

棉花资源群体结构的推测与纤维品质的关联分析

贺道华,邢宏宜,赵俊兴,李婷婷,汤 益,曾 舟

(西北农林科技大学 农学院,陕西 杨凌 712100)

[摘 要] 【目的】推测棉花资源群体材料的基因组血统与分子亲缘关系,并通过关联分析检测群体中影响表型的位点,为优质纤维品种的选育及其效率的提高提供理论依据。【方法】利用均匀分布于 26 条染色体上的 132 个 SSR 标记,对经挑选的 92 份栽培棉材料的基因组变异进行扫描,利用 STRUCTURE 2.3.1 软件的混合模型聚类分析该群体的遗传结构,采用 NTSYS-pc (Version 1.8) 软件进行 UPGMA 聚类,随后采用 TASSEL 软件的 GLM (General linear model) 程序,对 5 个纤维品质性状 2 年 $\times$ 3 重复的观测值进行标记与目标性状的关联分析。【结果】当指定 7 个亚群($K=7$)时似然值达到最大,该群体分为亚洲棉组(Cluster 5)、海岛棉组(Cluster 7)、陆地棉 1~3 组(Cluster 2、Cluster 4 和 Cluster 6)、混合血统组(Cluster 1 和 Cluster 3)。组群间的 Kullback-Leibler 和 UPGMA 聚类图显示了 7 个组群的遗传组成和材料间的分子亲缘关系。关联分析表明,该群体中累计有 21 个位点与目标性状相关,一些标记同时与 2 个或多个目标性状相关联,可能是目标性状相关乃至“一因多效”的遗传基础。【结论】明确由 92 份材料组成的自然群体的群体结构以及材料间的亲缘关系,并获得了 21 个与棉花纤维品质相关的 SSR 位点,将为这些育种材料的利用、标记辅助选择及下一步候选基因的关联作图提供参考信息。

[关键词] 棉花;SSR;群体结构;纤维品质;关联分析

[中图分类号] S562.024

[文献标识码] A

[文章编号] 1671-9387(2011)01-0103-10

Population structure inferring and association analysis of fiber quality in cultivated cotton

HE Dao-hua, XING Hong-yi, ZHAO Jun-xing, LI Ting-ting,
TANG Yi, ZENG Zhou

(College of Agronomy, Northwest A&F University, Yangling, Shaanxi 712100, China)

Abstract: 【Objective】The major objective was to infer the genome ancestry and the estimation of molecular relation in natural population, and then to detect the loci controlling the interested traits by means of association analysis, so as to lay a theoretic foundation for the effective improvement of fiber quality and the heightening of breeding efficiency. 【Method】One hundred and thirty-two SSR loci, evenly distributed in 26 chromosomes, were assayed for polymorphism among 92 cultivated cotton accessions which represent the high genetic diversity. The identified 494 alleles served as raw data for estimating population structure based on the ADMIXTURE model in the software STRUCTURE 2.3.1. The UPGMA dendrogram was calculated using the software of NTSYS-pc (Version 1.8). Then, the association analysis between SSR loci and 5 fiber quality traits (2 years $\times$ 3 replicate) was performed by using TASSEL GLM (general linear model) program. 【Result】Specifying seven subpopulations ($K=7$) led to a maximum likelihood value. Therefore, this natural population was inferred to be composed of 7 subpopulations, which were *Gossypium*

* [收稿日期] 2010-06-09

[基金项目] 国家自然科学基金项目(30971821);高等学校博士学科点专项科研基金项目(20090204120017);陕西省自然科学基金项目(2010JQ3005)

[作者简介] 贺道华(1975—),男,湖北随州人,讲师,博士,主要从事棉花育种与生物技术研究。E-mail:daohuahe@nwsuaf.edu.cn

[通信作者] 邢宏宜(1956—),男,陕西临潼人,副研究员,主要从事棉花遗传育种研究。E-mail:xinghongyi1169@163.com

arboreum (Cluster 5), *G. barbadense* (Cluster 7), *G. hirsutum* (Cluster 2, Cluster 4 and Cluster 6), and mixture (Cluster 1 and Cluster 3). The Kullback-Leibler distance of pairwise subpopulation was also inferred as the UPGMA dendrogram, which gave a detailed percentage of genetic composition of seven subpopulations and molecular relatedness of each accession. And twenty one SSR loci associated with the traits were screened out from this cultivated population. There were a few loci associated with two or more traits simultaneously, which might be the genetic reason of correlation among traits or pleiotropic phenomena. 【Conclusion】 The population structure of this natural population and the molecular relatedness of each accession were inferred; twenty one SSR loci associated with the traits were detected. These results would provide some referenced information for the utilization of these accessions, for marker assisted selection (MAS), and for the upcoming association studies of some candidate genes.

Key words: cultivated cotton; simple-sequence repeat (SSR); population structure; fiber quality; association analysis

关联分析 (Association analysis), 又称关联作图 (Association mapping), 最先广泛应用于人类的遗传学研究^[1]。该方法是以自然群体为对象, 以自然进化和人工进化过程中长期重组后积累的基因 (位点) 间连锁不平衡 (Linkage disequilibrium, LD) 为基础, 将目标性状表型的多样性与基因 (或标记) 位点的多态性结合起来分析, 从而鉴定出与表型多样性显著相关的基因位点或标记位点^[2]。与传统的 QTL 作图相比, 关联分析具有 3 个特点: ①以种质资源作为研究材料, 不需要专门构建作图群体, 从材料的角度使基础研究与育种工作紧密结合; ②广泛的遗传材料可同时考察任一目标性状大多数基因位点 (或连锁位点) 的各种等位变异, 不受传统 QTL 作图的仅“两等位型”的限制; ③自然群体经历了长期重组后, LD 距离严重衰减, 保证了定位的更高精确性^[3]。

