Theor Appl Genet (2011) 123:439–454
DOI 10.1007/s00122-011-1597-5

ORIGINAL PAPER

Differential gene expression and associated QTL mapping for cotton yield based on a cDNA-AFLP transcriptome map in an immortalized F_2

Renzhong Liu · Baohua Wang · Wangzhen Guo · Liguo Wang · Tianzhen Zhang

Received: 30 May 2010 / Accepted: 5 April 2011 / Published online: 22 April 2011
© Springer-Verlag 2011

Abstract cDNA-AFLP techniques have found new applications in recent years. Currently, the methodology is used to establish differential gene expression and construct linkage maps. In the present study, a transcriptome map based on cDNA-AFLP techniques was constructed using an immortalized F_2 (IF_2) population of 171 lines. The lines were derived from intercrosses between 180 recombinant inbred lines (RILs) of the cotton hybrid Xiangzamian 2 (*Gossypium. hirsutum* L.). A total of 302 transcriptome-derived fragments (TDFs) were mapped onto 26 linkage groups that covered 2,477.06 cM in length with an average distance of 8.23 cM between two markers. Seventy-one QTL for yield and yield component traits were detected by CIM procedures based on four environments, with 13 QTL identified in at least two environments. Some TDFs co-located with yield QTL were subsequently sequenced and analyzed by online homology searches. Potential candidate genes for yield and yield component traits were found to encode proteins involved in DNA replication, transcription, translation, and biosynthesis regulation. Furthermore, genes regulating metabolic processes signal transduction, transport, and structural components of organelles were identified. Correlation analysis between expression patterns of TDFs and trait performance detected six TDFs positively correlated to both yield and yield heterosis: six TDFs positively correlated to yield, and seven TDFs to yield heterosis. These TDFs have potential for cloning the functional genes responsible for each corresponding trait and have future value in marker-assisted selection.

Abbreviations

QTL	Quantitative trait locus/loci
RIL	Recombinant inbred line
IF_2	Immortalized F_2
cDNA-AFLP	cDNA amplified fragment length polymorphism
TDFs	Transcriptome derived fragments

Introduction

The first molecular genetic map for cotton was published in 1994 (Reinisch et al. 1994) and to date almost fully resolved cotton genetic maps have been constructed. Two major types of molecular markers have been used in map construction: hybridization-based restriction fragment length polymorphisms (RFLP) (Reinisch et al. 1994; Shappley et al. 1998; Rong et al. 2004) and PCR-based markers such as random amplified polymorphic DNA (RAPD) (Kohel et al. 2001), amplified fragment length polymorphisms (AFLP) (Abdalla et al. 2001; Mei et al. 2004), sequence tagged sites (STS) (Rong et al. 2004), and simple sequence repeats (SSR) (Han et al. 2004, 2006; Frelichowski et al. 2006). These markers provide valuable information regarding levels of genetic variability, but little

Communicated by G. Bryan.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-011-1597-5) contains supplementary material, which is available to authorized users.

R. Liu · B. Wang · W. Guo · T. Zhang (✉)
National Key Laboratory of Crop Genetics and Germplasm Enhancement, Cotton Research Institute, Nanjing Agricultural University, Nanjing 210095, China
e-mail: cotton@njau.edu.cn

R. Liu · L. Wang
Shandong Cotton Research Center, Jinan 250100, China

information in terms of biological function can be obtained. Therefore, despite their usefulness, it is generally unknown if these markers target coding or non-coding regions of the genome. Recently developed candidate gene markers represented by expressed sequence tags derived SSR (EST-SSR) (Chee et al. 2004; Park et al. 2005; Guo et al. 2007), cDNA probe-based STS or RFLP markers (Rong et al. 2004), single nucleotide polymorphisms (SNP) (An et al. 2007), and sequence-related amplified polymorphisms (SRAP) (Lin et al. 2005) provide information about candidate genes with biological function. However, the utilization of these markers is limited in cotton genome mapping either by their low rates or number of polymorphisms detected (Rong et al. 2004), or by inadequate genome sequence information and highly repetitive DNA content (An et al. 2008).

