Theor Appl Genet (2012) 125:1263–1274
DOI 10.1007/s00122-012-1911-x

ORIGINAL PAPER

Molecular diversity, genomic constitution, and QTL mapping of fiber quality by mapped SSRs in introgression lines derived from *Gossypium hirsutum* × *G. darwinii* Watt

Bin Wang · Yichun Nie · Zhongxu Lin · Xianlong Zhang · Junjie Liu · Jing Bai

Received: 30 September 2011 / Accepted: 5 June 2012 / Published online: 26 June 2012

Abstract Because the genetic basis of current upland cotton cultivars is narrow, exploring new germplasm resources and discovering novel alleles relevant to important agronomic traits have become two of the most important themes in the field of cotton research. In this study, *G. darwinii* Watt, a wild cotton species, was crossed with four upland cotton cultivars with desirable traits. A total of 105 introgression lines (ILs) were successfully obtained. By using 310 mapped SSRs evenly distributed across the interspecific linkage map of *G. hirsutum* × *G. barbadense*, these 105 ILs and their corresponding parents were analyzed. A total of 278 polymorphic loci were detected among the 105 ILs, and the average length of introgression segments accumulated to 333.5 cM, accounting for 6.7 % of the whole genome. These lines included many variations. However, high similarity coefficients existed between lines, even between those derived from different parents. Finally, all the ILs and their upland cotton parents were used for association mapping of fiber quality in three environments. A total of 40 SSRs were found to be associated with five fiber quality indexes ($P < 0.05$) with some being detected in multiple environments and traits. The contribution rate for trait variation was 6.31 % on average, ranging from 2.00 to 14.79 %. This study develops novel ILs for cotton genetics and breeding and provides the basis for future research into fine mapping of genes related to fiber quality, analyses of candidate gene association, and molecular marker-assisted breeding.

Communicated by M. Frisch.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-1911-x) contains supplementary material, which is available to authorized users.

B. Wang · Y. Nie · Z. Lin (✉) · X. Zhang · J. Liu · J. Bai
National Key Laboratory of Crop Genetic Improvement and National Centre of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, Hubei, People's Republic of China
e-mail: linzhongxu@126.com

Introduction

Introgression lines (ILs), which are also called chromosome segment substitution lines (CSSLs), are optimal tools for the fine mapping of QTLs and genes and for evaluating gene action without complicated dominant and epistatic effects (Zhu et al. 2009; Zhao et al. 2009; Ali et al. 2010). Due to a series of backcrossing and self-crossing events, most apparent traits except for a few ILs have been found to be similar to those of the receptor parent. These ILs are distinct from each other by several DNA polymorphisms, which were substituted by different segments from the donor parent. QTL mapping can be executed conveniently through simple statistical methods, such as one-way analysis of variance (ANOVA) and *t* testing. Both single-chromosome segment substitution (SCSS) lines and ILs with overlapping segments can be used to increase the precision of fine mapping (Takershi et al. 2005; Chen et al. 2009; Ye et al. 2010).

Molecular evaluation of germplasm resources includes the evaluation of genetic diversity, molecular identification, and the discovery of novel alleles. Genetic diversity and genomic structure of germplasm resources are the basis for germplasm identification and the application of those genetic resources. They could lay the foundation for the introduction, cultivation, and protection of germplasm resources suitable for breeding. The discovery of novel alleles could help the research community determine the laws that govern target traits, which would be useful for marker-assisted breeding.

The *Gossypium darwinii* Watt, $(AD)_5$ genome, allotetraploid cotton, is only found in the Galapagos Islands, and it belongs to the *Karpas rafinesque* subgenus with *G. hirsutum* (upland cotton), *G. barbadense*, *G. tomentosum*, and *G. mustelium*. According to the classification of cotton germplasm resources, *G. Darwinii* Watt and upland cotton belong to the first germplasm bank and can hybridize with each other directly (Pan 1998). *G. darwinii* Watt has genes that may confer drought resistance, nematode resistance, and fiber fineness (Liang 1999). These genes can be introduced into upland cotton to improve upland cotton cultivars. Because the genetic diversity of upland cotton is narrow, it is necessary to widen its genetic basis (Brubaker and Wendel 1994; Iqbal et al. 1997). *G. darwinii* Watt may be suitable for this purpose.

In order to widen the genetic basis of current upland cotton cultivars, to explore new germplasm resources, and to discover novel alleles with superior fiber quality, *G. darwinii* Watt was crossed with *G. hirsutum* cv. Ejing1, and the offspring were backcrossed with different upland cotton cultivars with outstanding traits as part of a multi-target breeding program. This project is aimed at overcoming the shortcomings of upland cotton cultivars while maintaining their superior traits. A set of 105 progeny ILs of advanced generations were obtained. We then evaluated the genetic diversity of these ILs and analyzed the distribution of chromosome fragments introgressed from *G. darwinii* Watt by mapped SSRs. QTL mapping of fiber quality was conducted in three different environments to determine QTLs for fiber quality from *G. darwinii* Watt.

Materials and methods

Plant materials and population development

The plant materials for population development and the scheme of population development are described in Fig. 1. The F_1 generation was made by crossing *G. darwinii* Watt as the male parent with Ejing1 as the female parent during the winter of 2003 in China's Hainan Province. Ejing1 is a high-yield upland cotton cultivar, notable for its high boll weight and lint percentage. Its maturity and fiber quality are also good. The F_1 generation was planted in the winter of 2004 in Hainan and backcrossed with Ejing1 as the male parent to produce the BC_1 population. The BC_1 population was backcrossed with Ejing1 and Jingchu201, respectively, to produce the BC_2 population in the summer of 2005 in Wuhan, Hubei Province. Jingchu201 is an upland cotton cultivar with high yield production and fine fiber quality. The BC_2 population was planted in the winter of 2005 in Hainan Province, and backcrossed with Ejing1, Jingchu201, $Ekang9_{Bt}$, and IR_3 to produce the BC_3 population. IR_3 is a highly insect-resistant line. $Ekang9_{Bt}$ is transformed from Ekang9 with two *Bt* genes. It is highly disease and insect resistant and has excellent yield (Nie et al. 2002; Li and Zhang 2004). From the BC_3 generation, ILs were selected according to agronomic traits such as yield and fiber quality and according to morphological markers such as epidermic villus, leaf shape, plant type, and anther color. During 2006 and 2007, the BC_3 plants were shuttle-planted in Wuhan and Hainan, where they were self-pollinated four times. Finally, a total of 105 ILs were obtained.