棉花的产量、品质及抗病性等重要性状都是数量性状, 以往常利用“家系连锁作图”进行 QTL 检测^[4-7]。由于“家系连锁作图”的群体材料多态性低, 导致 QTL 作图精度难以进一步提高, 随着棉花分子育种和基因组学研究的深入, 应用关联分析来深入发掘棉花的数量性状基因受到了研究者的关注。目前, 关联分析已在玉米、水稻、小麦等作物 QTL 的剖析中得到了应用^[8-10]。在棉花上, Abdurakhmonov 等^[11-12]利用 95 个 SSR 标记和陆地棉的自然群体, 开展了针对纤维品质性状的关联分析, 随后又增加标记至 202 个, 并扩大群体对纤维品质性状进行了新的关联分析, 获得了一些可供棉花标记辅助选择的关联位点。但这些研究都是在一定群体及特定环境中的研究结果。

本研究从棉花 3 个栽培种 (陆地棉、海岛棉、亚洲棉) 中根据已有的遗传多样性研究结果^[13-18], 选取代表性栽培棉种质组成具有代表性的样本, 对其的纤维品质进行有重复的表型鉴定, 利用在基因组上均匀分布的 132 个 SSR 标记, 获得该样本的标记 (基因型) 数据, 并在分析群体结构的基础上, 进行目标性状的关联分析, 发掘各类目标性状优良位点的关联标记, 以期为棉花优质纤维品种的选育及效率的提高提供理论依据。

1 材料与方法

1.1 供试材料

根据已发表的关于棉花遗传多样性的文献^[13-18], 选用我国 3 大主产棉区不同年代的有代表性的棉花栽培品种, 与部分国外引进品种组成一个自然群体, 从该自然群体中共选出 92 份供试棉花品种 (系), 均由中国农业科学院棉花研究所 (CRI, CAAS) 提供, 其中包括 7 份亚洲棉、74 份陆地棉和 11 份海岛棉 (表 1)。这些材料均在 2008 年播种于西北农林科技大学三原试验站, 随机区组设计, 3 次重复。株距 30 cm, 行距 60 cm, 常规田间管理。2009 年重复 1 次田间试验。

1.2 SSR 标记分析

棉花基因组 DNA 提取采用 CTAB 法^[4]。SSR 反应体系与程序参照 He 等^[4]的方法。采用垂直板变性聚丙烯酰胺凝胶电泳法对 SSR 扩增产物进行分离。对 SSR 扩增产物的电泳结果采用“0~1”系统记录谱带位置, 观察扩增标记 (条带) 的有无, 有条带标记为“1”, 无条带标记为“0”。同时参照 DNA Marker 推测并记录条带的大小 (bp)。

表 1 92 份供试棉花品种(系)的名称、代号和系谱

Table 1 No. and pedigree of 92 cultivars studied in this research

序号 No.	材料 Cultivar	系谱来源 Pedigree source
1	农林 6 NL 6	引自日本 From Japan
2	安徽青阳大子花毛子 AHQYDZHMZ	未知 Unknown
3	长顺代化小花 CSDHXX	未知 Unknown
4	安徽青阳大子花光子 AHQYDZHGZ	未知 Unknown
5	完县小白花 WXXBH	未知 Unknown
6	下寨小花 3 XZXH 3	未知 Unknown
7	长绒中棉 CRZM	未知 Unknown
8	新海 8 号 Xinhai 8	军海 1 Junhai 1
9	新海 12 号 Xinhai 12	军海 1 Junhai 1
10	PimaS-3	引自美国,不详 From American,unknown
11	Pima3-79	遗传标准系 The genetic standard line
12	埃棉 3 号 Aimian 3	未知 Unknown
13	5476 依 5476-И	引自前苏联,不详 From The former U. S. S. R. ,unknown
14	海 7124 Hai 7124	Mcixxill
15	Pima 90-53	引自美国,不详 From American,unknown
16	军海 1 号 Junhai 1	9122-И
17	新海 3 号 Xinhai 3	军海 1 Junhai 1
18	吉扎 80 Giza 80	引自埃及,不详 From Egypt,unknown
19	陕 1155 Shaan 1155	中长 1×徐州 1818 Zhongchang 1×Xuzhou 1818
20	冀棉 20 号 Jimian 20	冀棉 10×(海陆野)F Jimian 10×(barbadense×hirsumtum×wilt)F
21	辽棉 9 号 Liaomian 9	辽 661×68-3 Liao 661×68-3
22	洞庭 1 号 Dongting 1	岱字 15 Deltapine 15
23	86-1	陕棉 5 Shaanmian 5
24	T586	Multiple-markers lines
25	TM-1	岱字 14 Deltapine 14
26	57-681	岱字 15 Deltapine 15
27	52-128	德字棉 531 Delfos 531
28	PD 9364	引自美国,不详 From American,unknown
29	Acala 3080	引自美国,不详 From American,unknown
30	鲁棉 6 Lumian 6	72 系 Line 72
31	鄂沙 28 Esha 28	(荆棉 4×冈棉 1)×荆棉 4 (Jingmian 4×Gangmian 1)×Jingmian 4
32	鲁 1024 Lu 1024	山东 129×陕 5245 Shandong 129×Shaan 5245
33	中 521 Zhong 521	(徐州 209×海 910 依)×陕棉 4 (Xuzhou 209×Hai 910И)×Shaanmian 4
34	泗棉 3 号 Simian 3	新洋 76-75×泗 791 Xinyang 76-75×Si 791
35	冀棉 6 号 Jimian 6	徐州 1818×(蜀葵+徐州 1818) Xuzhou 1818×(Shukui+Xuzhou 1818)
36	川 243 Chuan 243	川 737×岱字 15 Chuan 737×Deltapine 15
37	冀棉 12 Jimian 12	冀棉 5×冀棉 1 Jimian 5×Jimian 1
38	冀棉 8 号 Jimian 8	冀棉 5×(岱字 45×紫花)辐射 Jimian 5×(Deltapine 45×Zihua) Coradiation
39	冀棉 14 Jimian 14	冀 75-7×冀 75-23 Ji 75-7×Ji 75-23
40	中 8004 Zhong 8004	(徐州 209×岱字 15)×(岱字 14×徐 209) (Xuzhou 209×Deltapine 15)×(Deltapine 14×Xuzhou 209)
41	苏棉 6 号 Sumian 6	冀合 24-1×AC 241 Jihe 24-1×AC 241
42	鲁棉 1 号 Lumian 1	中棉所 2×山东 1195,再辐射 CRI. CAAS 2×Shandong 1195,radiation
43	苏棉 2 号 Sumian 2	(中棉所 7×岱字 16)×(徐州 142. 邢台 68-71) (CRI. CAAS 7×Deltapine 16)×(Xuzhou 142. Xingtai68-71)
44	晋棉 7 号 Jinmian 7	陕棉 9×(中长 1+晋中 200) Shaanmian 9×(Zhongchang 1+Jinzhong 200)
45	鲁无 401 Luwu 401	中无 383×鲁 384 Zhongwu 383×Lu 384
46	军棉 1 号 Junmian 1	C 1470×多父本 C 1470×Many males
47	秦远 4 号 Qinyuan 4	中棉所 12×(86-1,等) CRI. CAAS 12×(86-1,et al)
48	斯字棉 4 Stoneville 4	斯字棉 1 号 Stoneville 1
49	苏棉 11 号 Sumian 11	湘棉 10×CAdcs-2-81 Xiangmian 10×CAdcs-2-81
50	苏棉 12 号 Sumian 12	(8004×冀 328)×江苏 9101 (8004×Ji 328)×Jiangsu 9101
51	豫棉 21 号 Yumian 21	中棉所 12×豫植 117 CRI. CAAS 12×Yuzhi 117
52	豫棉 10 号 Yumian 10	(商丘 40×86-1)×冀 366 (Shangqiu 40×86-1)×Ji 366