cDNA-AFLP fragments or transcriptome-derived fragments (TDFs) are a type of candidate gene marker with high levels of polymorphism and reproducibility (Bachem et al. 1996). TDFs are derived from mRNA sequences of constitutively expressed genes, and therefore generally represent sequences conferring biological function (Brugmans et al. 2002). Consequently, the cDNA-AFLP technique is used extensively in differential gene expression and gene discovery in a wide number of organisms (Breyne and Zabeau 2001; Samuelian et al. 2004; Ranik et al. 2006) and applied in genome-wide expression analysis (Breyne et al. 2003) and transcriptome map construction (Brugmans et al. 2002; Pan et al. 2007; Ritter et al. 2008).

A cDNA-AFLP fragment-based genome-wide transcriptome map provides a powerful tool to identify candidate genes involved in, or controlling, various biological processes (Breyne and Zabeau 2001). Ritter et al. (2008) analyzed TDFs co-located with known QTL on an integrated potato transcriptome map and identified two resistance gene homologs. Several candidate genes in potato were identified for early tuber production in a diploid population using bulked segregant analysis (BSA) coupled with transcriptome analysis and QTL mapping (Fernández-del-Carmen et al. 2007). More recently, a cotton fiber transcriptome map was constructed and six candidate genes represented by TDFs co-segregating with QTL for fiber strength were identified (Liu et al. 2009). These functional genes were characterized as integral in plant cell wall morphogenesis or cellulose synthesis metabolism processes, and considered as candidate genes controlling cotton fiber strength.

Identification and characterization of QTL and genes controlling agronomic and fiber quality traits in cotton has been a research focus for well over a decade. Almost all mapped QTL have been identified in interspecific or intrapecific $F_2/F_{2:3}$ (Ren et al. 2002; Ulloa et al. 2005; Guo et al. 2006; He et al. 2007) and BC_1 segregating families in a single or a limited number of environments; however, mapping precision with these transient populations is limited. Immortalized mapping populations such as recombinant inbred lines (RILs) (Shen et al. 2007; Wang et al. 2007; Wu et al. 2009) and doubled haploid (DH) (Zhang et al. 2002) lines have been considered as alternatives to solve this problem. However, since alleles in these populations are homozygous, dominant effects cannot be detected. Hua et al. (2002, 2003) first proposed the immortalized F_2 (IF_2) population. Intercrosses between RILs selected by random permutations are used to construct an IF_2. The genetic composition of the population is similar to an F_2; however, replicated experiments in multiple environments can be achieved (Hua et al. 2003). Wang et al. (2007) recently constructed an immortalized F_2 population in cotton, and data from two environments were used to map fiber quality QTL. In this report, an immortalized F_2 population, Xiangzamian2 (XZM2), the most widely cultivated cotton hybrid planted both in F_1 and F_2 population before transgenic Bt hybrids were extensively released in China, derived from the intraspecific cross "CRI12 × J8891", was used to construct a cDNA-AFLP based transcriptome map and QTL for yield and yield components were identified. Concurrently, candidate genes for these important quantitative traits were identified by sequencing TDFs co-located with the corresponding QTL.

Materials and methods

Plant materials

XZM2 is a cotton hybrid with high yield potential, developed from the intraspecific cross "CRI12 × J8891". XZM2 was released in the Hunan province in 1997 (Li et al. 1997) and in the Yangtze River cotton growing region in 2001. It is the most widely cultivated cotton hybrid planted both in F_1 and F_2 before transgenic Bt hybrids were extensively released in China; CRI12 and J8891 exhibit a high level of combining ability and yield production. Furthermore, CRI12 and J8891 are inbreds derived from self-pollination. CRI12 was crossed with J8891 to generate the F_1 in 1998. A F_8 population of 180 recombinant inbred lines (RILs) was generated by a bulk-selfing technique from this cross in 2002, and the immortalized F_2 (IF_2) population was constructed by crossing between RILs randomly selected by permutations in 2003 (Wang et al. 2007).