These 105 ILs were classed into four populations according to the materials involved in population development. They were named Pop1, Pop2, Pop3, and Pop4, which included 38, 33, 16, and 18 lines, respectively.

DNA extraction and marker analysis

The 105 ILs were planted in the experimental field of Huazhong Agricultural University (E1) in 2008. Total genomic DNA was extracted from young leaves with the modified CTAB method (Paterson et al. 1993). Markers were selected at an average of 10 cM from the interspecific linkage map of cotton constructed in our laboratory (Zhang

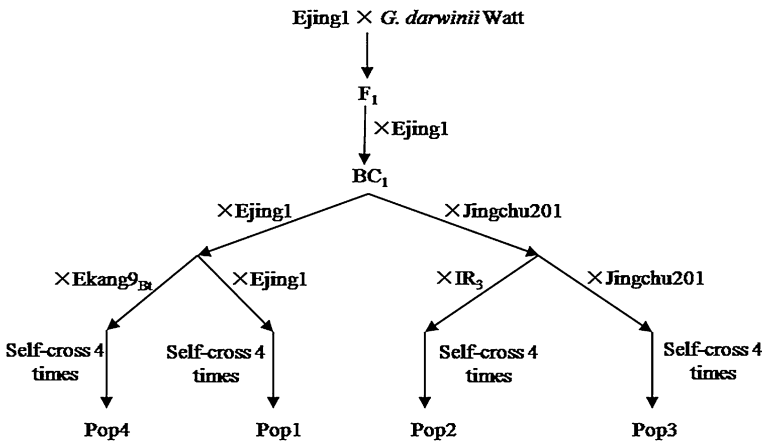


Fig. 1 Development of introgression lines

et al. 2008). A total of 310 markers were selected. Polymerase chain reaction (PCR), electrophoresis and silver staining were performed based on the protocols described by Lin et al. (2005).

Measurement of fiber quality and statistical analysis

In 2008, 20 bolls were hand harvested from the middle portion of the plants in E1 and fibers were tested for quality. In the next year, all the materials mentioned above were planted in the experimental fields in Ezhou in China's Hubei Province (E2) and Huangzhou, also in Hubei Province (E3). Cotton fibers were harvested within the same month and sent to be tested. Five indexes of fiber quality were tested, including fiber up-half mean length (UHML), fiber uniformity index (UI), fiber elongation (ELO), fiber strength (STR), and fiber micronaire value (MIC).

Basic statistical parameters of fiber quality of the 105 ILs and their upland cotton parents were calculated using Data Processing System 7.05 software (http://www.chinadps.com). These data included mean value, standard deviation, variation range, skew, kurt, and correlations of one index in different environments.

Analysis method of molecular diversity and genome structure

SSR fragments were coded "1" for presence, "0" for absence, and "–" for missing data. Genetic diversity, the average number of genotype and alleles on each locus, actual heterozygosity, and polymorphism information content (PIC) were calculated by using PowerMarker software V 3.25 (Liu and Muse 2005).

Polymorphism percentage (P), Shannon index (I), number of valid alleles (Ne), genetic differentiation (PhiPT), gene flow (Nm), and expected heterozygosity (He) were estimated using the AMOVA function of GENALEX6.2 software (Peakall and Smouse 2006).

The genetic similarity matrices were determined using SIMQUAL from the NTSYS-pc 2.10e statistical package based on Jaccard's algorithms (Adams and Rolf 2000).

According to the positions of SSR markers in the interspecific linkage map, Graphical GenoType2.0 software (Berloo 2008) was used to analyze the introgression fragments from all parents except Ejing1. The genotype of Ejing1 served as background. The number and length of introgression fragments were calculated using Microsoft Excel®.

QTL mapping method

Because fiber quality is a quantitative trait and these ILs were a biased population, the $Q + K +$ MLM model was used for association mapping of fiber quality. To evaluate the population structure for further association mapping, STRUCTURE version 2.1 (Pritchard et al. 2000a, b) software package was used to analyze the population structure of these ILs. MCMC was set to 50,000, while burn-in was set to 100,000 and K value was set to $2 \sim 10$. Each K value was repeated three times. According to the largest LnP (D) and the most stable principles of α, we could not find suitable K value for the germplasm lines. So, we calculated Q value matrix of $K = 4$ according to the population history and took the Q value as a covariance to correct false association analysis for association mapping. Kinship was observed among the lines. Because both the population structure and the kinship certainly reflected the linkage disequilibrium, the Q value matrix and kinship matrix were taken into account as part of the covariance matrix. The TASSEL package was used to analyze molecular marker data and fiber quality in three different environments individually by $Q + K +$ MLM model (Yu et al. 2006). For association mapping, the 5 % minor allele frequency-filtered SSR datasets were used. The P values of markers associated with QTLs were regulated by the method of controlling false discovery rate (Benjamini and Hochberg 1995).

Results

Polymorphism between parents

A total of 310 polymorphic loci containing 603 alleles were detected by 295 SSR primer pairs between upland cotton parents and *G. darwinii* Watt. Among them, 25 loci showed polymorphism between upland cotton parents. All the 310 loci that were found to be polymorphic between parents were located on the interspecific genetic linkage map (Zhang et al. 2008). They were found to be uniformly distributed over 44 linkage groups of the 26 chromosomes, covering the whole genome.

Genetic diversity of ILs

A total of 278 polymorphic loci between all parents were also detected in the ILs. According to the definition of polymorphic loci for a population in which all allele frequency is $\leq$99 %, 22 rare loci and 256 polymorphic loci were detected in the ILs; the allele frequency was >5 % for 119 loci (42.8 %) and <5 % for 159 loci (57.2 %).

The number of alleles on each locus was 2.109 on average. The average effective number of alleles (Ne) was 1.185. The mean number of genotypes per locus was 2.831; the average gene diversity was 0.135; and the average observed heterozygosis (Ho) was 0.054. The average PIC

value was 0.122, and the Shannon index (I) was 0.225. The overall expected heterozygosity (He) was 0.127.

The results of the analysis of molecular variance of populations revealed that the variation among populations accounted for only 5.66 % of the total variance and variation within populations accounted for 94.34 % within populations (Table 1). The pronounced homology between populations may result from common parents, and the biggest differences within populations were attributed to different parents in later backcrossing events. The genetic differentiation (PhiPT) was 0.057, and gene flow (Nm) was 4.16 ($P < 0.01$). This indicated that the division of these 105 ILs was random. There was strong gene flow between populations.