续表 1 Continued table 1

序号 No.	材料 Cultivar	系谱来源 Pedigree source
53	晋棉 18 号 Jinmian 18	晋棉 6 Jinmian 6
54	晋棉 19 号 Jinmian 19	82-87×石 711 82-87×Shi 711
55	晋棉 12 号 Jinmian 12	冀棉 11×中 6331 Jimian 11×Zhong 6331
56	陕棉 11 Shaanmian 11	陕 3619 Shaan 3619
57	鄂抗棉 7 Ekangmian 7	鄂 7701×荆 1418 E7701×Jing 1418
58	陕棉 6 号 Shaanmian 6	(岱福棉×63-105)×中 3 (Daifumian×63-105)×Zhong 3
59	陕棉 10 Shaanmian 10	陕 3619×(3719+陕棉 4) Shaan 3619×(3719+Shaanmian 4)
60	陕棉 4 号 Shaanmian 4	中棉所 3×(57-681+辽 2) CRI. CAAS 3×(57-681+Liao 2)
61	石远 321 Shiyuan 321	(86-1×F3)×中 12 (86-1×F3)×Zhong 12
62	晋棉 26 号 Jinmian 26	晋棉 7(转 Bt 基因) Transgenic Bt from Jinmian 7
63	陕棉 8 号 Shaanmian 8	52-128×高密 1195 52-128×Gaomi 1195
64	陕棉 2365 Shaanmian 2365	未知 Unknown
65	陕棉 5 号 Shaanmian 5	7538×陕棉 3 7538×Shaanmian 3
66	陕棉 9 号 Shaanmian 9	陕棉 3×52-128 Shaanmian 3×52-128
67	徐州 142 Xuzhou 142	徐州 58 Xuzhou 58
68	珂字棉 201 Coker 201	未知 Unknown
69	新陆中 3 Xinluzhong 3	{[(108 夫×C1470)×108 夫]×137 夫}×陕 401 {(108F×C 1470)×108F}×137F}×Shaan 401
70	新陆中 5 Xinluzhong 5	陕棉 11×108 夫 Shaanmian 11×108F
71	新陆早 6 号 Xinluzao 6	85-174×贝尔斯洛 85-174×Beilslow
72	辽棉 13 号 Liaomian 13	辽 7440×6051 Liao 7440×6051
73	德字棉 531 Delfos 531	福字棉 Foster
74	苏远 7235 Suyuan 7235	未知 Unknown
75	岱字 15 Deltapine 15	岱字 14 Deltapine 14
76	塔什干 1 号 Tashkent-1	C4727×墨西哥半野生棉 C-4727× <i>G. hirsutum</i> ssp. <i>Mexicanum</i>
77	湘棉 10 号 Xiangmian 10	岱红岱×陕棉 4 号 Daihongdai×Shaanmian 4
78	斯字棉 2B Stoneville 2B	斯字棉 2A Stoneville 2A
79	徐州 1818 Xuzhou 1818	徐州 209 Xuzhou 209
80	中棉所 12 CRI. CAAS 12	乌干达 4×邢台 68-71 Uganda 4×Xingtai 68-71
81	中棉所 36 CRI. CAAS 36	H109×662
82	邢台 68-71 Xingtai 68-71	徐州 1818 Xuzhou 1818
83	中棉所 17 CRI. CAAS 17	(中 7259×中 6651)×中 10 (Zhong 7259×Zhong 6651)×Zhong 10
84	中棉所 16 CRI. CAAS 16	中 211×辽 4086 Zhong 211×Liao 4086
85	中棉所 23 CRI. CAAS 23	{[(5658×陕 5245)×4067]×中 10}×冀棉 8 {(5658×Shaan 5245)×4067}×Zhong 10}×Jimian 8
86	中棉所 21 CRI. CAAS 21	斯字棉 213×(中 758×849) Stoneville 213×(Zhong 758×849)
87	中棉所 30 CRI. CAAS 30	中棉所 16×Bt 基因种质系 Transgenic Bt from CRI. CAAS 16
88	中棉所 9 号 CRI. CAAS 9	(中棉所 4×乌干达 3)×陕 3778 (CRI. CAAS 4×Uganda 3)×Shaan 3778
89	中棉所 19 CRI. CAAS 19	中棉所 17×(中 7263×中 6429) CRI. CAAS 17×(Zhong 7263×Zhong 6429)
90	中棉所 3 号 CRI. CAAS 3	岱字 15 Deltapine 15
91	中棉所 35 CRI. CAAS 35	中 23021×(中 12×川 174) Zhong 23021×(Zhong 12×Chuan 174)
92	运城 1729 Yuncheng 1729	[(辽棉 7×贝无)×(中 3×斯字棉 7)]×86-1[(Liaomian 7×Beiwu)×(Zhong 3×Stoneville7)]×86-1