CRI12 and J8891 (parents), F_1 and 171 IF_2 were planted in 2004 and 2005 in Jiangpu Experiment Station, Nanjing Agriculture University (JES/NAU), the Yangtze River cotton growing region in China, and in 2008 in Linqing, Shangdong, Yellow River cotton growing region in China.

Each site-year represented an individual environment (i.e., four environments). A randomized complete block design with two replications was applied in each environment. Cotton seeds were directly planted into nutritive pots in seed-bed plots in small arches covered by polyethylene mulch to control the bed temperature of 25–30°C. Seedlings exhibiting growth of three to four leaves were transplanted to the field, with a planting distance of 30 cm and a row spacing of 80 cm. Five plants in the middle of each single-row plot were tagged for trait measurement. The following yield and yield related traits were investigated: seed cotton yield (SY, g/plant), lint yield (LY, g/plant), bolls/plant (BP), boll weight (BW, g), lint percentage (LP, %), seed index (SI, g), lint index (LI, g) and fruit branch number (FBN). The IF_2 population mid-parent heterosis for SY and LY was calculated using $H_{MP} = (IF_2-(P_1 + P_2)/2) \times 100\%$, where P_1 and P_2 indicated the original parents CRI12 and J8891, respectively.

cDNA-AFLP analysis

Total RNA was extracted from topmost unfolded leaves of the parents, F_1 and IF_2 individuals at the florescence stage in one of the replications in JES/NAU in 2008, with a modified CTAB-acidic phenolic method described by Jiang and Zhang (2003). At least three leaves for each sample were pooled for RNA extraction. A spectrophotometer was used to estimate RNA concentration and quality and RNA was visualized on 1% agarose gels. Residual DNA was digested with DNaseI. First- and second-strand cDNAs were synthesized according to the M-MLV RTase cDNA Synthesis Kit protocol (TaKaRa Company, China). Double-stranded cDNA (ds-cDNA) was digested with *Mse*I/*Eco*RI followed by ligation of *Mse*I/*Eco*RI adapters with T4 DNA ligase (using the same adapters as Vos et al. 1995). cDNA-AFLP reactions were performed according to Bachem et al. (1996, 1998). Primers contained one and two selective nucleotides for preamplification and selective amplification, respectively. Selective amplification products were denatured and separated on 6% denaturing polyacrylamide (19:1) gels and visualized by silver staining.

A total of 256 *Mse*I/*Eco*RI selective primer combinations were tested (Table S1). All primer combinations that exhibited polymorphisms between the two parents and showed band type stability and consistency among 12 random samples in pilot reactions were selected to screen the immortalized F_2 populations. The presence/absence of polymorphic cDNA fragments, also referred to as transcriptome derived fragments (TDFs), were scored and a subsequent χ^2 test ($P < 0.05$) was performed to determine if the segregation ratio fit the expected 3:1 ratio.

Transcriptome map construction

JoinMap3.0 was used to conduct linkage analysis (Van Ooijen and Voorrips 2001), with a maximum recombination fraction of 0.40 and a minimum LOD score of 4.0. The Kosambi function (Kosambi 1944) was used to convert recombination frequency to map distance (centiMorgan, cM). The linkage groups were denoted as LGXX, where XX represented serial numbers. Each TDF was designated according to the primer combination, followed by the estimated fragment size in bp. MapChart 2.2 created the graphic representation of linkage groups (Voorrips 2006).

QTL analysis

Eight yield and yield component traits were examined in this study, including seed cotton yield (SY, g/plant), lint yield (LY, g/plant), bolls/plant (BP), boll weight (BW, g), lint percentage (LP), seed index (SI, g), lint index (LI, g), and fruit branch number (FBN). The mean for each individual phenotypic measurement from two replications in each environment were calculated to assess trait performance. QTL detection was conducted by the Composite Interval Mapping (CIM) procedure (Zeng 1994) using Windows QTL Cartographer 2.5 (Basten et al. 2001). The standard model (Model 6) was adopted. The window size was set at 5 cM and the walk speed 1 cM. The maximum 10 background markers were used for genetic background control. LOD threshold values were estimated by 1,000 permutations to declare significant QTL (Churchill and Doerge 1994). QTL confidence intervals (90%/95%) were set as map intervals corresponding to two or one LOD decline on either side of the peak. The genetic QTL modes were classified as described by Stuber et al. (1987). QTL nomenclature referred to McCouch et al. (1997). The designation begins with "q", followed by an abbreviation of the trait name, the linkage group, and finally the serial number.