The genetic relationship among these ILs was very close. Among the 105 ILs, the Jaccard genetic similarity coefficients (data not shown) ranged from 0.562 to 0.958 with an average of 0.759. The lowest genetic similarity coefficient, 0.562, was observed between HD698 and HD722. It belonged to Pop3 and Pop4, respectively. The highest genetic similarity coefficient, 0.958, was observed between HD717 and HD718, both of which were from Pop4 and derived from the same parents.

Introgression analysis

The 310 markers covered 91.94 % of the cotton genome. As shown in Fig. 2, there were a maximum of 20 introgression loci on chromosome 12 and a minimum of 4 introgression loci on chromosomes 7 and 15, respectively. On average, there were 11 introgression loci on each chromosome. A total of 144 introgression loci were located in the At sub-genome and 134 in the Dt sub-genome. The introgression length of At and Dt sub-genomes were 2,311.0 and 2,289.3 cM, respectively. The introgression percentages of the At and Dt sub-genomes were 83.6 and 85.2 %, respectively. The longest introgression was 328.7 cM in length on chromosome 19, and the shortest was 53.4 cM on chromosome 7. The average introgression length of each chromosome was 176.9 cM. The percentage of chromosomal introgression length was highest on chromosome 24 (99.96 %) and lowest on chromosome 13 (52.23 %).

Graphical genotype analysis (GGT) showed that the distribution of the 278 introgressions accounted for 84.38 % of the cotton genome (Fig. 3). Among the 105 ILs, the average length of introgression segments derived from parents other than Ejing1 accumulated to 333.5 cM, accounting for 6.7 % of the whole genome. The length of introgression fragment from *G. darwinii* Watt was 241.6 cM per line on average, accounting for 4.8 % of the whole genome. There were 1,596 introgression segments from *G. darwinii* Watt and the number of single-line introgression segments ranged from 3 to 29 with an average of 15 (supplementary file 1). The background, Ejing1, accumulated to 4,602.9 cM, accounting for 91.8 % of the whole genome.

Among the lines that had Jingchu201 as a backcross parent, the average introgression length from Jingchu201 was 54.2 cM per line, accounting for 1.1 % of the whole genome. The average introgression lengths from IR_3 and $Ekang9_{Bt}$ were 72.1 and 1.1 cM, respectively, accounting for 1.4 and 0.02 % of the whole genome.

Table 1 AMOVA of populations (pops)

Source	*df*	SS	Est. var.	Percentage of variance	*P* value
Among pops	3	479.456	3.806	5.66	<0.001
Within pops	102	6,466.610	663.398	94.34	<0.001
Total	105	6,946.066	67.204	100	

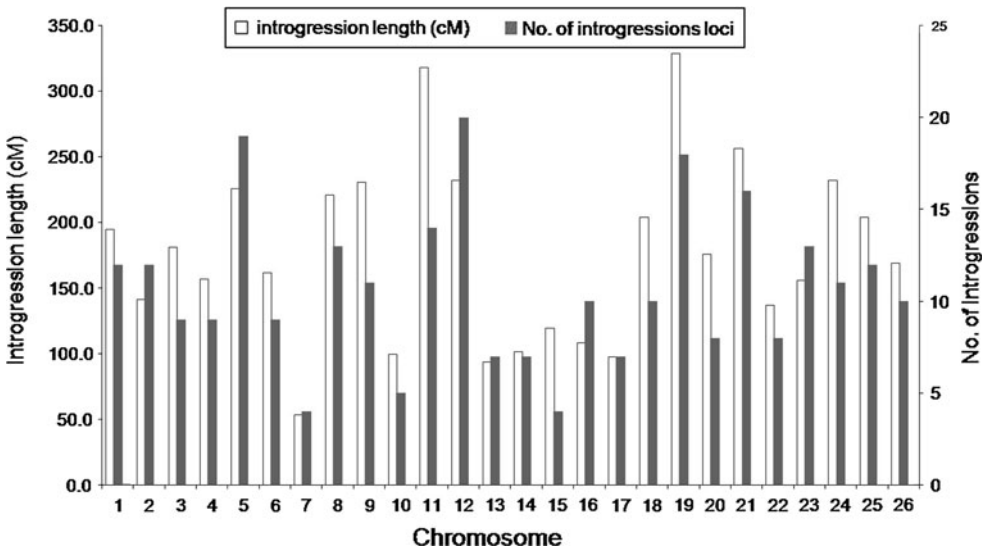


Fig. 2 Distribution of introgression length and the number of introgressions on chromosomes

