

基于 BSA-seq 法的纤维衣分相关候选基因的鉴定

赵 岩^{1,2}, 冯加加^{1,2}, 肖向辉², 黄晋玲¹, 卢全伟², 渠云芳¹

(¹ 山西农业大学农学院, 晋中 030801; ² 安阳工学院生物与食品工程学院, 安阳 455000)

摘要: 衣分是影响棉纤维产量的重要指标。本研究以高衣分棉花陆海渐渗系材料 MBI7747-14 和中棉所 45 为亲本, 构建包含 2403 个单株的 F₂ 分离群体, 挑选衣分性状极端单株构建 2 个混池, 采用 BSA-seq 方法进行衣分相关基因定位分析。结果表明, 在 D2 染色体上得到 4 个置信指数高于 95% 的候选区间, 4 个区间共包含 236 个基因, 总长度为 5.47 Mb。在以上基因中, 200 个基因中含有 SNP, 190 个基因中含有 InDel, 其中 70 个基因含有非同义突变位点。通过转录组数据的基因表达模式分析, 筛选出 19 个可能与衣分相关的候选基因。GO 功能富集分析表明, 19 个候选基因集中于 NADP⁺ 活性、醛糖醇代谢、碳利用和调控细胞发育等功能条目。本研究为进一步研究棉纤维衣分形成的遗传机制奠定了一定的基础。

关键词: 衣分; BSA-seq; QTL

Identification of Candidate Genes Associating with Fiber Lint Percentage Using BSA-seq

ZHAO Yan^{1,2}, FENG Jia-jia^{1,2}, XIAO Xiang-hui², HUANG Jin-ling¹, LU Quan-wei², QU Yun-fang¹

(¹ College of Agriculture, Shanxi Agricultural University, Jinzhong 030801;

² School of Biological and Food Engineering, Anyang Institute of Technology, Anyang 455000)

Abstract: Lint percentage is an important index affecting cotton fiber yield. In this study, an F₂ segregation population containing 2403 individual plants was constructed derived from chromosome introgression line MBI7747-14 (*Gossypium barbadense* × *Gossypium hirsutum*) with high lint percentage crossing with CCRI45. The BSA-seq was conducted using two bulked samples by pooling individual plants showing higher and lower lint percentage, respectively. As a result, four candidate physical regions showing confidence indices higher than 95% were obtained in a total length of 5.47 Mb on chromosome D2, which contained 236 annotated genes. Out of them, 200 genes contain SNPs, 190 genes contain Indel, and 70 genes contain non-synonymous mutation sites. Through gene expression pattern analysis of transcriptome data, 19 candidate genes possibly associating with lint percentage were identified. GO functional enrichment analysis showed that the 19 candidate genes were enriched in NADP⁺ activity, aldol metabolism, carbon utilization and regulation of cell development. This study laid a foundation for further analyzing the genetic mechanism of cotton fiber lint percentage formation.

Key words: lint percentage; BSA-seq; QTL

收稿日期: 2021-04-21 修回日期: 2021-05-14 网络出版日期: 2021-06-16

URL: <http://doi.org/10.13430/j.cnki.jpgr.20210421002>

第一作者研究方向为棉花种质创新与遗传工程, E-mail: 1761695407@qq.com

通信作者: 渠云芳, 研究方向为棉花远缘杂交育种, E-mail: quyunfang@163.com

卢全伟, 研究方向为棉花分子遗传学, E-mail: daweiayang@163.com

基金项目: 国家自然科学基金 (U1804103); 棉花生物学开放基金国家重点实验室 (CB2020A10); 山西农业大学博士科研启动项目 (2020BQ46)

Foundation projects: National Natural Science Foundation of China (U1804103), Sponsored by State Key Laboratory of Cotton Biology Open Fund (CB2020A10), Doctoral Research Project of Shanxi Agricultural University (2020BQ46)

棉花是世界上重要的经济作物^[1-2],衣分是构成棉花纤维产量的因素之一。衣分作为棉花品种固有的性状特性,遗传较为稳定^[3-4]。研究棉花衣分发育相关基因,对培育高产棉花新品种具有重要意义。

通过构建高密度的遗传连锁图谱,对感兴趣的数量性状位点(QTL, quantitative trait loci)基因进行精细定位是当前常用的定位方法^[5]。朱亚娟等^[6]利用 SSR 标记在 15 号染色体上检测到 2 个衣分 QTLs,其中 qLP-15-1 在 F₂ 群体和 F_{2,3} 家系中都被检测到且位于相同的分子标记 JESPR152~NAU3040 区间内,qLP-15-2 仅在 F_{2,3} 家系中检测到,位于分子标记 NAU5302~NAU2901 区间内。Yu 等^[7]利用 SSR 标记分别在 5、7、16 号染色体上检测出 5 个衣分 QTLs,解释表型变异率为 15.19%~18.06%,其中有 2 个 QTLs 距离很近但不重合。Qin 等^[8]利用 SSR 标记对 241 份棉花种质资源进行分析,定位到 17 个与衣分相关的 QTLs,分别位于 A2、A3、A5、A6、A9、A12、A13、D1、D2、D7、D11、D12 和 D13 染色体上,其中有 5 个 QTLs 与前人研究一致。焦梦佳等^[9]利用 SSR 标记分别在 A10、A12、D6 染色体上各鉴定出 1 个 QTL,分别位于标记 DPL0310、NAU2672 和 NAU3588 附近,解释表型变异率为 5.92%~16.82%。以上定位过程往往需要通过构建大分离群体和数目较多的 DNA 分子标记对群体进行遗传分析,费时费力^[10-11]。利用二代测序技术(NGS, next-generation sequencing)与混合分组分析法(BSA, bulked segregation analysis)相结合的方法,简称为 BSA-seq,可以快速有效地定位与特定性状关联的基因^[12]。近年来该技术已经应用在多种作物重要农艺性状基因定位的研究中^[13-15],Zhu 等^[16]利用 BSA-seq 技术和病毒诱导的基因沉默(VIGS, virus-induced gene silencing)方法对芽黄基因 *virescent-1* 进行精细定位,并成功鉴定出与该性状相关的基因 *GhCHL1*。陈鹏云^[17]联合 BSA-seq 法和构建连锁图谱的方法对棉花早花性状进行 QTL 定位,最终将候选区间定位在 17 号染色体的 2 个区段内:15.96~25.68 Mb、40.33~41.40 Mb。对 2 个区段内中的基因进行了基因结构分析和同源基因进化分析,确定了 3 个与开花相关的候选基因。