1.3 棉花群体遗传结构的分析

为了估测该棉花自然群体的遗传结构,应用 STRUCTURE 2. 3. 1 软件^[19],进行基于数学模型的类群划分,并计算材料相应的 Q 值(即第 *i* 材料基因组变异源于第 *k* 群体的概率)。

先设定群体数(*K*)为 2~10,并假定位点均是独立的,本研究对 132 个 SSR 位点的分型数据进行分析,将 MCMC (Markov chain monte carlo)开始时的不作数迭代 (Length of burn-in period) 设为 10 000 次,再将不作数迭代后的 MCMC 设为 100 000 次,然后依据似然值最大的原则选取一个合适的 *K*

值,并绘出基于模型的群体遗传结构图。利用 NT-SYS-pc (Version 1. 8) 软件^[20],采用离差平方和进行系统聚类分析。

1.4 棉花纤维品质性状调查

于 09-10—09-20 对每个棉花品种取同一节位的纤维送至农业部棉花纤维品质监督检验测试中心进行检测,检测温度 20 ℃,相对湿度 65% (HVI900)。纤维品质性状包括上半部平均长度 (UHML)、整齐度指数 (UI)、马克隆值 (MIC)、断裂比强度 (STR) 和伸长率 (ELG)。

1.5 关联分析

使用 TASSEL 软件的 GLM (General linear model)程序^[21],以各个体的 Q 值作为协变量进行群体矫正,将 5 个性状 2 年平均的表型数据分别与标记位点的等位变异进行回归分析。

2 结果与分析

132 个 SSR 标记在 92 个棉花品种(系)中共扩增出 494 条多态性条带,每个位点平均扩增 3.69 条多态性带,各引物的平均多态信息量 (PIC) 为 0.01~0.871 4,平均为 0.546 6。

2.1 棉花资源的群体遗传结构

基于连锁位点的 SSR 分型数据,利用 STRUC-TURE 2.3.1 软件中的混合模型聚类法来推断群体结构。根据似然值的变化,群体数 $K=7$ 时,似然值最大 $[\lg Pr(X|K)=-9\ 805.2]$,当 $K=8$ 和 $K=9$ 时,似然值分别为 $-10\ 155$, $-18\ 691.8$,可见 $K>7$ 时似然值显著下降。因此,推断此群体的遗传结构由 7 类血缘组成(图 1),分别以红色、鲜绿、蓝色、黄色、粉红、青绿、金色 7 种颜色表示,表明此群体种质来源具有明显的异质性和多元特性。

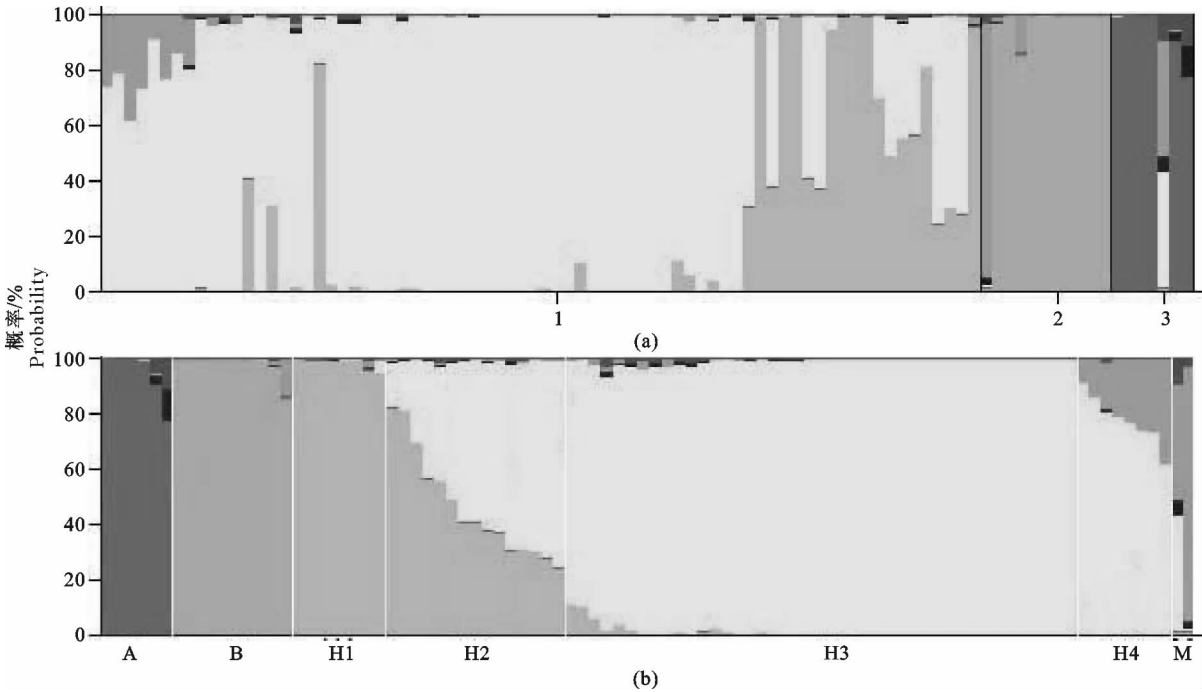


图 1 基于 STRUCTURE 2.3.1 软件中混合模型聚类法推断的供试棉花品种的群体遗传结构

(a)按材料所属的种归类,1,2,3 分别表示陆地棉、海岛棉、亚洲棉;(b)根据 Q 值来排列;