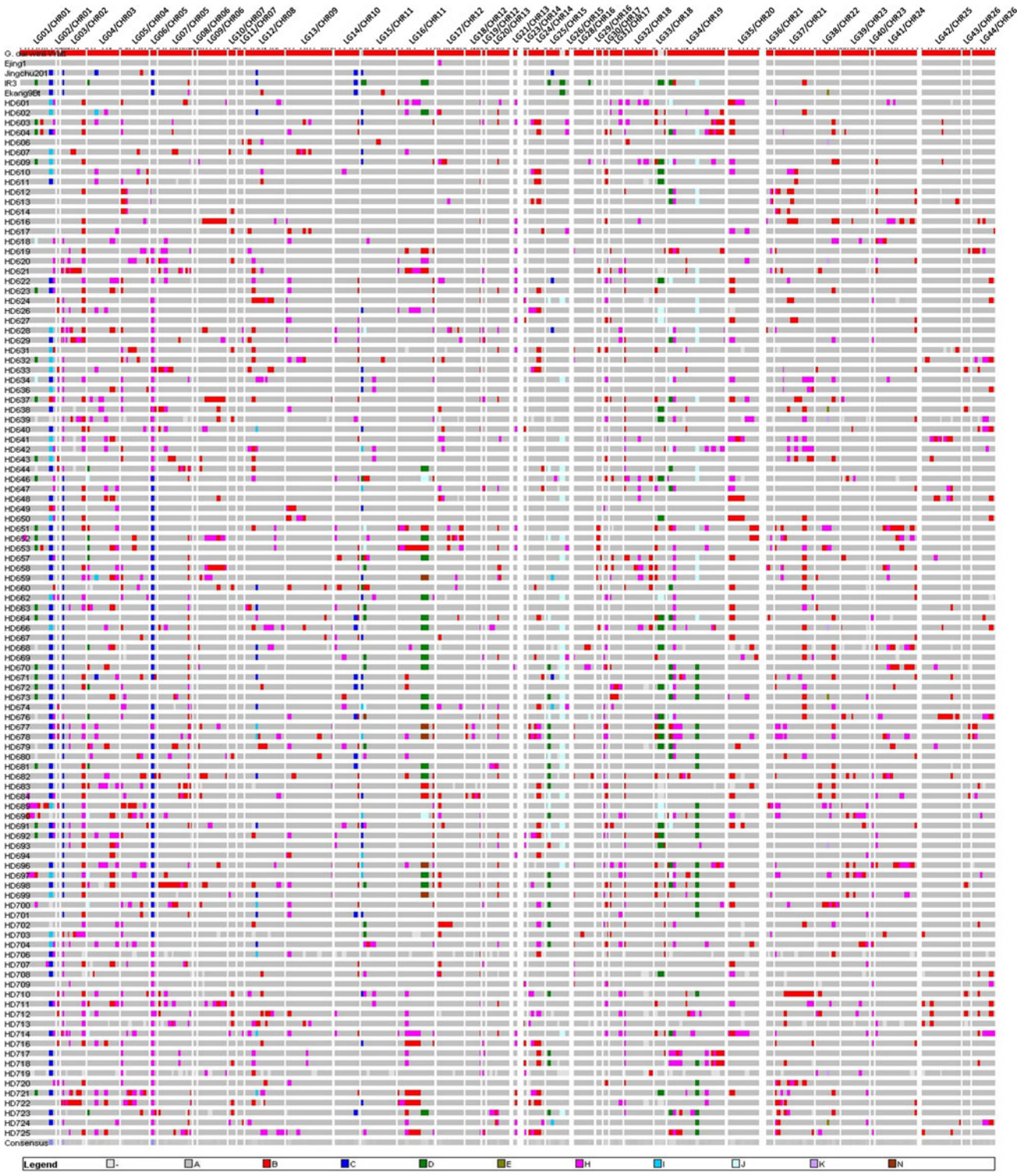


Fig. 3 Genotypes of five parents and 105 ILs. *A*, *B*, *C*, *D*, and *E* indicate the genotypes of Ejing1, *G. darwinii* Watt, Jingchu201, IR_3, and $Ekang9_{Bt}$, respectively. *H*, *I*, *J*, *K*, and *N* indicate allele combinations of *A* and *B*, *A* and *C*, *A* and *D*, *A* and *E*, and *B* and *D*, respectively. "–" indicates missing fragments

Table 2 Basic parameters of five fiber quality indexes of 105 ILs and four upland cotton parents in three environments

Indexes	Environments	Introgression lines						Ejing1	Jingchu201	IR_3	$Ekang9_{Bt}$
		Mean	SD	Range	CV (%)	Skew	Kurt				
UHML	E1	27.50	1.84	23.05–32.29	6.69	0.18	−0.21	27.11	29.07	28.97	28.90
	E2	26.91	1.66	21.65–31.50	6.17	0.06	1.09	27.14	28.28	28.21	27.00
	E3	27.71	1.45	24.48–31.57	5.23	0.32	−0.28	26.57	28.97	28.74	28.97
UI	E1	84.38	1.48	77.10–86.40	1.75	1.57	4.63	84.40	86.57	85.80	84.63
	E2	84.19	1.73	78.10–87.55	2.05	−0.87	1.46	85.24	85.75	85.73	84.58
	E3	83.52	1.39	79.95–86.60	1.66	−0.15	−0.47	82.58	84.70	84.88	84.70
MIC	E1	5.63	0.75	4.17–7.15	13.32	0.14	−1.07	5.38	5.33	5.74	5.31
	E2	5.25	0.69	3.60–6.90	13.14	0.56	−0.30	4.86	5.28	5.70	4.75
	E3	5.38	0.60	4.18–6.70	11.15	0.44	−0.50	5.61	5.25	5.70	5.25
ELO	E1	6.44	0.14	6.10–6.80	2.17	0.02	−0.29	6.30	6.43	6.47	6.47
	E2	6.25	0.09	5.95–6.50	1.44	−0.31	1.16	6.21	6.28	6.31	6.18
	E3	6.64	0.14	6.25–6.95	2.11	0.00	−0.46	6.46	6.65	6.71	6.65
STR	E1	27.50	1.92	23.10–34.40	6.98	0.50	0.96	26.45	27.97	27.20	27.33
	E2	27.85	2.30	22.50–34.10	8.26	0.16	0.17	26.78	28.23	27.29	26.85
	E3	29.11	1.62	26.05–34.50	5.57	0.42	0.25	27.20	29.25	28.86	29.25

Statistical analysis of fiber quality

The basic parameters of the five indexes in the three environments were different, but their trends were almost consistent (Table 2). The correlations of one single index in three environments were significant, almost very significant (supplementary file 2). The micronaire value, the comprehensive index of fiber fineness and maturity, is more easily influenced by environment. This is probably why the CV % of micronaire value was the largest. As shown in Table 2, the CVs % of UHML, MIC, and STR were larger, indicating that the variation of the three indexes among lines grown in the same environment was very large and that more phenotypic variation gathered in these lines. Most of the frequency distributions of these five indexes in the three environments were continuous, which revealed that the fiber quality was controlled by quantitative trait locus (supplementary file 3). MIC was not correlated with ELO and negatively correlated with the other three indexes. Apart from this, correlations between every two indexes were found to be very positively significant. This was true of all the three environments (supplementary file 4).

Association mapping of fiber quality

When the 0.05 minor allele frequency filtered SSR datasets was used, the number of markers associated with fiber quality in the three environments was 36, 40, and 28, respectively. A total of 40 SSRs were associated with these five fiber quality indexes and 21 of them were multi-detected in different environments (Fig. 4). The contribution rate for fiber quality variation was 6.31 % on average, ranging from 2.00 to 14.79 %, while the mean contribution rate of the MLM model for fiber quality variations in the three environments was 37.3, 30.7, and 26.7 % per marker, respectively.

A total of 18 markers were found to be associated with several traits in three environments, and 15 of them were detected in different environments for more than one index. The other three markers were found to be associated with only single trait in different environments. Nine markers were found to be associated with multiple traits in E1, ten in E2, and six in E3.

Fiber upper-half mean length (UHML)

A total of ten markers were associated with UHML, and the percentage of phenotypic variation explained by the markers (R^2) ranged from 2.31 to 11.93 %. Among them, four markers were detected in more than one environment.

Fiber uniformity index (UI)

There were 15 markers associated with UI in the three environments and R^2 ranged from 2.00 to 14.79 %. Four markers were detected in more than one environment.