前期本课题组从 1 个纤维品质优异渐渗系材料 MBI7747(BC₄F_{3,5})与中棉所 45 进一步回交、自交构建的 BC₆F₂ 群体中筛选到与衣分相关 QTL 的单片段纯合单株(单株号:9174669-16,方便起见后统

称为 MBI7747-14,衣分为 45%),该单株衣分高于亲本材料 MBI7747(衣分为 34.63%)与中棉所 45(衣分为 36.35%)^[18-19]。本研究以中棉所 45 和高衣分渐渗系材料 MBI7747-14 构建的 F₂ 分离群体为研究材料,以其群体中高衣分 DNA 混池(H-pool)和低衣分 DNA 混池(L-pool)为研究对象,采用 BSA-seq 法确定候选区间,结合中棉所 45 和渐渗系材料 MBI7747 在纤维发育 6 个时期转录测序结果,筛选可能的衣分发育候选基因,为棉花衣分发育相关遗传机制研究奠定基础。

1 材料与方法

1.1 试验材料

本研究以 MBI7747-14 为供体亲本与轮回亲本中棉所 45 杂交、自交构建了一个含有 2403 个单株的 F₂ 分离群体,于 2018 年种植在中国农业科学院棉花研究所试验农场(河南省安阳市),常规管理。收获后考种,检测每个单株的籽棉和皮棉,计算衣分。

1.2 DNA 提取及检测

2019 年 8 月,在 F₂ 分离群体中,选择衣分极端性状单株各 50 株,构建高衣分池(H-pool)和低衣分池(L-pool),使用植物 DNA 提取试剂盒(宝生物工程(大连)有限公司,方法参考说明书,有所改进)提取单株 DNA。

由北京贝瑞和康生物技术有限公司将 H-pool 和 L-pool 单株 DNA 经过 Qubit 定量和纯度、完整性鉴定后,对合格单株的 DNA 按照等量混合,之后用标准的 Illumina 流程进行测序工作。对测序得到的 Raw Reads 进行质量评估并过滤得到 Clean Reads,随后利用 BWA 软件把 H-pool 和 L-pool 的 Clean data 比对到棉花参考基因组上^[20],比对结果经 SAMTOOLS 去除重复后进行单核苷酸多态性(SNP, single nucleotide polymorphisms)和插入缺失(InDel, insertion-deletion)标记的检测及注释。采用 GATK 软件进行多个样本 SNP 和 InDel 标记检测;利用 snpEff 软件对 SNP 和 InDEL 进行注释^[21]。

1.3 候选区段分析和变异基因筛选

为了更加直观反映 SNP-index(Single nucleotide polymorphisms index)在染色体的分布,分别计算两个混池中 SNP-index 和 Δ (SNP-index)。以 1 Mb 为窗口,100 kb 为步长,计算每个窗口 SNP-index 的平均值来反映 H-pool 和 L-pool 的 SNP-index 分布,

H-pool 和 L-pool 作差得到 $\Delta(\text{SNP-index})$ 。进行 1000 次置换检验, 选取 95% 置信水平为筛选阈值, 得到两混池的 $\Delta(\text{SNP-index})$ 在染色体上的分布, 利用 QTL-seq 统计学方法中 $\Delta(\text{SNP-index}) > 95\%$ 的置信水平和 G 统计学方法中的 $q\text{-value} < 0.01$ 确定候选区间。在候选区间内, 提取 snpEff 的注释结果, 优先挑取含有非同义突变位点所在的基因。

1.4 候选基因筛选

利用课题组前期对亲本 (MBI7747 和中棉所 45) 在纤维发育的 6 个时期 (5 DPA (Days post anthesis)、7 DPA、10 DPA、15 DPA、20 DPA、25 DPA) 的转录组结果^[18], 设定 GFOLD 阈值, 在候选区段内筛选至少在 1 个纤维发育时期差异表达的基因作为差异表达基因, 把差异基因比对到候选变异基因上, 筛选含有非同义突变位点且差异表达的基

因作为候选基因^[22], 通过 Omicshare 网站 (<https://www.omicshare.com/>) 对基因进行 GO (Gene ontology) 富集分析。

2 结果与分析

2.1 衣分性状统计分析

使用 Microsoft Excel 2010 软件对 F_2 群体中单株衣分进行统计, Origin 软件绘制高低衣分混池箱线图。其中单株棉花衣分最小值为 20.40%, 最大值为 49.20%, 平均值为 32.34%, 峰度系数为 1.27, 偏度系数为 0.01 (表 1)。说明 F_2 群体中棉花单株衣分离散程度低, 表现为单峰分布且不存在比例关系, 符合正态分布, 表现出数量性状遗传特点。L-pool 衣分在 24.62%~29.12% 之间, H-pool 衣分在 38.00%~43.61% 之间, 2 个极端混池中衣分性状存在显著差异 (图 1), 符合 BSA 建池要求。

表 1 衣分性状统计表
Table 1 Statistics of lint percentage characters

材料 Material	样本数 Number of samples	最小值 (%) Min.	最大值 (%) Max.	平均值 (%) Mean	变异系数 (%) CV	峰度 Kurt	偏度 Skew
F_2	2403	20.40	49.20	32.34	9.20	1.27	0.01
高衣分池 H-pool	50	38.00	43.61	39.35	2.93	2.88	1.58
低衣分池 L-pool	50	24.62	29.12	27.87	3.90	0.38	-1.02