Cloning and sequencing of cDNA-AFLP fragments

Targeted TDFs over 200 bp in length, which were co-located with QTL of yield and yield component traits, were excised from polyacrylamide gels and treated with diffusion solution containing 0.5 mM ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA and 0.1% SDS at 65°C overnight. The gel extraction mini kit (Watson Biotechnologies Inc., Shanghai, China) was then used following the manufacturer's instructions. Aliquots of the solution were used as templates to re-amplify the fragments, using PCR conditions and primer combinations identical to that used in the corresponding selective amplifications.

TDFs were cloned and sequenced using standard protocols (Sambrook et al. 1989), with slight modifications. Two microliters of the re-amplified products were ligated into pMD18-T Vectors (Takara Bio. Inc., Dalian, China), and ligates were further transformed into *E. coli* strain TOP10 cells. The cells were spread on Luria–Bertani (LB) agar plates containing ampicillin, X-gal and IPTG, and grown at 37°C overnight. Individual positive colonies were selected and grown in LB/ampicillin medium for 3–4 h at 37°C, with 300 r/min shaking. Two microliters of the culture solutions were used as templates to validate the presence of the putative inserted fragments via PCR amplifications, using M13 primer pairs and the corresponding AFLP primer combinations. GeneScript Corporation (Nanjing, China) sequenced at least two clones of each TDF in order to determine the consensus sequence.

DNAStar software was applied to remove all vector sequences. Subsequently, sequence homology searches were performed using the BLASTx algorithms (Altschul et al. 1997) in the GenBank non-redundant protein database (Benson et al. 2008) accessed through the NCBI homepage (http://www.ncbi.nlm.nih.gov/) and using the AmiGO search engine on http://www.geneontology.org/ under a significant *e* value $<10^{-4}$. For TDFs resulting in no significant homolog hits, BLASTn algorithms were used to search EST Others databases in GenBank under a significant *e* value $<10^{-8}$. ESTs with the highest similarity to each of these TDFs were selected as alternatives for a second protein database search. The correlation between functional TDF expression patterns and trait performance and/or yield heterosis were analyzed with SPSS 16.0 software using the rank-based Kendall's tau-b method (Kendall 1948). A large sample approximation for the test statistic was calculated with the mean value of trait measurement in each environment and TDF presence/absence as 1/0, respectively. The threshold for declaring significant/highly significant correlations was $P \leq 0.05$and/or 0.01.

Results

Construction of transcriptome map

Each cDNA-AFLP primer combination generated 10–30 readable bands, and most bands were monomorphic. A total of 627 polymorphic cDNA-AFLP fragments over 100 bp in length were generated from 256 primer combinations executed in the analyses, with 295 and 332 fragments from CRI12 and J8891 parents, respectively. Most polymorphic fragments (552/627) exhibited the clear presence/absence of dominant band types. A small portion of fragments (75/627) showed polymorphisms with expression levels indicated by a clear cut-off in band intensity, which made it possible to distinguish the two types of genotypes: high-intensity bands were scored as present and low-intensity bands as absent. Of the total 627 polymorphic fragments generated in the present study, 105 (16.75%) showed significant distortions, consistent with 14.4% reported by Fan et al. (2008). All polymorphic TDFs were used to construct the transcriptome map except for the 54 severely distorted TDFs.

JoinMap 3.0 was used to construct a transcriptmome map with 26 linkage groups at a threshold of 4.0 (Table S2). The total length of the map was 2747.01 cM, and the average individual linkage group length was 95.27 cM. Among the 573 TDF markers processed, 302 (52.70%) were mapped with an average interval of 8.23 cM between two adjacent markers. Distances between two adjacent TDFs were within 5.0 cM for approximately 45.0% of all markers. Individual linkage groups contained between three and 23 markers with an average 12 markers per linkage group. All cDNA-AFLP markers were distributed uniformly among linkage groups, with no significant clusters (Fig. 1). Obvious clusters were not detected in each linkage group, which is incongruent with most genetic maps constructed by DNA markers. Seventeen larger linkage groups with lengths ranging from 84.17 to 151.87 cM contained 85.8% (259/302) of the markers, with two large gaps over 30 cM on LG09 and LG14.