Fiber micronaire value (MIC)

Fifteen markers were found to be associated with MIC. The R^2 of these markers ranged from 3.23 to 13.72 %. Four markers were detected in more than one environment.

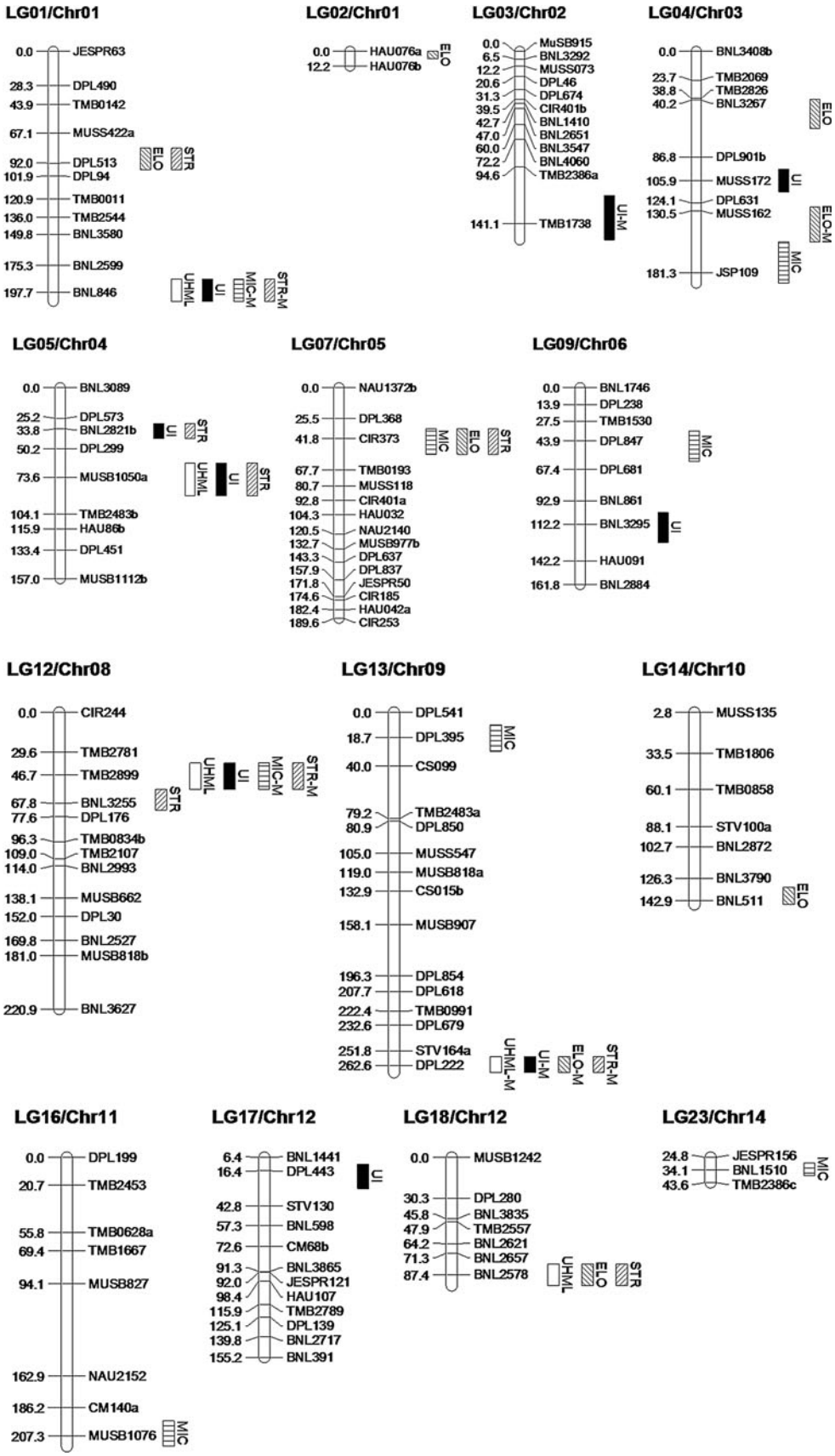
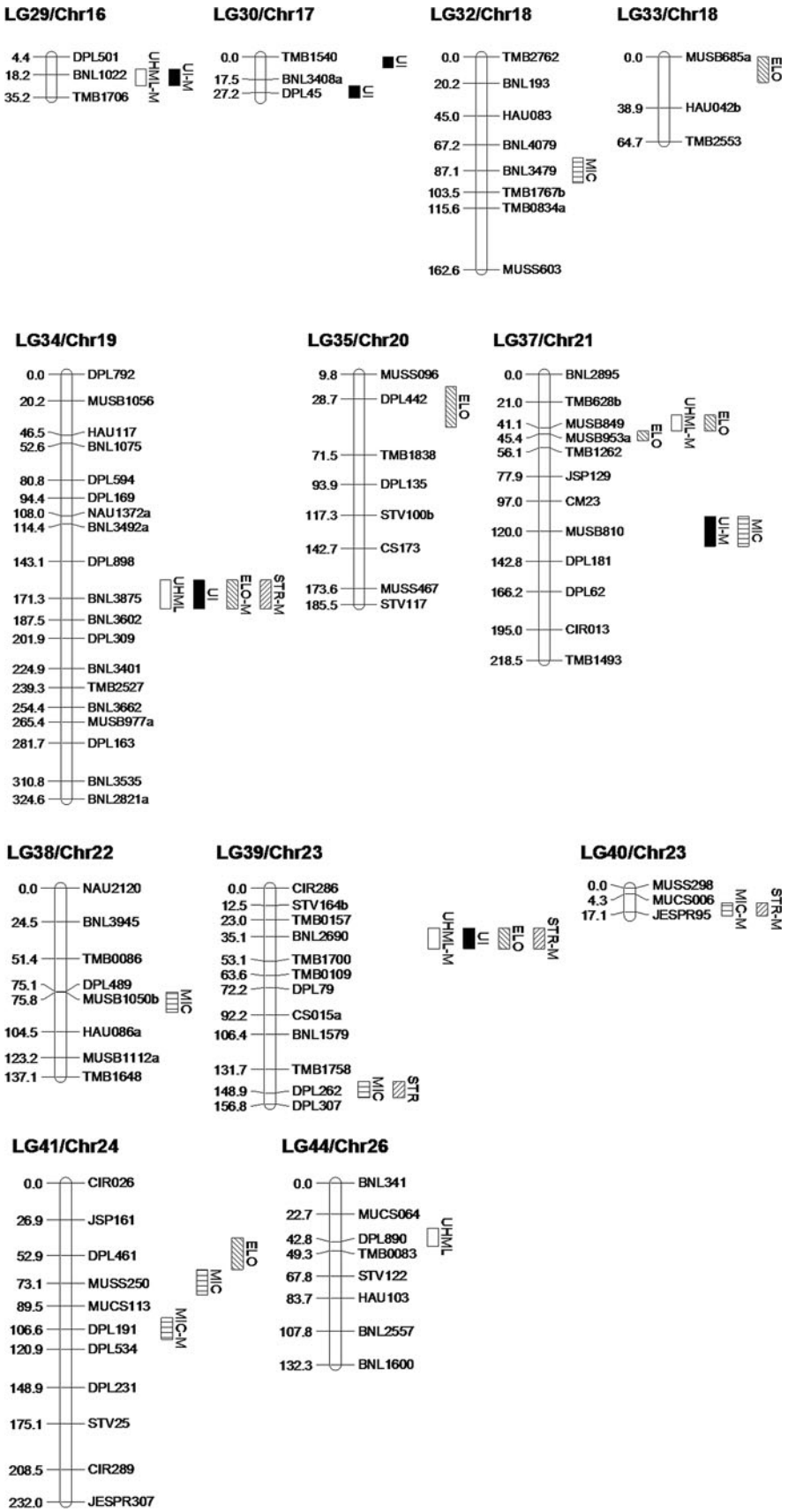