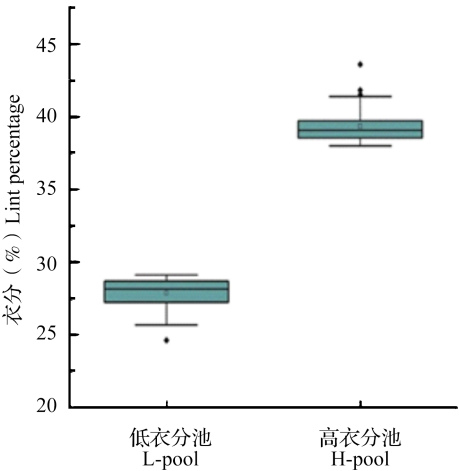


图 1 BSA 混池衣分表型值
Fig.1 The phenotypic value of lint percentage in both pools

2.2 BSA-seq 测序数据分析

对 H-pool 和 L-pool 分别进行重测序, 过滤后共得到 284.21 G 的 clean reads, 2 个样本中质控后的数据能够比对到参考基因组上的比例非常高,

均在 96% 以上, GC 含量正常, 分别是 37.69% 和 37.70%, 样品平均测序深度在 $58.76 \times$ 以上, $1 \times$ reads 覆盖度比率在 99.48% 以上, $10 \times$ reads 覆盖度比率在 99.12% 以上 (表 2), 由此可见本实验所获得的测序数据不仅数据质量高而且没有污染, 所有样本的数据量足够, 测序正常, 测序数据与棉花参考基因组比对结果正常, 可用于后续的分析。

对 2 个混池中所有存在差异的变异位点进行汇总, SNP 的检测与注释结果表明 H-pool 和 L-pool 共获得 4080280 个 SNPs, 其中错义突变的 SNP 有 26536 个, 同义突变的 SNP 有 15007 个, 终止子提前 SNP 有 618 个, 终止子丢失 SNP 有 174 个, 转换 / 颠倒是 1.94。InDel 的检测与注释结果表明 H-pool 和 L-pool 共获得 878785 个 InDels, 其中引起移码突变的有 1464 个。SNP 和 InDel 的变异类型主要是基因上游、基因下游、基因间区和内含子, 基因编码区的变异少于基因非编码区 (表 3)。

表 2 BSA-seq 测序数据统计
Table 2 BSA-seq sequencing data statistics

样本 Sample	总 reads 数 Number of reads	比对 reads 数 Mapped reads	比对比例 (%) Mapped ratio	GC 含量 (%) GC content	平均测序深度 (×) Average sequencing depth	1× reads 覆盖 度比率 (%) 1× reads coverage ratio	10× reads 覆盖 度比率 (%) 10× reads coverage ratio
高衣分池 H-pool	869281932	851907314	98.00	37.69	58.76	99.61	99.12
低衣分池 L-pool	1016199300	985046259	96.93	37.70	66.58	99.48	99.21

表 3 变异位点注释统计表
Table 3 Statistical table of notes of variant sites

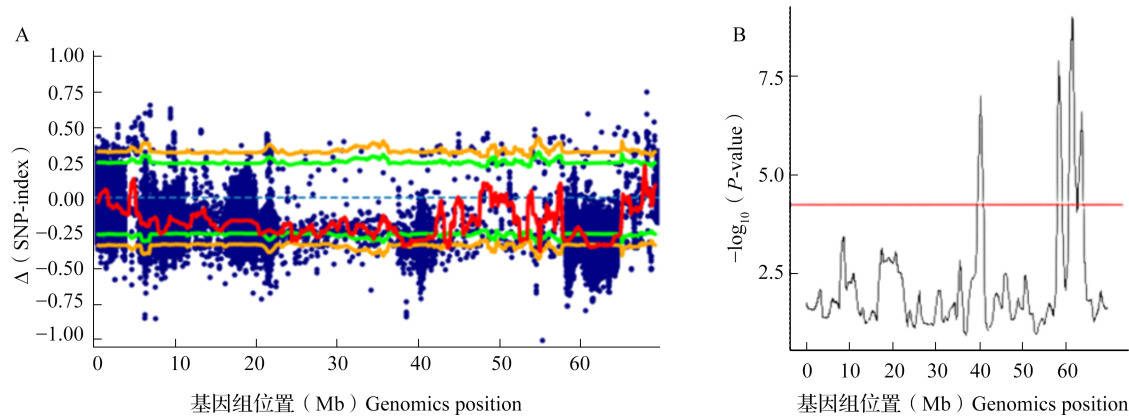
变异位点信息 Variant site information	SNP 数目 Number of SNP	InDel 数目 Number of InDel	变异位点信息 Variant site information	SNP 数目 Number of SNP	InDel 数目 Number of InDel
外显子 Exon	42164	2288	错义突变 Missense	26536	-
内含子 Intron	117647	29142	同义突变 Synonymous	15007	-
基因上游 Upstream	719041	160073	移码突变 Frameshift	-	1464
基因下游 Downstream	588230	196944	转换 Transitions	3452682	-
基因间区 Intergenic	2608578	488954	颠倒 Transversions	1777911	-
终止子提前 Stop gained	618	72	转换 / 颠倒 Ts/Tv	1.94	-
终止子丢失 Stop lost	174	22	总数 Total	4080280	878785

- 表示数据不存在
- Indicates that data do not exist

2.3 候选区间及分析

利用 Δ (SNP-index) 和 G-value 结合的方法,在 D2 染色体得到 4 个 QTL 候选区间 (QTL1: 37.80~39.09 Mb, QTL2: 55.33~56.50 Mb, QTL3: 57.85~59.58 Mb, QTL4: 59.88~61.16 Mb),总大小为 5.47 Mb

(图 2)。这 4 个候选区间内共包含 236 个基因, 198 个被注释。在候选区间内,共有 7912 个 SNP 变异位点分布在 200 个基因上, 2077 个 InDel 变异位点分布在 190 个基因上,含有非同义突变位点的基因有 70 个。