QTL analysis of yield and yield component traits based on transcriptome map

The mean value of traits from both mapping parents and the IF_2 based on two replications in each environment was calculated. Student's *t* test detected significant differences between parents for all traits with the exception of LP, SI, and FBN (data not shown). All yield and yield components in the IF_2 expressed transgressive segregation in both directions, suggesting polygenic inheritance (Table 1; Fig. 1).

Seed cotton yield

Thirteen QTL were detected for SY, which explained between 6.33 and 35.99% of the phenotypic variation (PV) and LOD scores ranged from 2.52 to 4.83. Three dominant or overdominant QTL (*qSY-LG09-1*, *qSY-LG12-2* and *qSY-LG16-1*) were detected in two environments. Another QTL *qSY-LG08-1* showed additive or overdominance in two environments. The other nine QTL were detected only once, seven of which exhibited partial to complete dominant effects and two showed overdominant effects. Alleles from J8891 contributed to an increase in SY for seven QTL, and alleles from CRI12 had positive effects for six QTL. A major QTL *qSY-LG12-2* was associated with TDFs E4M16_100, E3M12_300 and E12M12_400.

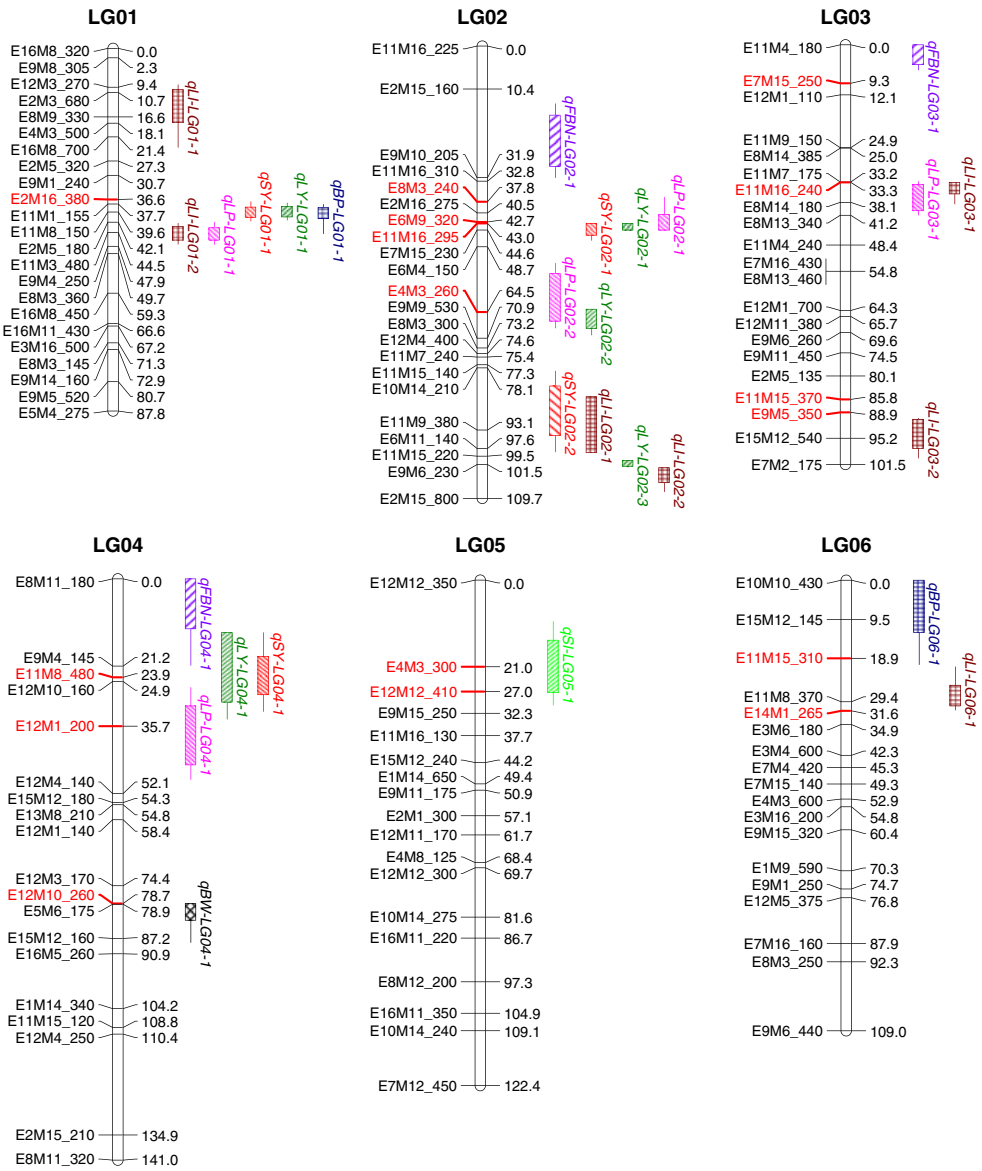


Fig. 1 Transcriptome map and QTL position. The map of the immortalized F_2 population derived from crosses between RILs of XZM2 (CRI12 × J8891F_1) was constructed by JoinMap 3.0 with the Kosambi function and a LOD ratio of 3.0. Significant QTL for yield and yield component traits were indicated by *bars* and *whiskers*, which are 1- and 2-LOD likelihood intervals, respectively. Sequenced cDNA-AFLP fragments co-located with QTL were shown in *brick red color*