Fig. 4 The localization of QTLs for fiber quality in the genetic map which was described by Zhang et al. (2008). "Trait-M" refers to markers found to be associated with one fiber quality index in more than one environment

Fig. 4 continued



Fiber elongation (ELO)

A total of 15 markers were found to be associated with ELO in the three environments, and R^2 ranged from 4.29 to 14.43 %. Four markers were detected in more than one environment.

Fiber strength (STR)

A total of 13 markers were found to be associated with STR. R^2 ranged from 3.34 to 11.78 %. Six of these markers were detected in more than one environment.

Among the 54 EST-SSRs, 4 EST-SSRs were found to be significantly associated with fiber quality (Table 3) and two of them (MUSS162 and MUSS250) were found to be highly homologous to certain proteins. MUSS250 was found to be highly homologous to cotton integral membrane protein PIP1-3.

Discussion

The companion ILs developed in this study to previously developed ILs

Due to their utility for fine mapping and map-based cloning, ILs have become a hot spot in cotton research. Most of the previously reported cotton ILs were derived from interspecific backcrosses between *G. barbadense* (such as 3-79, hai1, and hai7124) as the donor parent and upland cotton cultivars (such as upland cotton genetic standard line TM-1, cultivar CCRI45, and early maturity cultivar CCRI36) as the receptor parent. In order to develop ILs with fewer segments and single-segment sustained lines (SSSLs) and introgressions covering the whole cotton genome, molecular marker-assisted selection was used to choose ILs during backcrossing (Saha et al. 2004; Wang et al. 2008; Yang et al. 2009). However, most of these ILs were derived from two parents, which limited their variance. It would be difficult to choose new lines capable of surpassing their parents with respect to agronomic traits. If the segments from donor parents contributed less to these traits than the receptor parents did, then the traits of these ILs might be the same as those of the receptor parents and therefore difficult to use in breeding.

This main goal of this study was to introduce the variations from *G. darwinii* Watt into upland cotton and to select ILs suitable for future breeding. As a result of complicated mating design, these ILs came to possess many exotic chromosome segments derived from their five parents. Some loci showed multiple polymorphisms, which would be beneficial to QTL mapping. The selection method adopted in this project not only improved target traits across the whole population, but also maintained high variances (Table 2). We assert that these ILs would be useful in both genetic research and breeding.

Molecular diversity and genomic structure

By exploring the genetic diversity of the ILs using mapped SSRs, we found that these ILs had pronounced genetic diversity and that polymorphic loci were abundant among all the ILs. Loci with allele frequencies >5 % accounted for only 42.8 % of the 278 polymorphic loci, indicating that allele frequencies of more than half of the loci were under 5 % and that few of the ILs had specific fragments. A comparison of the average effective number of alleles (Ne) to the total number of alleles (Na) showed a large gap. This also illustrated the uneven distribution of alleles caused by the role of artificial selection during population development.

Clustering results showed that the genetic similarity coefficients of these ILs from different populations were relatively small and they included relatively large numbers of variations. They could be used to select variations suitable for cotton breeding. AMOVA results showed that the differences among lines within the population constituted the highest proportion of the total variations. This may have been caused by backcrossing or existing gene flow (Nm) among the four populations.

All the 278 introgressions accounted for 84.38 % of the cotton genome, and the introgression length of all segments in 105 ILs accounted for 6.4 times the cotton genome. The length and number of introgressions in At sub-genome were equal to the Dt sub-genome, indicating that the variance of these ILs was genome-wide. However, 32 introgressions were not found in the ILs. The major reason for these lost introgressions is that the selection was based on traits, and the alleles that did not foster those traits were discarded during selection. In spite of loss of some donor segments, the ILs introduced not only chromosome segments from *G. darwinii* Watt, but also chromosome segments from other upland cotton cultivars, which might be valuable for the improvement of upland cotton. There is no doubt that the rich molecular diversity of these ILs would be beneficial to QTL mapping and material selection.

Markers associated with fiber quality

Association results showed that QTLs of fiber quality were distributed in 26 linkage groups and formed clusters in some intervals, such as LG01/CHR01, LG39/CHR23, and LG37/CHR21. All of these are beneficial to fine mapping.

In this study, some markers, such as TMB1738, MUSS162, and DPL191, were found to be associated with single fiber indexes in multiple environments, indicating

Table 3 Summary of EST-SSR loci associated with fiber quality identified

Marker	LG/CHR	EST ID	ORF	Putative function	Association index
MUSS162	04/03	CON_001_07795	N/A	Putative protein	ELO (E2, E3)
MUSS172	04/03	CON_001_10145	1–354	–	UI (E3)
MUSS250	41/24	CON_014_04369	643–939	Putative aquaporin PIP1-3	MIC (E2)
HAU076a	02/01	EE592816	112–507	–	ELO (E1)

that these QTLs were more stable in these environments. Markers associated with multiple traits were also frequently observed. Many markers, such as BNL3875, TMB2899, and DPL222, were associated with several fiber indexes. This phenomenon was similar to the high correlation of fiber quality indexes. However, we have no evidence to verify the relationship between them or prove whether they are pleiotropic genes or gene clusters. Fine mapping of these intervals may tell us the truth about these complicated correlations.