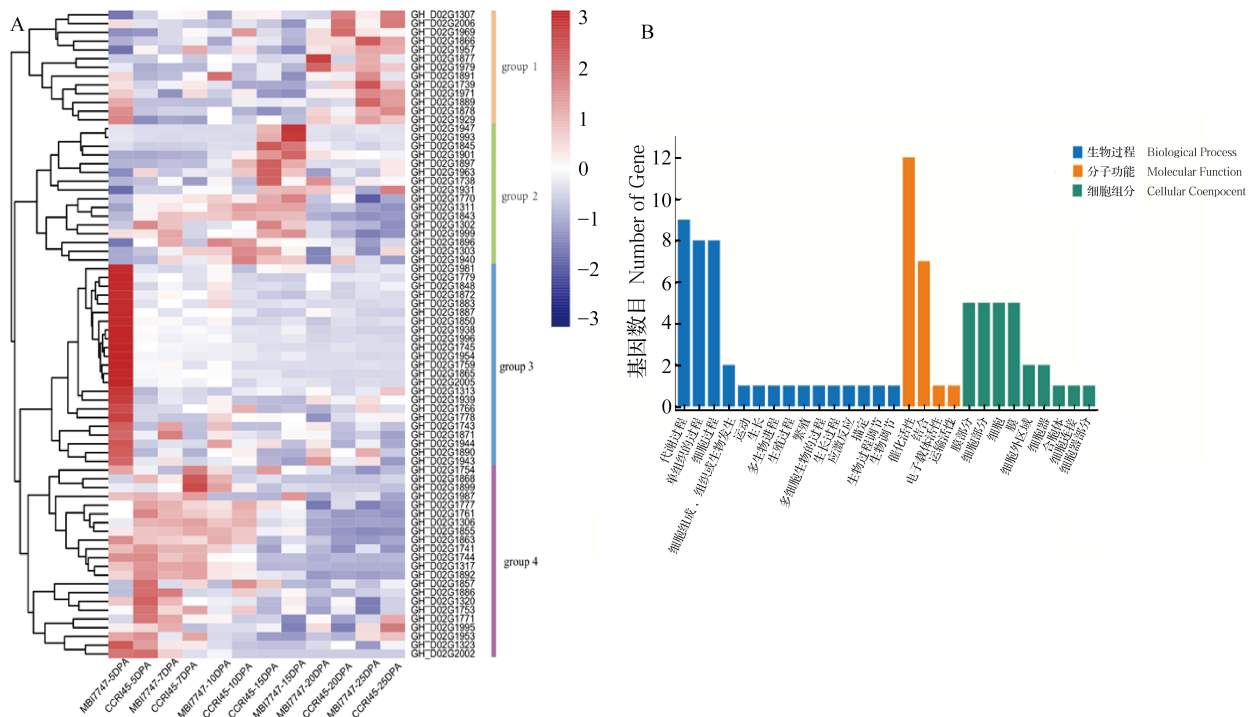
A: Δ (SNP-index) 在 D2 染色体的分布,蓝色点为 Δ (SNP-index),橙色线为 99% 置信线,绿色线为 95% 置信线,红色线为 Δ (SNP-index) 划窗后的拟合线; B: G value 在 D2 染色体的分布,红色线为 $-\log_{10}(P\text{-value})$ 阈值线
A: The distribution of Δ (SNP-index) on chromosome D2, blue dot is Δ SNP index, orange line is 99% confidence line, green line is 95% confidence line, red line is fitting line after Δ (SNP-index) window,
B: The distribution of G value on chromosome D2, the red line indicate the $-\log_{10}(P\text{-value})$ threshold line

图 2 候选区间在 D2 染色体的分布
Fig.2 Distribution of candidate intervals on chromosome D2

2.4 候选区间内差异表达基因分析

根据前期课题组对 MBI7747 和中棉所 45 不同纤维发育时期的转录组数据,寻找同一时期两材料间的差异表达基因。GFOLD 绝对值大于 0.5 的基因被视为差异表达基因。候选区间内共有 236 个基因,其中有 74 个基因至少在纤维发育某一个时期差异表达。根据上述基因的表达量 (FPKM, fragments per kilobase of exon per million mapped) 使用 R 包绘制热图 (图 3),并以表达量的聚类特征将基因分

为 4 组。从图 4 可知,第 1 组含有 13 个基因,主要在 20 DPA 和 25 DPA 的纤维组织中差异表达;第 2 组含有 16 个基因,主要在 10 DPA 和 15 DPA 的纤维组织中差异表达;第 3 组含有 23 个基因,主要在 5 DPA 的纤维组织中差异表达;第 4 组包括 22 个基因,主要在 5 DPA、7 DPA 和 10 DPA 的纤维组织中差异表达。根据纤维发育过程可以发现,以上 4 组的基因主要参与纤维细胞的伸长期和加厚期过程。



A: 热图; B: GO 富集分析

A: Heatmap, B: GO terms

图 3 候选基因的功能和表达分析

2.5 候选基因分析

在候选区间内,将 74 个差异表达基因和 70 个含有非同义突变的基因进行取交集,得到 19 个非同义突变的差异表达基因,将含有非同义突变的差异表达基因预测为调控纤维衣分的候选基因。并通过 CottonFGD (<https://cottonfgd.org/search/>) 进行基因注释 (表 4)。对候选基因进行 GO 富集分析 (图 3B)。GO 富集分析结果显示,19 个基因涉及 28 个条目,生物过程包含的 GO 功能分类最多,共 15 种,大多与代谢过程、单组织的过程和细胞过程有关;在分子功能中,大多数基因富集在催化活性和结合功能条目;在细胞组分中,注释到的基因多定位于细胞膜部分。