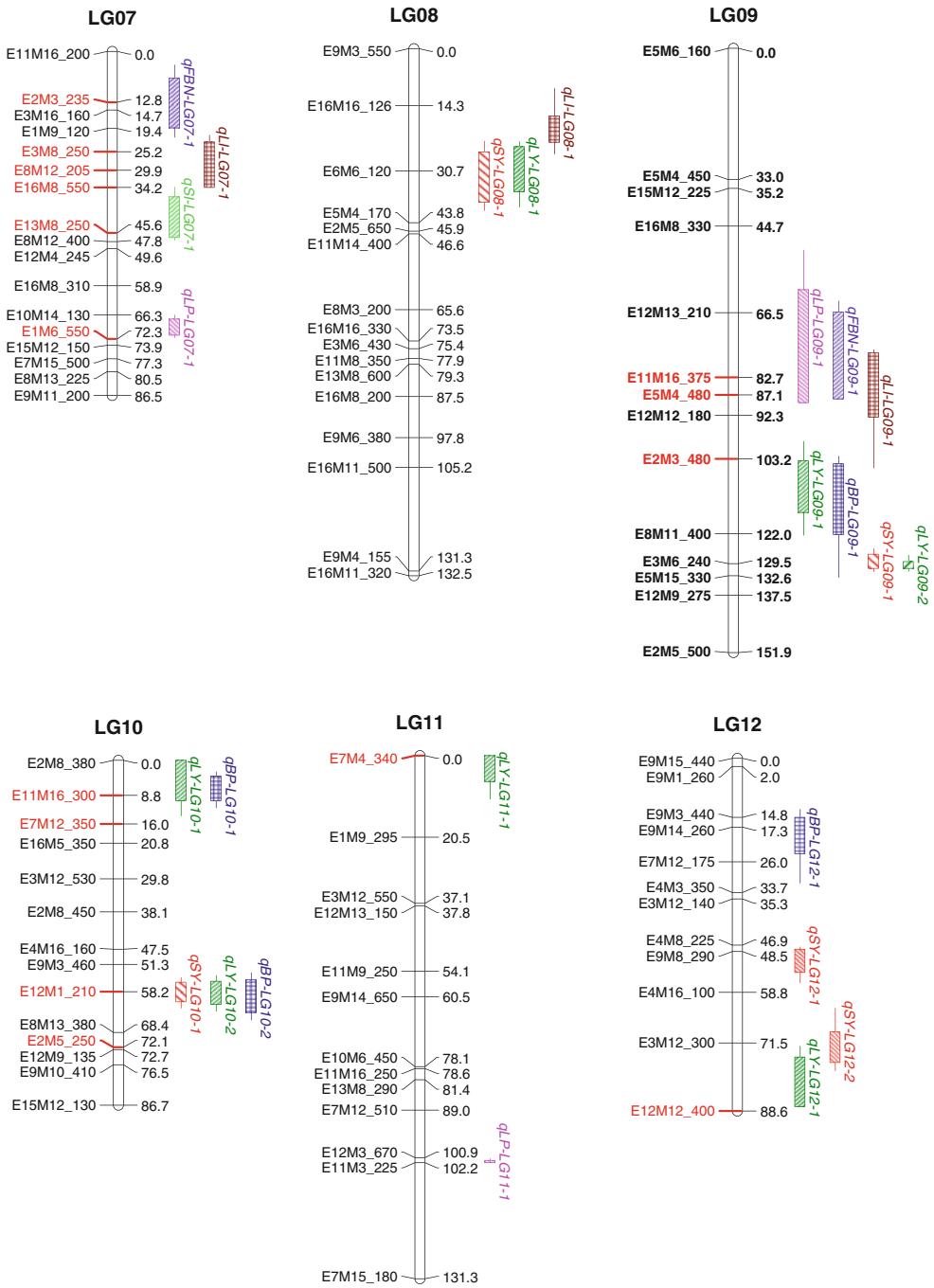
Lint yield

Fourteen QTL were detected for LY, which explained 6.93% of the PV in *qLY-LG09-2* to 35.05% in *qLY-LG12-1*. LOD scores ranged from 2.5 to 4.68. Two QTL (*qLY-LG01-1* and *qLY-LG08-1*) were detected in two environments, both exhibiting at least partial dominance. The other 12 QTL were identified in one environment, with four showing partial to complete dominance and eight overdominance. Alleles from J8891 contributed to an increase in LY for six QTL, and alleles from CRI12 had positive effects for eight QTL. The major QTL *qLY-LG12-1* was co-located with TDFs E3M12_300 and E12M12_400 on LG12.

Bolls per plant

Twelve QTL were detected for BP with LOD scores ranging from 2.52 to 5.15; and one QTL was detected twice exhibiting dominance and overdominance. Four partial to complete dominant and seven overdominant QTL were detected in a single environment. The PV explained by these QTL ranged from 10.38% for *qBP-LG01-1* to 40.37% for *qBP-LG22-1*. CRI12 alleles were associated with increased BP for five QTL, and J8891 contributed positive alleles for seven QTL. Major QTL *qBP-LG22-1* was associated with E16M8_250 and E7M15_700, and *qBP-LG10-1* with E2M8_380 and E11M16_300.