Two markers, BNL3255 and DPL45, were detected in specific environments. Obviously, the expression levels of the QTLs associated with these markers were clearly influenced by environmental conditions. However, these QTLs are very important and valuable, because they might be expressed differently in multiple genetic backgrounds, and their effects might be detected in different studies.

Effects of ILs on QTL mapping

The development of ILs in this study was based on phenotype selection, and the selection of agronomic traits might have many positive or negative effects on the QTL mapping. The first point of this was that the selection method in advanced generation may lead to a reduction in the diversity of the population phenotype. The range of the phenotype may narrow under the selection. However, the frequency of the phenotype of these ILs remained with normal distribution, except fiber uniformity in E1. A lower phenotype variance could indirectly improve the phenotype variance explanations (PVE) and facilitate the discovery of QTLs (Li et al. 2010). In this study, markers found to be associated with main effect QTLs ($R^2 > 5.0$ %) accounted for 48.5, 71.3, and 79.2 % in three environments. There was a strong ability to detect the main effect QTLs and more QTLs would be found in this population. Selection of unfavorable traits in advanced generation had a positive effect on QTL mapping in this population. The positive effects of selection of unfavorable traits in advanced generation were also supported by the results of studies of advanced backcrossing QTL (AB-QTL) (Septiningsih et al. 2003; Tanksley and Nelson 1996; Thomson et al. 2003; Jing et al. 2008).

However, selection also had some negative effects on QTL mapping. During the development of the population, artificial selection only affected limited genomic regions and eliminated many alleles on them that would lead to significantly lower genetic diversity at these loci compared to that at the whole genome (Barrero et al. 2011). In our study, the loss of donor segments and selection of unfavorable traits also led to omission of the QTLs or reduction effects in some genomic regions. As a result, QTL mapping in these genomic regions in this population became too difficult to be accurate or effective. To solve this problem, some ILs should be chosen from earlier generation lines to assess the ILs containing these lost introgressions. Further study can then be carried out on the improved population, which accounts for more of the cotton genome. QTL mapping in the later population could be more accurate and effective.

Comparisons of QTLs to previous studies

A total of seven markers out of all of the associated markers were consistent with previous reports (supplementary file 6). The functions predicted for MUSS250 showed that it was highly homologous to integral membrane proteins PIP1-3, a kind of aquaporins, exclusively expressed in cotton fiber cells (Liu et al. 2008). In this study, MUSS250 was found to be associated with MIC. For this reason, we speculated that the expression of GhPIP1-3 might influence MIC, but this is yet to be proved.

BNL3255 was found to be associated with fiber strength and located in the QTL intervals controlling fiber elongation and fiber strength (Shen et al. 2005). However, some contrasting results were also reported. BNL3255 was linked with the QTLs for fiber length, fiber micronaire value, fiber elongation, and fiber strength (Abdurakhmonov et al. 2008; An et al. 2010; Zhang et al. 2010; He et al. 2011). This might be a pleiotropic phenomenon or due to different alleles and growing environments.

There were two QTL markers in our results similar to those of previous studies, but we also recorded some new findings. BNL3875 and TMB2899 showed distinct levels of stability in different genetic backgrounds, different environments, and different studies. Although three other markers were significantly associated with fiber quality,

they were also found to be associated with other fiber indicators (supplementary file 6). This might be due to the different environments and materials. In brief, the results of this study are reliable and may have a high reference value for the fine mapping of fiber quality in cotton.

This study successfully developed 105 ILs for cotton genetics and breeding. These ILs were high-generation, multi-parent hybrids and they had aggregated a lot of variations. According to the strong genetic similarity between these ILs, it may be easier for us to screen and obtain certain chromosomes of near-isogenic lines (NILs) for fine mapping. Our results may make it more convenient for researchers to study segments from *G. darwinii* Watt and also be useful for molecular breeding.

Acknowledgments This work was financially supported by the National Basic Research Program (No. 2011CB109303) and Genetically Modified Organisms Breeding Major Project of China (No. 2008ZX08009-001).

References

Abdurakhmonov IY, Kohel RJ, Yu JZ (2008) Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. Genomics 92:478–487. doi:10.1016/J.YGENO.2008.07.013

Adams DJ, Rolf FJ (2000) Ecological character displacement in Plethodon: biomechanical differences found from a geometric study. Proc Natl Acad Sci USA 97:4106–4111

Ali ML, Paul L, Yu SB, Lorieux M, Eizenga GC (2010) Chromosome segment substitution lines: a powerful tool for the introgression of valuable genes from *Oryza* wild species into cultivated rice (*O. sativa*). Rice 3:218–234. doi:10.1007/s12284-010-9058-3

An CF, Jenkins JN, Wu JX, Guo YF, Mccarty JC (2010) Use of fiber and fuzz mutants to detect QTL for yield components, seed, and fiber traits of upland cotton. Euphytica 172:21–34. doi:10.1007/s10681-009-0009-2

Barrero RA, Bellgard M, Zhang XY (2011) Diverse approaches to achieving grain yield in wheat. Funct Integr Genomics 11:37–48. doi:10.1007/s10142-010-0208-x

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser C Appl Stat 57(1):289–300

Berloo RV (2008) GGT2.0: Versatile software for visualization and analysis of genetic data. J Hered 99:232–238. doi:10.1093/jhered/esm109

Brubaker CL, Wendel JF (1994) Reevaluating the origin of domesticated cotton (*Gossypium hirsutum* Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). Am J Bot 81:1309–1326

Chen L, Zhang ZS, Hu MC (2008) Genetic linkage map construction and QTL mapping for yield and fiber quality in upland cotton (*Gossypium hirsutum* L.). Acta Agr Sinica 34:1199–1205. doi:10.3724/sp.j.1006.2008.01199

Chen H, Qian N, Guo WZ, Song QP, Li BC, Deng FJ, Dong CG, Zhang TZ (2009) Using three overlapped RILs to dissect genetically clustered QTL for fiber strength on Chro.D8 in upland cotton. Theor Appl Genet 119:605–612. doi:10.1007/s00122-009-1070-x

He DH, Xing HY, Zhao JX, Li TT, Tang Y, Zeng Z (2011) Population structure inferring and association analysis of fiber quality in cultivated cotton. J Northwest A & F Univ (Nat Sci Ed) 39:103–112