3 讨论

近年来,高通量测序的发展促进了 BSA-seq 方法在水稻、玉米、油菜等作物的重要性状相关基因定位中的应用。Wang 等^[23]利用 BSA-seq 在油菜 A6 染色体鉴定出 1 个与油菜枝分叉角相关的主效 QTL (ba1),随后用 InDel 标记在 F₂ 群体中进行作图分析,最终得到 82 个与油菜枝分叉角相关的基因,并根据其表达情况及拟南芥同源基因注释结果,将 *BnaA0639380D* 作为油菜分枝角的候选基因。Song 等^[24]利用 BSA-seq 在大豆 1、11 号染色体上定位到 2 个控制大豆种子子叶颜色的 2 个候选区间: qCC1 和 qCC2,在 2 个候选区间内使用 SSR 标记完成了精细定位,最终 qCC1 被定位到 1

表 4 候选基因功能注释
Table 4 Functional annotation of candidate genes

候选区间 Candidate interval	基因 ID Gene ID	基因名 Gene Name	功能注释 Function annotation	序列长度 (bp) Sequence length	非同义突变位点数目 Number of non synonymous variant sites
QTL1	GH_D02G1303	<i>At3g17800</i>	UV-B 诱导蛋白 At3g17800	1658	1
	GH_D02G1306	<i>ABCG11</i>	ABCG 11	2941	1
	GH_D02G1307	<i>DHBK</i>	3, 4- 二羟基 -2- 丁酮酶	8374	1
	GH_D02G1311	<i>SKOR</i>	钾离子通道 SKOR	21061	2
QTL2	GH_D02G1759	<i>TBL2</i>	Ⅱ类毛状体蛋白	2654	1
	GH_D02G1766	<i>PCMP-E14</i>	含五肽重复序列的蛋白质 At5g27110	1973	2
	GH_D02G1771	<i>At1g60420</i>	核氧还蛋白 I	2152	1
	GH_D02G1778	<i>NA</i>	碳酸酐酶	6753	1
	GH_D02G1779	<i>AOR</i>	NADPH 依赖性烯醛酶	4150	1
QTL3	GH_D02G1845	<i>LBD42</i>	LOB 结构域的蛋白 42	780	2
	GH_D02G1855	<i>CSE</i>	咖啡酰基酯酶	3224	1
	GH_D02G1857	<i>AtMg00310</i>	线粒体蛋白 AtMg00310	1663	4
	GH_D02G1866	<i>alxA</i>	ALG-2 互作蛋白 X	7401	1
	GH_D02G1871	<i>nadD</i>	烟酸核苷酸腺苷酸转移酶	5692	1
	GH_D02G1877	<i>LRK10L-1.2</i>	叶锈病 10 抗病基因座受体样蛋白 激酶样 1.2	1187	1
	GH_D02G1878	<i>NA</i>	葡萄糖 -6- 磷酸 1- 脱氢酶	3474	2
QTL4	GH_D02G1892	<i>Os08g0536000</i>	丙酮酸脱氢酶 E1 亚单位 β-1	5035	1
	GH_D02G1971	<i>STA1</i>	STA1 蛋白	3101	2
	GH_D02G2006	<i>petC</i>	细胞色素 b6-f 复合铁硫亚单位	1820	1

个包含 4 个注释基因的 30.7 kb 区域, qCC2 被定位到 1 个包含 9 个基因的 67.7 kb 区域。Guo 等^[25]利用 BSA-seq 在 6、9 号染色体定位到大小 0.82 Mb 的候选区域, 其中包含 98 个基因, 结合 Longjing25 和 Longjing11 转录组结果在候选区间内找到 50 个差异表达基因, 经过 KEGG 代谢通路富集分析, 最终在水稻孕穗期确定 4 个耐冷候选基因。本研究中利用 BSA-seq 法在 D2 染色体找到 4 个与纤维衣分相关的 QTL 候选区间, 邓晓英^[26]在 D2 染色体 57.79~59.16 Mb 和 60.54~60.58 Mb 2 个区间内各定位到 1 个与衣分相关 QTL, 与其衣分 QTL 定位结果相比并通过软件 e-PCR 锚定到参考基因组上发现, 重叠区间分别是 1.31 Mb 和 0.04 Mb。另外, 张保才^[27]在 D2 染色体 58.89 Mb 附近定位到 1 个与衣分相关的 QTL。贾永斌^[28]在 D2 染色体 60.08 Mb 附近定位到 1 个与衣分相关的 QTL。刘桂珍^[29]对 180 份优良陆地棉品系产

量、纤维品质、熟性、株型、种子品质等数量性状进行关联分析, 在 D2 染色体发现 1 个衣分关联的标记位点 CGR5657。张淑文^[30]对 115 个渐渗系材料进行产量和纤维品质评价, 在 D2 染色体发现 1 个衣分关联的标记位点 HAU3236, 但是 CGR5657 和 HAU3236 标记并没有锚定在 D2 染色体上, 具体原因未知。另外在本研究中 QTL1 和 QTL2 两个候选区间内并未发现有 QTL 报道。以上说明了本研究中 QTL 定位的准确性, 并证实 BSA-seq 法在数量性状定位中的可行性。

衣分是棉花纤维产量和籽棉产量的百分比率, 纤维产量的形成与纤维的生长过程密切相关^[31]。棉纤维是胚珠表皮层的单细胞伸长加厚形成的, 每个单细胞都可以发育成单根棉纤维^[32-33]。棉纤维伸长期起始于开花当天, 在伸长期结束前, 纤维细胞会形成含有中腔的细长薄壁管状物; 次生壁加厚期主要是细胞壁中纤维素的不断累积, 最终形