图中不同颜色表示不同的组群;每条竖线代表一份种质,不同颜色所占比例越大,则该种质被划分到相应组群的可能性越大;

A. 亚洲棉组群;B. 海岛棉组群;H1、H2、H3、H4. 分别表示血统不同的 4 个陆地棉组群;M. 种间混合血统的组群

Fig. 1 Bar plot of the genetic composition of individual lines,based on 132 SSR linked multilocus data by using admixture model in STRUCTURE 2.3.1

(a). Group by pre-defined POP Id;1,2 and 3 denote *G. arboretum*,*G. barbadense* and *G. hirsutum*, respectively;

(b). Sort by Q. Each column (92 columns in total) represents the genotype of an accession,and is partitioned into K colored segments,and the length of a segment represents the estimated genetic fraction in every line of each of the 7 inferred subpopulations;A. Cluster *G. arboretum*; B. Cluster *G. barbadense*;

H1,H2,H3 and H4. Four clusters for 4 ancestries in *G. hirsutum*;M. Cluster for mixed ancestries between species

图 1(b)和表 2 显示,有 10 个海岛棉归属 cluster7 血统(图 1(b)中 B),但有 1 个海岛棉即吉扎 80 比较特殊(图 1(b)中 M),这是来源于埃及的一个品种,其血统来源与其他 10 个海岛棉材料存在很大差异。cluster5 代表着亚洲棉血统(图 1(b)中 A),但也存在一个较特殊的亚洲棉材料即长绒中棉(图 1(b)中 M),这可能是由于长绒中棉更接近四倍体栽

培棉的绒长,故其遗传基础(血统)与普通亚洲棉不同,但从图 1(b)可以看出,海岛棉吉扎 80 与亚洲棉长绒中棉在血统上有 41.8%的近似性,且 Q 值矩阵(数据略)显示,吉扎 80 与长绒中棉基因组变异源于 cluster2 组群的概率分别为 0.91 和 0.418。陆地棉材料的血统比较复杂,不但存在 3 个较大的血统(cluster2、cluster4 和 cluster6),而且存在 2 种在此

群体中比例较低的血统,即 cluster1 和 cluster3。根据血缘组成,陆地棉可分为 4 个组群:①“鲜绿 cluster2+黄色 cluster4”组群(图 1(b)中 H4),共包含 8 个材料,即晋棉 7 号、中棉所 35、陕 1155、鲁棉 6、冀棉 20 号、T586、斯字棉 4 号和 Acala3080,这些材料多具有斯字棉血缘,如晋棉 7 号、中棉所 35、陕 1155、鲁棉 6、冀棉 20 号均含有经徐州 209 而来的斯字棉血缘;②“黄色 cluster4”组群(图 1(b)中

H3),共包含 39 个材料,这些材料多具有岱字棉、斯字棉血缘;③“黄色 cluster4+青绿 cluster6”组群(图 1(b)中 H2),共包含 19 个材料,这些材料多具有斯字棉、德字棉血缘;④“青绿 cluster6”组群(图 1(b)中 H1),共包含 8 个材料,即中棉所 16、辽棉 13 号、豫棉 10 号、徐州 142、中棉所 17、中棉所 21、豫棉 21 号和中棉所 19,这些材料多具有德字棉、脱字棉血缘。

表 2 棉花先验群体(3 个种)归属于各个组群的比例

Table 2 Proportion of membership of each pre-defined population (three species) in each of the 7 clusters

先验群体(物种) Given pop (species)	推断的组群 Inferred cluster							材料数 Number of accessions
	组群 1 Cluster 1	组群 2 Cluster 2	组群 3 Cluster 3	组群 4 Cluster 4	组群 5 Cluster 5	组群 6 Cluster 6	组群 7 Cluster 7	
陆地棉 <i>G. hirsutum</i>	0.004	0.027	0.003	0.757	0.002	0.205	0.001	74
海岛棉 <i>G. barbadense</i>	0.005	0.096	0.005	0.001	0.002	0.002	0.889	11
亚洲棉 <i>G. arboreum</i>	0.038	0.061	0.030	0.061	0.807	0.003	0.001	7

2.2 棉花组群间的 Kullback-Leibler 距离

组群间的 Kullback-Leibler 距离可以衡量 2 个组群间的血统(遗传)差异。表 3 显示,7 个组群两

两间的 Kullback-Leibler 距离为 0.000 1~1.493 3, 21 个两两组群间的平均距离为 0.679 6。该结果显示,将此群体分为 7 个组群是最佳的聚类结果。

表 3 棉花组群(Cluster)间的 Kullback-Leibler 距离

Table 3 Matrix of the Kullback-Leibler distance of pairwise subpopulations

组群 Cluster	组群 1 Cluster 1	组群 2 Cluster 2	组群 3 Cluster 3	组群 4 Cluster 4	组群 5 Cluster 5	组群 6 Cluster 6
组群 2 Cluster 2	0.314 1					
组群 3 Cluster 3	0.000 1	0.314 1				
组群 4 Cluster 4	0.495 8	0.800 3	0.497 9			
组群 5 Cluster 5	0.565 9	0.824 6	0.566 4	1.422 0		
组群 6 Cluster 6	0.215 4	0.555 9	0.216 1	0.602 3	0.823 1	
组群 7 Cluster 7	0.599 9	1.087 3	0.602 0	1.493 3	1.264 3	1.011 6