Iqbal MJ, Aziz N, Saeed NA (1997) Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. Theor Appl Genet 94:139–144. doi:10.1007/s001220050392

Jing ZB, Pan DJ, Qu YY, Fan ZL, Chen Y, Chen JY, Chen F, Li C (2008) Application of AB-QTL to exploiting and utilizing the elite genes in wild rice. Mol Plant Breed 6(4):637–644

Li DG, Zhang WY (2004) Study on main economic characters heterosis of upland cotton hybrids F_1 generation. J Hubei Agric Coll 24:253–257

Li HH, Zhang LY, Wang JK (2010) Analysis and answers to frequently asked questions in quantitative trait locus mapping. Acta Agro Sinica 36(6):918–931. doi:10.3724/SP.J.1006.2010.00918

Liang ZL (1999) Genetic and breeding science of cotton hybridization. Science Press, Beijing

Lin ZX, He DH, Zhang XL (2005) Linkage map construction and mapping QTL for cotton fiber quality using SRAP, SSR and RAPD. Plant Breed 124:180–187. doi:10.1111/j.1439-0523.2004.01039.x

Liu K, Muse SV (2005) Powermarker: integrated analysis environment for genetic marker data. Bioinformatics 21:2128–2129. doi:10.1093/bioinformatics/bti282

Liu D, Tu L, Wang L (2008) Characterization and expression of plasma and tonoplast membrane aquaporins in elongating cotton fibers. Plant Cell Rep 27:1385–1394. doi:10.1007/s00299-008-0545-6

Nie YC, Zhou XR, Zhang XL (2002) Comparison on yield, pest-resistance and fiber quality of transgenic pest-resistant cotton. Plant Genet Res Sci 3:8–12

Pan JJ (1998) Breeding science for cotton. Agricultural Publishing House of China, Beijing

Park YH, Alabady MS, Ulloa M, Sickler B, Wilkins TA, Yu J, Stelly DM, Kohel RJ, Osama ME, Canrell RG (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. Mol Gen Genomics 274:428–441. doi:10.1007/s00438-005-0037-0

Paterson AH, Brubaker C, Wendel JF (1993) A rapid method for extraction of cotton (*Gossypium spp.*) genomic DNA suitable for RFLP or PCR analysis. Plant Mol Biol Rep 11:122–127. doi:10.1007/BF02670470

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295. doi:10.1111/j.1471-8286.2005.01155.x

Pritchard JK, Stephens M, Donnelly P (2000a) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b) Association mapping in structured populations. Am J Hum Genet 67:170–181

Saha S, Wu J, Jenkins JN, McCarty JC, Jr, Gutierrez OA, Stelly DM, Percy RG, Raska DA (2004) Effect of chromosome substitutions from *Gossypium barbadense* L. 3-79 into *G. hirsutum* L. TM-1 on agronomic and fiber traits. J Cotton Sci 8:162–169

Septiningsih EM, Prasetiyono J, Lubis E, Tai TH, Tjubaryat T, Moeljopawiro S, McCouch SR (2003) Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. Theor Appl Genet 107(8):1419–1432

Shen XL, Guo WZ, Zhu XF, Yuan YL, Yu ZJ, Kohel RJ, Zhang TZ (2005) Molecular mapping of QTLs for fiber qualities in three

diverse lines in upland cotton using SSR markers. Mol Breed 15:169–181. doi:10.1007/S11032-004-4731-0

Takershi E, Yoshinobu T, Yasunori N, Toshio Y, Kasumi T, Masahiro Y (2005) Construction and evaluation of chromosome segment substitution lines carrying overlapping chromosome segments of *indica* rice cultivar 'Kasalath' in a genetic background of *japonica* elite cultivar 'Kooshihikari'. Breed Sci 55:65–73

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92(2):191–203

Thomson MJ, Tai TH, McClung AM, Hinga ME, Lobos KB, Xu Y, Martinez C, McCouch SR (2003) Mapping quantitative trait loci for yield, yield components, and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. Theor Appl Genet 107(3):479–493

Wang P, Ding YZ, Lu QX, Guo WZ, Zhang TZ (2008) Development of *Gossypium barbadense* chromosome segment substitution lines in the genetic standard line TM-1 of *Gossypium hirsutum*. Chin Sci Bull 53:1512–1517. doi:10.1007/s11434-008-0220-x

Yang ZM, Li JZ, Li AG, Zhang BC, Liu GP, Li JW, Shi YZ, Liu AY, Jiang JX, Wang T, Yuan YL (2009) Developing chromosome segment substitution lines (CSSLs) in cotton (*Gossypium*) using advanced backcross and MAS. Mol Plant Breed 7:233–241

Ye GY, Liang SS, Wan JM (2010) QTL mapping of protein content in rice using single chromosome segment substitution lines. Theor Appl Genet 121:741–750. doi:10.1007/s00122-010-1345-2

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhang YX, Lin ZX, Xia QZ, Zhang XL (2008) Characteristics and analysis of simple sequence repeats in the cotton genome based on a linkage map constructed from a BC_1 population between *Gossypium hirsutum* and *G. barbadense*. Genome 51:534–546. doi:10.1139/G08-033

Zhang ZS, Hu MC, Zhang J (2009) Construction of a comprehensive PCR-based marker linkage map and QTL mapping for fiber quality traits in upland cotton (*Gossypium hirsutum* L.). Mol Breed 24:49–61. doi:10.1007/s11032-009-9271-1

Zhang K, Zhang J, Ma J, Tang SY, Liu DJ, Teng ZH, Liu DX, Zhang ZS (2010) Genetic mapping and quantitative trait locus analysis of fiber quality traits using a three-parent composite population in upland cotton (*Gossypium hirsutum* L.). Mol Breed. doi:10.1007/s11032-011-9549-y

Zhao LN, Zhou HJ, Lu LX, Liu L, Li XH, Lin YJ, Yu SB (2009) Identification of quantitative trait loci controlling rice mature seed cultivability chromosomal segment substitution lines. Plant Cell Rep 28:247–256. doi:10.1007/s00299-008-0641-7

Zhu WY, Lin J, Yang DW, Zhao L, Zhang YD, Zhu Z, Chen T, Wang CL (2009) Development of chromosome segment substitution lines derived from backcross between two sequenced rice cultivars, Indica recipient 93–11 and Japonica donor Nipponbare. Pant Mol Biol Rep 27:126–131. doi:10.1007/s11105-008-0054-3