成两端细中间粗的一根棉纤维^[34]。伸长期和次生壁加厚期在纤维的整个生长期占据的时间比例较大,说明这 2 个时期对纤维产量的形成具有重要作用。本研究在 BSA-seq 法确定的 4 个 QTL 候选区间内,结合亲本转录组结果和功能注释分析,初步确定了 19 个与调控衣分发育的候选基因,其中 3 个基因 (*GH_D02G1845*、*GH_D02G1855* 和 *GH_D02G1857*) 均在纤维伸长生长期 (0~20 DPA) 差异表达,且都在 QTL3 内,*GH_D02G1845* 被注释为含 LOB 结构域蛋白 42,属于 AS2/LOB 家族,参与拟南芥侧生器官的近轴细胞命运调控^[35-36]; *GH_D02G1855* 被注释为咖啡酰莽草酸酯酶,参与木质素单体生物合成特异途径^[37-39]; *GH_D02G1857* 功能未知;表明上述 3 个基因可能在纤维伸长过程发挥重要作用。另外,有 2 个基因 (*GH_D02G1866* 和 *GH_D02G1971*) 均在纤维加厚生长期差异表达,且都在 QTL4 内,*GH_D02G1866* 被注释为 ALIX 蛋白,在拟南芥中参与 ABA 的调控过程^[40]; *GH_D02G1971* 被注释为 STA1 蛋白,参与响应植物的非生物胁迫^[41-42];因此,QTL4 内的 2 个基因可能对纤维加厚过程起到重要作用。另外,在 QTL1 内预测到基因 *GH_D02G1311*,该基因被注释为钾通道蛋白,参与纤维细胞的快速伸长^[43];在 QTL2 内预测到基因 *GH_D02G1759*,该基因被注释为 TBL2,属于 TBL 家族,参与细胞壁多糖乙酰化修饰,研究发现 TBL38 在纤维伸长发育过程中具有促进作用^[44];推测这 2 个基因通过促进纤维伸长生长来影响纤维衣分。其余的候选基因对纤维衣分的影响还有待进一步的研究。后续可进行基因序列分析,进一步明确基因中突变位点引起的氨基酸序列变化或蛋白功能变化,通过 VIGS、利用基因编辑技术结合棉花遗传转化进行基因功能验证等。

参考文献

- [1] Shi Y Z, Li W T, Li A G, Ge R H, Zhang B C, Li J Z, Liu G P, Li J W, Liu A Y, Shang H H, Gong J W, Gong W K, Yang Z M, Tang F, Liu Z, Zhu W P, Jiang J X, Yu X N, Wang T, Wang W, Chen T, Chen T B, Zhang Z S, Zhang Z L. Constructing a high-density linkage map for *Gossypium hirsutum* x *Gossypium barbadense* and identifying QTLs for lint percentage. *Journal of Integrative Plant Biology*, 2015, 57 (5): 450-467
- [2] Liu D X, Liu F, Shan X R, Zhang J, Tang S Y, Fang X M, Liu X Y, Wang W W, Tan Z Y, Teng Z H, Zhang Z S, Liu D J. Construction of a high-density genetic map and lint percentage and cottonseed nutrient trait QTL identification in upland cotton (*Gossypium hirsutum* L.). *Molecular Genetics and Genomics*, 2015, 290 (5): 1683-1700
- [3] 王宁,苏桂兰,周红,许庆华,黄群,严根土. 中国与美国棉花品种的铃质量和衣分差异分析. *河南农业科学*, 2015, 44 (7): 43-47
Wang N, Su G L, Zhou H, Xu Q H, Huang Q, Yan G T. Difference analysis of boll weight and lint index of cotton between Chinese and American varieties. *Journal of Henan Agricultural Sciences*, 2015, 44 (7): 43-47
- [4] 喻树迅,范术丽,王寒涛,魏恒玲,庞朝友. 中国棉花高产育种研究进展. *中国农业科学*, 2016, 49 (18): 3465-3476
Yu S X, Fan S L, Wang H T, Wei H L, Pang C Y. Progresses in research on cotton high yield breeding in China. *Scientia Agricultura Sinica*, 2016, 49 (18): 3465-3476
- [5] 袁文娟,覃鸿妮,王旭,王国强,王久光,刘志斋,蔡一林. 玉米雄穗颜色的 QTL 分析. *植物遗传资源学报*, 2013, 14 (2): 289-297
Yuan W J, Qin H N, Wang X, Wang G Q, Wang J G, Liu Z Z, Cai Y L. Identification of QTL for tassel color in Maize. *Journal of Plant Genetic Resources*, 2013, 14 (2): 289-297
- [6] 朱亚娟,王鹏,郭旺珍,张天真. 利用海岛棉染色体片段导入系定位衣分和籽指 QTL. *作物学报*, 2010, 36 (8): 1318-1323
Zhu Y J, Wang P, Guo W Z, Zhang T Z. Mapping QTLs for lint percentage and seed index using *Gossypium barbadense* chromosome segment introgression lines. *Acta Agronomica Sinica*, 2010, 36 (8): 1318-1323
- [7] Yu J W, Zhang K, Li S Y, Yu S X, Zhai H H, Wu M, Li X L, Fan S L, Song M Z, Yang D G, Li Y H, Zhang J F. Mapping quantitative trait loci for lint yield and fiber quality across environments in a *Gossypium hirsutum* x *Gossypium barbadense* backcross inbred line population. *Theoretical and Applied Genetics*, 2013, 126 (1): 275-287
- [8] Qin H D, Chen M, Yi X D, Bie S, Zhang C, Zhang C Y, Lan J Y, Meng Y Y, Yuan Y L, Jiao C H. Identification of associated SSR markers for yield component and fiber quality traits based on frame map and Upland cotton collections. *PLoS ONE*, 2015, 10 (1): e118073
- [9] 焦梦佳,陈煜,宋章强,霍雪寒,高阳,周娟,王静静,潘奥,赵程杰,王芙蓉,张军. 利用棉花优质渐渗系进行纤维品质性状和衣分的 QTL 定位. *植物遗传资源学报*, 2020, 21 (3): 716-726
Jiao M J, Chen Y, Song Z Q, Cui X H, Gao Y, Zhou J, Wang J J, Pan A, Zhao C J, Wang F R, Zhang J. QTL mapping for fiber quality and lint percentage traits of cotton using elite germplasm with introgression of *Gossypium barbadense* L.. *Journal of Plant Genetic Resources*, 2020, 21 (3): 716-726
- [10] Teng C C, Du D Z, Xiao L, Yu Q L, Shang G X, Zhao Z G. Mapping and identifying a candidate gene (*Bnmfs*) for female-male sterility through whole-genome resequencing and RNA-seq in rapeseed (*Brassica napus* L.). *Frontiers in Plant Science*, 2017, 8: 2086
- [11] 徐剑文,刘剑光,赵君,王希睿,肖松华. 利用 BSA-seq 发掘棉花适宜机采的果枝长度相关 QTL. *棉花学报*, 2019, 31 (4): 319-326
Xu J W, Liu J G, Zhao J, Wang X R, Xiao S H. The identification of QTL associated with cotton fruit branch length suitable for mechanized harvest utilizing BSA-seq. *Cotton Science*, 2019, 31 (4): 319-326
- [12] Pandey M K, Khan A W, Singh V K, Vishwakarma M K,