2.3 棉花品种间的聚类分析

基于连锁位点的 SSR 分型数据,利用 NTSYS-pc (Version 1.8) 软件^[20],采用离差平方和方法进行系统聚类分析,获得的 UPGMA 树状聚类图如图 2 所示。由图 2 可以看出,在遗传相似系数为 0.773 6~0.777 7 时,可将 92 个棉花品种划分成 6 个大类:Ⅰ类包含 6 个品种,全为亚洲棉;Ⅱ类包含 10 个品种,全部为海岛棉;Ⅲ类包含 54 个品种,其中 1 个为亚洲棉(长绒中棉),其余 53 个均为陆地棉;Ⅳ类包含 19 个陆地棉品种;Ⅴ类包含 2 个陆地棉品种(中棉所 19 和辽棉 1);海岛棉吉扎 80 独自归为一类(Ⅵ类)。在相似系数为 0.825 0~0.828 6 时,Ⅲ类又可分为 6 个亚类,其中Ⅲ-1 亚类为长绒中棉;Ⅲ-2 亚类含陕 1155、鲁棉 6、冀棉 20、中棉所 35、Acala3080、斯字棉 4 号和 T586 共 7 个材料;Ⅲ-3 亚类为晋棉 7;Ⅲ-4 亚类含 52-128、川 243 和冀棉 12 共 3 个材料;Ⅲ-5 亚类含 41 个材料;Ⅲ-6 亚类为德字棉 531。聚类结果显示,有 44.57% 的棉花材料,即Ⅲ-5 中的 41 个材料的相似系数为 0.871 9±

0.035 6,表明陆地棉的种内遗传多样性偏低。

2.4 与棉花纤维品质性状相关联的 SSR 标记分析

上述研究表明,本研究中的 92 份棉花自然群体由 7 个组群组成,现将各个体相应的 Q 值作为协变量,分别进行 5 个纤维品质性状的表型变异对标记变异的回归分析,探求与纤维品质性状 QTL 相关联的标记及其效应的大小(表型变异解释率)。结果表明,本研究检测的 132 个 SSR 位点中,分布于 15 个染色体上的 21 个位点的变异与 5 个纤维品质性状变异相关,其中 Chr10 上有 4 个位点与纤维品质性状显著相关,Chr5、Chr8 和 Chr20 上均有 2 个位点与纤维品质性状显著相关,其余 11 个染色体上各有 1 个位点与纤维品质性状相关。另外,Chr10 上有 6 个达到显著相关的“位点-性状关联”,Chr8、Chr20 上分别有 5 和 4 个显著的“位点-性状关联”。

表 4 列出了所有关联标记及其对相应性状及表型变异的解释率。由表 4 可以看出,与棉花上半部平均长度相关的位点有 6 个,分别位于染色体箱 (Chromosomal Bin)^[22] Chr04_02、Chr05_05、Chr08

_04、Chr10_01、Chr10_05 和 Chr20_05 中,表型变异解释率为3.75%~8.66%;与整齐度指数相关的位点有 9 个,分别位于染色体箱 Chr05_02、Chr06_01、Chr08_04、Chr08_05、Chr10_04、Chr16_11、Chr17_06、Chr20_05 和 Chr24_01 中,表型变异解释率为 5.75%~13.51%;与马克隆值相关的位点有 6 个,分别位于染色体箱 Chr03_01、Chr08_04、Chr10_01、

Chr12_01、Chr20_01 和 Chr25_04 中,表型变异解释率为 3.26%~7.20%;与伸长率相关的位点有 6 个,分别位于染色体箱 Chr03_01、Chr06_01、Chr08_04、Chr10_01、Chr18_06 和 Chr20_05 中,表型变异解释率为 7.28%~18.17%;与断裂比强度相关的位点有 3 个,分别位于染色体箱 Chr01_08、Chr10_08 和 Chr19_09 中,表型变异解释率为 3.27%~3.72%。



图 2 根据 Jaccard's 相似系数所得的 92 个棉花材料的 UPGMA 树状聚类图
Fig.2 UPGMA dendrogram based on Jaccard's similarity coefficients of the 92 cotton accessions

表 4 显示,上述 30 个位点中存在 5 个“一因多效”的 SSR 位点,如 BNL3255 同时对纤维上半部平均长度、整齐度指数、马克隆值和伸长率 4 个性状有显著效应;CER0077 同时对整齐度指数和伸长率 2 个性状有显著效应;CGR5349 同时对纤维上半部平均长度、马克隆值和伸长率 3 个性状有显著效应;

表 4 与棉花纤维品质性状极显著相关($P<0.01$)的标记位点及其染色体箱和表型变异的解释率

Table 4 Marker loci associated with traits and explanation of their chromosomal bin and phenotypic variation

