子指相关联,说明这些标记可以实现多个目标性状的同步改良;另外,本研究检测到产量相关性状的位点中,发现有 7 个位点与前人通过连锁分析定位得到的产量及相关性状的 QTL 结果一致,其中有 3 个 QTL 与本研究具有相同关联的性状,分别为 CIR0183、JESPR0153、NAU3995,说明这些标记具有较好的可重复性和稳定性;其余 32 个标记位点可能为新开发的标记位点,这些标记可为进一步理解棉花产量性状的遗传机理提供理论依据,并可为后期棉花产量性状的分子标记辅助选择育种提供实践指导。

参考文献

- [1] 中国农业科学院棉花研究所. 中国棉花遗传育种学[M]. 济南:山东科学技术出版社,2003:562
- [2] 殷剑美,武耀廷,张天真,等. 陆地棉产量性状 QTLs 的分子标记及定位[J]. 生物工程学报,2002,18(3):162-166
- [3] Abdurakhmonov I Y, Buriev Z T, Saha S, et al. Microsatellite markers associated with lint percentage trait in cotton, *Gossypium hirsutum*[J]. Euphytica,2007,156:141-156
- [4] Ma X X, Ding Y Z, Zhou B L, et al. QTL mapping in A-genome diploid Asiatic cotton and their congruence analysis with AD-genome tetraploid cotton in genus *Gossypium* [J]. J Genet Genomics,2008,35:751-762
- [5] Qin H D, Guo W Z, Zhang Y M, et al. QTL mapping of yield and fiber traits based on a four-way cross population in *Gossypium hirsutum* L[J]. Theor Appl Genet,2008,117:883-894
- [6] 秦永生,刘任重,梅鸿献,等. 陆地棉产量相关性状的 QTL 定

- 位[J]. 作物学报,2009,35(10):1812-1821
- [7] Liu R Z, Wang B H, Guo W Z, et al. Quantitative trait loci mapping for yield and its components by using two immortalized populations of a heterotic hybrid in *Gossypium hirsutum* L. [J]. Mol Breed, 2012, 29: 297-311
- [8] Yu J W, Yu S X, Gore M, et al. Identification of quantitative trait loci across interspecific F_2 , $F_{2,3}$ and testcross populations for agronomic and fiber traits in tetraploid cotton [J]. Euphytica, 2013, 191: 375-389
- [9] Szalma S J, Buckler I S, Snook M E, et al. Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks [J]. Theor Appl Genet, 2005, 110: 1324-1333
- [10] Narayanan S, Vara Prasad P V, Shroyer K, et al. Characterization of a spring wheat association mapping panel for root traits [J]. Agron J, 2014, 106: 1593-1604
- [11] 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
- [12] Li C Q, Ai N J, Zhu Y J, et al. Association mapping and favorable allele exploration for plant architecture traits in upland cotton (*Gossypium hirsutum* L.) accessions [J]. J Agric Sci, 2016, 154: 567-583
- [13] 郭志军, 赵云雷, 陈伟, 等. 陆地棉 SSR 标记遗传多样性及其与农艺性状的关联分析 [J]. 棉花学报, 2014, 26(5): 420-430
- [14] Mei H X, Zhu X F, Zhang T Z. Favorable QTL alleles for yield and its components identified by association mapping in Chinese upland cotton cultivars [J]. PLoS One, 2013, 8: e82193
- [15] Cai C P, Ye W X, Zhang T Z, et al. Association analysis of fiber quality traits and exploration of elite alleles in upland cotton cultivars/accessions (*Gossypium hirsutum* L.) [J]. J Integr Plant Biol, 2014, 56: 51-62
- [16] 王莉萍, 孙国清, 梁亚军, 等. 棉花纤维品质性状与 SSR 标记的关联分析 [J]. 中国农业科技导报, 2013, 15(4): 110-120
- [17] Saeed M, Guo W Z, Zhang T Z. Association mapping for salinity tolerance in cotton (*Gossypium hirsutum* L.) germplasm from US and diverse regions of China [J]. Aust J Crop Sci, 2014, 8: 338-346
- [18] Zhao Y, Wang H, Chen W, et al. Genetic structure, linkage disequilibrium and association mapping of Verticillium wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm population [J]. PLoS One, 2014, 9: e86308
- [19] 杜雄明, 周忠丽. 棉花种质资源描述规范和数据标准 [M]. 北京: 中国农业出版社, 2005
- [20] Zhao L, Lv Y D, Cai C P, et al. Toward allotetraploid cotton genome assembly: integration of a high-density molecular genetic linkage map with DNA sequence information [J]. BMC Genomics, 2012, 13: 539
- [21] Paterson A H, Brubaker C L, Wendel J F. A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP and PCR analysis [J]. Plant Mol Biol, 1993, 11: 122-127
- [22] 张军, 武耀廷, 郭旺珍, 等. 棉花微卫星标记的 PAGE/银染快速检测 [J]. 棉花学报, 2000, 12(5): 267-269
- [23] 李小胜, 陈珍珠. 如何正确应用 SPSS 软件做主成分分析 [J]. 统计研究, 2010, 27(8): 105-108
- [24] Liu K J, Muse S V. PowerMarker: an integrated analysis environment for genetic marker analysis [J]. Bioinformatics, 2005, 21: 2128-2129
- [25] Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study [J]. Mol Ecol, 2005, 14: 2611-2620
- [26] Bradbury P J, Zhang Z, Kroon D E, et al. TASSEL: software for association mapping of complex traits in diverse samples [J]. Bioinformatics, 2007, 23: 2633-2635
- [27] Yang X H, Gao S B, Xu S T, et al. Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize [J]. Mol Breed, 2011, 28: 511-526
- [28] Wang C, Zhang T Z, Guo W Z. The mutant gene negatively affects many aspects of fiber quality traits and lint percentage in cotton [J]. Crop Sci, 2013, 53: 27-37
- [29] Chen, H, Qian N, Guo W Z, et al. Using three selected overlapping RILs to fine-map the yield component QTL on Chro. D8 in upland cotton [J]. Euphytica, 2010, 176: 321-329
- [30] Qin H D, Chen M, Yi X D, et al. Identification of associated SSR markers for yield component and fiber quality traits based on frame map and upland cotton collections [J]. PLoS One, 2015, 10: e0118073
- [31] Lacape J M, Nguyen T B, Thibivilliers S, et al. A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* × *Gossypium barbadense* backcross population [J]. Genome, 2003, 46: 612-626
- [32] Pritchard J K, Stephens M, Rosenberg N A, et al. Association mapping in structured populations [J]. Am J Hum Genet, 2000, 67: 170-181
- [33] Zhao K, Aranzana M J, Kim S, et al. An *Arabidopsis* example of association mapping in structured samples [J]. PLoS Genet, 2007, 3(1): 71-82