- Shasidhar Y, Kumar V, Garg V, Bhat R S, Chitikineni A, Janila P, Guo B Z, Varshney R K. QTL-seq approach identified genomic regions and diagnostic markers for rust and late leaf spot resistance in groundnut (*Arachis hypogaea* L.). *Plant Biotechnol Journal*, 2017, 15 (8): 927-941
- [13] Illa-Berenguer E, Van H J, Huang Z J, van der K E. Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. *Theoretical and Applied Genetics*, 2015, 128 (7): 1329-1342
- [14] Ogiso-Tanaka E, Tanaka T, Tanaka K, Nonoue Y, Sasaki T, Fushimi E, Koide Y, Okumoto Y, Yano M, Saito H. Detection of novel QTLs *qDTH4.5* and *qDTH6.3*, which confer late heading under short-day conditions, by SSR marker-based and QTL-seq analysis. *Breeding Science*, 2017, 67 (2): 101-109
- [15] Borovsky Y, Monsonogo N, Mohan V, Shabtai S, Kamara I, Faigenboim A, Hill T, Chen S Y, Stoffel K, van Deynze A, Paran I. The zinc-finger transcription factor *CcLOLI* controls chloroplast development and immature pepper fruit color in *Capsicum chinense* and its function is conserved in tomato. *The Plant Journal*, 2019, 99 (1): 41-55
- [16] Zhu J K, Chen J D, Gao F K, Xu C Y, Wu H T, Chen K, Si Z F, Yan H, Zhang T Z. Rapid mapping and cloning of the virescent-1 gene in cotton by bulked segregant analysis-next generation sequencing and virus-induced gene silencing strategies. *Journal of Experimental Botany*, 2017, 68 (15): 4125-4135
- [17] 陈鹏云. 棉花早熟性状相关基因的定位与鉴定. 长沙: 湖南农业大学, 2017
- Chen P Y. Fine mapping and identification of early maturity related genes in cotton (*Gossypium hirsutum* L.). Changsha: Hunan Agricultural University, 2017
- [18] 卢全伟. 陆海渐渗系纤维长度 QTL (qFL-12-2) 精细定位与候选基因鉴定. 太谷: 山西农业大学, 2017
- Lu Q W. Fine mapping and candidate gene identification of qFL-12-2 in chromosome introgression line. Taigu: Shanxi Agricultural University, 2017
- [19] 马留军. 陆海杂种回交高代染色体片段代换系的评价及 QTL 定位. 北京: 中国农业科学院, 2013
- Ma L J. The evaluation and QTL identifying of chromosome segment substitution lines in advanced backcross of *Gossypium Hirsutum* × *G. Barbardense*. Beijing: Chinese Academy of Agricultural Sciences, 2013
- [20] Hu Y, Chen J D, Fang L, Zhang Z Y, Ma W, Niu Y C, Ju L Z, Deng J Q, Zhao T, Lian J M, Baruch K, Fang D, Liu X, Ruan Y L, Rahman M, Han J L, Wang K, Wang Q, Wu H T, Mei G F, Zang Y H, Han Z G, Xu C Y, Shen W J, Yang D F, Si Z F, Dai F, Zou L F, Huang F, Bai Y L, Zhang Y G, Brodt A, Hilla B H, Zhu X F, Zhou B L, Guan X Y, Zhu S J, Chen X Y, Zhan T Z. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nature Genetics*, 2019, 51 (4): 739-748
- [21] Cingolani P, Platts A, Wang L L, Coon M, Nguyen T, Wang L, Land S J, Lu X Y, Ruden D M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸; iso-2; iso-3. *Fly (Austin)*, 2012, 6 (2): 80-92
- [22] Liu G Y, Pei W F, Li D, Ma J J, Cui Y P, Wang N H, Song J K, Wu M, Li L B, Zang X S, Yu S X, Zhang J F, Yu J W. A targeted QTL analysis for fiber length using a genetic population between two introgressed backcrossed inbred lines in upland cotton (*Gossypium hirsutum*). *The Crop Journal*, 2018, 3 (7): 273-282
- [23] Wang H, Cheng H T, Wang W X, Liu J, Hao M Y, Mei D S, Zhou R J, Fu L, Hu Q. Identification of *BnaYUCCA6* as a candidate gene for branch angle in *Brassica napus* by QTL-seq. *Scientific Reports*, 2016, 6: 38493
- [24] Song J, Li Z, Liu Z X, Guo Y, Qiu L J. Next-generation sequencing from bulked-segregant analysis accelerates the simultaneous identification of two qualitative genes in *Soybean*. *Frontiers in Plant Science*, 2017, 8: 919
- [25] Guo Z H, Cai L J, Chen Z Q, Wang R Y, Zhang L M, Guan S W, Zhang S H, Ma W D, Liu C X, Pan G J. Identification of candidate genes controlling chilling tolerance of rice in the cold region at the booting stage by BSA-Seq and RNA-Seq. *Royal Society Open Science*, 2020, 7 (11): 201081
- [26] 邓晓英. 杂交棉中棉所 70 产量和纤维品质性状的 QTL 定位. 太谷: 山西农业大学, 2016
- Deng X Y. Identification of QTL for yield and fiber quality traits in hybrid cotton CCRI70. Taigu: Shanxi Agricultural University, 2016
- [27] 张保才. AB-QTL 法定位海岛棉优异纤维品质基因和抗黄萎病基因. 北京: 中国农业科学院, 2006
- Zhang B C. QTL analysis of fiber quality and resistance to verticillium wilt using *Gossypium hirsutum* × *Gossypium barbadense* backcross populations. Beijing: Chinese Academy of Agricultural Sciences, 2006
- [28] 贾永斌. 海岛棉产量、纤维品质性状 QTL 有利等位基因鉴定. 重庆: 西南大学, 2020
- Jia Y B. Identification of favorable alleles of QTL for yield and fiber quality traits of *Gossypium barbadense*. Chongqing: Southwest University, 2020
- [29] 刘桂珍. 中国陆地棉品种重要性状的关联分析及 DNA 指纹条形码构建. 南京: 南京农业大学, 2015
- Liu G Z. Association analysis of important traits and construction of DNA fingerprinting barcode in Chinese upland cotton cultivars. Nanjing: Nanjing Agricultural University, 2015
- [30] 张淑文. 陆地棉 TM-1 为背景的两个陆地棉半野生种染色体片段渐渗系的创建及鉴定. 南京: 南京农业大学, 2015
- Zhang S W. Development and evaluation of introgression lines from two feral landraces into *Gossypium hirsutum* ACC TM-1. Nanjing: Nanjing Agricultural University, 2015
- [31] 马军, 王芳, 郭玉平, 郭显, 陈霞, 孙学振, 宋宪亮. 高衣分染色体片段导入系 IL-10-1 衣分形成的研究. 泰安: 山东农业科学, 2012, 44 (10): 25-29
- Ma J, Wang F, Guo Y P, Guo X, Chen X, Sun X Z, Song X L. Study on lint percentage formation mechanism of chromosome segment introgression line IL-10-1 with high lint percentage. Taian: Shandong Agricultural Sciences, 2012, 44 (10): 25-29
- [32] 薛宇. 雷蒙德氏棉 *HD-Zip* 基因家族进化及选择性剪切分析. 杭州: 浙江农林大学, 2014
- Xu Y. Genome-wide analysis of evolution and alternative splicing event of *HD-Zip* gene family in *Gossypium ramondii*. Hangzhou: Zhejiang A & F University, 2014.
- [33] 毛玮. 亚洲棉纤维发育关键基因 *GaHDI* 上游调控因子解