品质性状 Fiber quality	SSR 位点 SSR site	染色体箱 Bin No.	F 值 F value	P 值 P value	表型变异 解释率/%
上半部平均长度 Upper-half mean length	CIR253	Chr05_05	15.662 0	0.001 4	8.22
	BNL3255	Chr08_04	29.355 0	6.49×10^{-7}	8.66
	CER0063	Chr10_05	10.424 0	0.002 00	3.75
	CER0164	Chr04_02	14.474 0	0.000 34	5.31
	CGR5349	Chr10_01	13.787 0	0.000 45	4.79
	CGR5740	Chr20_05	16.043 0	0.000 14	5.43
整齐度指数 Uniformity index	BNL2597	Chr24_01	11.493 0	0.003 10	10.83
	CIR244	Chr08_05	18.690 0	0.000 83	12.28
	BNL2732	Chr05_02	22.694 0	0.000 37	13.51
	BNL3255	Chr08_04	10.097 0	0.002 10	5.75
	BNL3895	Chr10_04	8.759 7	0.004 40	6.45
	CER0077	Chr06_01	8.535 7	0.004 90	6.08
	CGR5181	Chr16_11	8.650 7	0.004 60	6.25
	CGR5453	Chr17_06	8.748 0	0.004 40	6.21
	CGR5740	Chr20_05	9.366 8	0.003 00	5.96
马克隆值 Micronaire value	BNL2553	Chr20_01	10.847 5	0.003 50	7.20
	BNL4059	Chr12_01	19.551 6	2.98×10^{-5}	6.36
	BNL1440	Chr25_04	14.358 1	0.000 31	5.37
	BNL3255	Chr08_04	10.836 3	0.001 50	4.06
	CGR5349	Chr10_01	15.303 1	0.000 23	6.19
	CGR5620	Chr03_01	8.960 6	0.003 60	3.26
伸长率 Elongation	BNL2544	Chr18_06	25.623 6	3.97×10^{-6}	15.55
	BNL3255	Chr08_04	41.681 4	8.37×10^{-9}	18.17
	CER0077	Chr06_01	11.076 3	0.001 50	7.86
	CGR5349	Chr10_01	10.014 9	0.002 40	7.39
	CGR5740	Chr20_05	21.393 5	1.42×10^{-5}	13.01
	CGR5620	Chr03_01	12.780 6	0.000 59	7.28
断裂比强度 Strength	BNL4096	Chr19_09	8.891 3	0.004 1	3.27
	CGR5417	Chr01_08	7.574 9	0.009 0	3.42
	CGR5565a	Chr10_08	9.999 1	0.002 4	3.72

3 讨 论

由于进化(驯化)历史及交配体系的不同,经育种专家长期的杂交、选育,植物自然群体积聚了大量的遗传重组,为高分辨率地剖析数量性状基因提供了遗传学前提。同时,复杂的育种历史产生了复杂的群体结构,对于常异花授粉的棉花,其群体结构的复杂性介于自花授粉植物(如水稻、小麦、大豆)与异花授粉植物(如玉米)之间。亚洲棉、陆地棉和海岛棉各自独立的起源及不同生态区陆地棉的复杂育种史,均使栽培棉群体内形成了特定的群体结构。

群体结构分析是开展关联作图的基础。由于群

CGR5620 同时对马克隆值和伸长率 2 个性状有显著效应;CGR5740 同时对纤维上半部平均长度、整齐度指数和伸长率 3 个性状有显著效应。这些位点的“多效性”提示:①可能存在多效性基因,该基因同时影响多个性状;②影响这些性状的部分基因可能紧密连锁,导致标记位点变异的多效性。

体结构会通过影响位点连锁不平衡(LD)而影响关联分析的准确性,本研究根据已公开发表的遗传图谱^[4,22],从每个连锁群上不同的标记框^[15]中选择 SSR 标记,结果从 26 个连锁群上共选择出 132 个 SSR 标记(分布于不同的染色体框),利用其对群体进行了标记分析。对多位点基因型数据采用基于模型的聚类来分析群体结构,并计算出各个体归入各亚群的概率(Q 值),再将各个体 Q 值作为协变量纳入 TASSEL^[21]的回归分析,从而规避因亚群混合而造成的伪关联^[3]。本研究的基因组血统与分子亲缘数据,显示了各组群、各材料间的亲缘关系,将对各供试材料在育种工作中的应用有指导作用。对于

没有系谱档案或具有混合血统的材料,基于分子标记的 STRUCTURE^[19] 分析将使之归于特定的族群。

本研究基于少量分子标记(132 个位点),进行了粗略意义上的全基因组关联分析,目的在于为即将开展的候选基因关联作图奠定基础。严格意义上的全基因组关联分析需使用高密度标记对全基因组进行扫描,特别是对于棉花这样的常异交植物,由于其 LD 衰减的距离较短(相对于自花授粉作物拟南芥、水稻、小麦等而言),所需的标记数量很多,因此要开展棉花全基因组关联分析,工作量及试验耗费将十分庞大。

目前常用的 QTL 定位方法较多,本研究采用 TASSEL 软件的 GLM 程序,检测表型变异与 SSR 位点等位变异间关联的显著性,由于不依赖于图谱进行 QTL 分析,比较简单直观,且对 QTL 的检测能力较高,直接表现出 30 个位点与 5 个性状相关联。与其他 QTL 作图结果^[4-7,23-24] 相比,本研究检出的位点既有一致性,也有互补性。

本研究对栽培种群体进行关联分析,与文献[11]无相同的关联位点,与文献[12]只有 1 个关联位点相同,而大多数关联位点不同,说明不同群体遗传基础有相当大的差异,应针对不同目标性状从不同材料中发掘有利基因。Abdurakhmonov 等^[12] 发现,BNL3255 与马克隆值显著关联。而本研究结果显示,BNL3255 除与马克隆值显著关联外,还与纤维上半部平均长度、整齐度指数和伸长率 3 个性状关联。He 等^[4] 在 BNL2544 和 BNL1440 附近分别检测出了 *qNSPBchr18* 和 *qNSPBA02*(控制单铃种子数),本研究发现 BNL2544 的变异对伸长率也有一定影响,而 BNL1440 的变异则对马克隆值有一定影响。Ma 等^[23] 在 BNL4059 附近检测出了影响纤维强度的 QTL(*qFS-12-1*) 和伸长率的 QTL(*qFE-12-1*)。An 等^[24] 发现,BNL3255 与纤维强度有显著的关联,而本研究发现 BNL4059 和 BNL3255 分别与马克隆值和上半部平均长度显著关联。本研究所用标记与 Abdurakhmonov 等^[11-12] 的部分重合,密度偏稀,加密后相同位点数可能会有所增多。所用群体不同,因其遗传基础存在差异,因而在不同类性状上关联位点的检出率也存在差异。有可能是因为使用标记有限,故未检出共同位点。此外,本研究还发现,某些单一标记同时与 2 个或多个性状相关联,该结果可用于解释数量性状之间可能存在的遗传相关,即性状的相互关联可能是控制