- 析. 杭州: 浙江农林大学, 2018
- Mao W. Analysis of upstream regulatory factors of *GaHDL*, a key gene for fiber development in *Gossypium arboreum*. Hangzhou: Zhejiang A & F University, 2018
- [34] Xu Z Y, Kohel R J, Song G L, Cho J, Alabady M, Yu J, Koo P, Chu J, Yu S X, Wilkins T A, Zhu Y X, Yu J Z. Gene-rich islands for fiber development in the cotton genome. *Genomics*, 2008, 92 (3): 173-183
- [35] Matsumura Y, Iwakawa H, Machida Y, Machida C. Characterization of genes in the ASYMMETRIC LEAVES2/ LATERAL ORGAN BOUNDARIES (AS2/LOB) family in *Arabidopsis thaliana*, and functional and molecular comparisons between AS2 and other family members. *The Plant Journal*, 2009, 58 (3): 525-537
- [36] Wang Y B, Song J P, Wang Z B. ASYMMETRIC LEAVES2-LIKE38, one member of AS2/LOB gene family, involves in regulating ab-adaxial patterning in *Arabidopsis* lateral organs. *Acta Physiologiae Plantarum*, 2015, 37 (9): 185
- [37] Vanholme R, Cesarino I, Rataj K, Xiao Y G, Sundin L, Sundin L, Goeminne G, Kim H, Cross J, Morreel K, Araujo P, Welsh L, Hastraete J, McClellan C, Vanholme B, Ralph J, Simpson G G, Halpin C, Boerjan W. Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. *Science*, 2013, 341 (6150): 1103-1106
- [38] 晁楠. 毛白杨木质素单体生物合成途径关键酶基因家族的功能分化研究. 北京: 北京林业大学, 2017
- Chao N. Functional divergence of key genes involved in monolignol biosynthesis in *Populus tomentosa*. Beijing: Beijing Forestry University, 2017
- [39] 王雪纯. 日本落叶松 CSE 和 DHN 基因功能分析. 北京: 北京林业大学, 2020
- Wang X C. Functional analysis of CSE and DHN in *Larix kaempferi*. Beijing: Beijing Forestry University, 2020
- [40] García-León M, Cuyas L, El-Moneim D A, Rodriguez L, Belda-Palazón B, Sanchez-Quant E, Fernández Y, Roux B, Zamarreño Á M, García-Mina J M, Nussaume L, Rodriguez P L, Paz-Ares J, Leonhardt N, Rubio V. *Arabidopsis* ALIX regulates stomatal aperture and turnover of abscisic acid receptors. *The Plant Cell*, 2019, 31 (10): 2411-2429
- [41] Kim G D, Yoo S D, Cho Y H. STABILIZED1 as a heat stress-specific splicing factor in *Arabidopsis thaliana*. *Plant Signaling & Behavior*, 2018, 13 (2): e1432955
- [42] Lee B H, Kapoor A, Zhu J, Zhu J K. STABILIZED1, a stress-upregulated nuclear protein, is required for pre-mRNA splicing, mRNA turnover, and stress tolerance in *Arabidopsis*. *The Plant Cell*, 2006, 18 (7): 1736-1749
- [43] Ruan Y L, Llewellyn D J, Furbank R T. The control of single-celled cotton fiber elongation by developmentally reversible gating of plasmodesmata and coordinated expression of sucrose and K⁺ transporters and expansin. *The Plant Cell*, 2001, 13 (1): 47-60
- [44] 张弛. 陆地棉纤维品质和产量性状的全基因组关联分析及候选基因的功能验证. 咸阳: 西北农林科技大学, 2020
- Zhang C. Genome-wide association analysis of fiber quality and yield and functional analysis of candidate genes in upland cotton. Xianyang: Northwest A&F University, 2020