该性状的 QTL 相互连锁或某 QTL 的“一因多效”所致。

[参考文献]

- [1] March R E. Gene mapping by linkage and association analysis [J]. Mol Biotechnol,1999,13:113-122.
- [2] Mackay I, Powell W. Methods for linkage disequilibrium mapping in crops [J]. Trends Plant Sci,2007,12:57-63.
- [3] 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.
Wen Z X, Zhao T J, Zheng Y Z, et al. Association analysis of agronomic and quality traits with SSR markers in *Glycine max* and *Glycine soja* in China: I. Population structure and associated markers [J]. Acta Agronomica Sinica,2008,34(7):1169-1178. (in Chinese)
- [4] He D H, Lin Z X, Zhang X L, et al. QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of *Gossypium hirsutum* × *G. barbadense* [J]. Euphytica,2007,153:181-197.
- [5] 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.
- [6] Lacape J M, Nguyen T B, Courtois B, et al. QTL analysis of cotton fiber quality using multiple *Gossypium hirsutum* × *Gossypium barbadense* backcross generations [J]. Crop Sci,2005,45:123-140.
- [7] Wang B, Guo W, Zhu X, et al. QTL mapping of fiber quality in an elite hybrid derived-RIL population of Upland cotton [J]. Euphytica,2006,152:367-378.
- [8] Flint-Garcia S A, Thuillet A, Yu J, et al. Maize association population: A high resolution platform for QTL dissection [J]. Plant J,2005,44:1054-1064.
- [9] Eizenga G C, Agrama H A, Lee F N, et al. Identifying novel resistance genes in newly introduced blast resistant rice germplasm [J]. Crop Sci,2006,46:1870-1878.
- [10] Maccaferri M, Sanguineti M C, Enrico N, et al. Population structure and long-range linkage disequilibrium in a durum wheat elite collection [J]. Mol Breed,2005,15:271-289.
- [11] Abdurakhmonov I Y, Kohel R J, Yu J Z, et al. Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm [J]. Genomics,2008,92(6):478-487.
- [12] Abdurakhmonov I Y, Saha S, Jenkins J N, et al. Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm [J]. Genetica,2009,136(3):401-417.
- [13] 徐秋华,张献龙,聂以春,等. 我国棉花抗枯萎病品种的遗传多样性分析 [J]. 中国农业科学,2002,35(3):272-276.
Xu Q H, Zhang X L, Nie Y C, et al. Genetic diversity evaluation of cultivars (*G. hirsutum* L.) resistant to *Fusarium* wilt by RAPD markers [J]. Scientia Agricultural Sinica,

2002,35(3):272-276. (in Chinese)

[14] 刘文欣,孔繁玲,郭志丽,等. 建国以来我国棉花品种遗传基础的分子标记分析 [J]. 遗传学报,2003,30(6):560-570.
Liu W X,Kong F L,Guo Z L,et al. An analysis about genetic basis of cotton cultivars in China since 1949 with molecular markers [J]. Acta Genetica Sinica,2003,30(6):560-570. (in Chinese)

[15] 温小杰,马峙英,王省芬,等. 中国抗枯、黄萎病陆地棉材料分子水平的遗传差异评价 [J]. 中国农业科学,2005,38(5):936-943.
Wen X J,Ma Z Y,Wang X F,et al. Molecular divergence estimation of Chinese Upland cottons with *Fusarium* and *Verticillium* wilt resistance [J]. Scientia Agricultura Sinica,2005,38(5):936-943. (in Chinese)

[16] 宋宪亮,孙学振,张天真,等. 棉花遗传多态性研究进展 [J]. 西北植物学报,2004,24(12):2393-2397.
Song X L,Sun X Z,Zhang T Z,et al. Advances on genetic diversity of cotton (*Gossypium*) [J]. Acta Botanica Boreali-Occidentalia Sinica,2004,24(12):2393-2397. (in Chinese)

[17] 王省芬,张桂寅,李喜焕,等. 黄河、长江流域棉区棉花抗病品种的 AFLP 分析 [J]. 遗传学报,2004,31(12):1426-1433.
Wang X F,Zhang G Y,Li X H,et al. AFLP analysis of cotton with *Fusarium* and *Verticillium* wilts from the Huanghe and Changjiang Valleys [J]. Acta Genetica Sinica,2004,31(12):1426-1433. (in Chinese)

[18] 王省芬,马峙英,张桂寅,等. 我国棉花抗枯、黄萎病骨干品种(系)基于 AFLP 的遗传多样性 [J]. 棉花学报,2005,17(1):23-28.
Wang X F,Ma Z Y,Zhang G Y,et al. Genetic diversity of Chinese key cottons with *Fusarium* and *Verticillium* wilts resistance based upon AFLPs [J]. Cotton Science,2005,17(1):23-28. (in Chinese)

[19] Pritchard J K,Wen X,Falush D. Documentation for structure software;Version 2. 3[EB/OL]. (2009-04-21)[2010-01-25]. <http://pritch.bsd.uchicago.edu/structure.html>.

[20] Rohlf F J. NTSYS-pc numerical taxonomy and multivariate analysis system [M]. New York: Version 1. 8 Exter Publication,1990.

[21] Buckler Edward Lab. Maize diversity research [EB/OL]. (2009-12-08)[2010-02-25]. http://www.maizegenetics.net/tassel/tassel2.1_standalone.zip.

[22] Xiao J,Wu K,Fang D D,et al. New SSR markers for use in cotton (*Gossypium* spp.) improvement [J]. The Journal of Cotton Science,2009,13:75-157.

[23] Ma X X,Ding Y Z H,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.

[24] An C,Jenkins J N,Wu J,et al. Use of fiber and fuzz mutants to detect QTL for yield components,seed,and fiber traits of upland cotton [J]. Euphytica,2010,172:21-34.