Fig. 1 continued



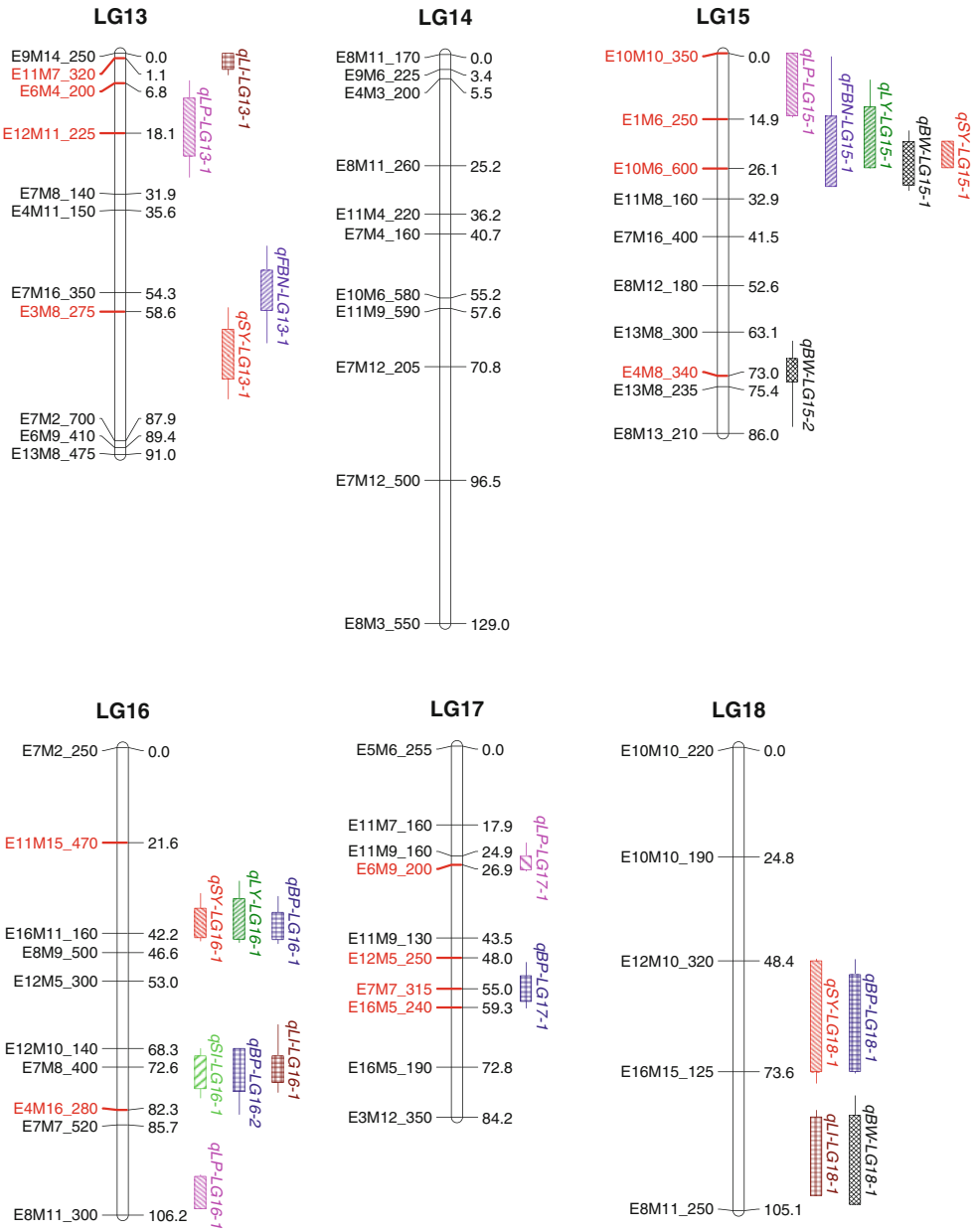
Boll weight

Four QTL for BW were detected in a single environment with LOD scores between 2.55 and 3.38. These QTL explained from 7.55% for *qBW-LG15-2* to 47.88% for *qBW-LG18-1* of the total phenotypic variation. *qBW-LG15-2* showed partial dominance and overdominance was exhibited by the three other QTL. J8891 alleles were responsible for an increase in BW for three QTL, and CRI12 alleles were associated with increased BW for one QTL. The major QTL *qBW-LG18-1* explained the highest percentage of PV and was co-located with TDFs E16M15_125 and E8M11_250 on LG18.

Lint percentage

Fourteen LP QTL were identified, which explained 7.59–39.80% of the PV. LOD scores ranged from 2.53 to 4.45. The overdominant QTL *qLP-LG07-1* was detected in three environments, and the partial dominant *qLP-LG15-1*

Fig. 1 continued



and additive *qLP-LG19-1* QTL were detected twice. The remaining 11 QTL were detected in a single environment, of which three showed partial to complete dominant effects and eight showed overdominant effects. CRI12 alleles were associated with increased LP for six QTL and J8891 alleles for the remaining eight QTL. The major QTL *qLP-LG16-1* was co-located with TDFs E7M7_520 and E8M11_300, and *qLP-LG17-1* with E11M9_160 and E6M9_200.

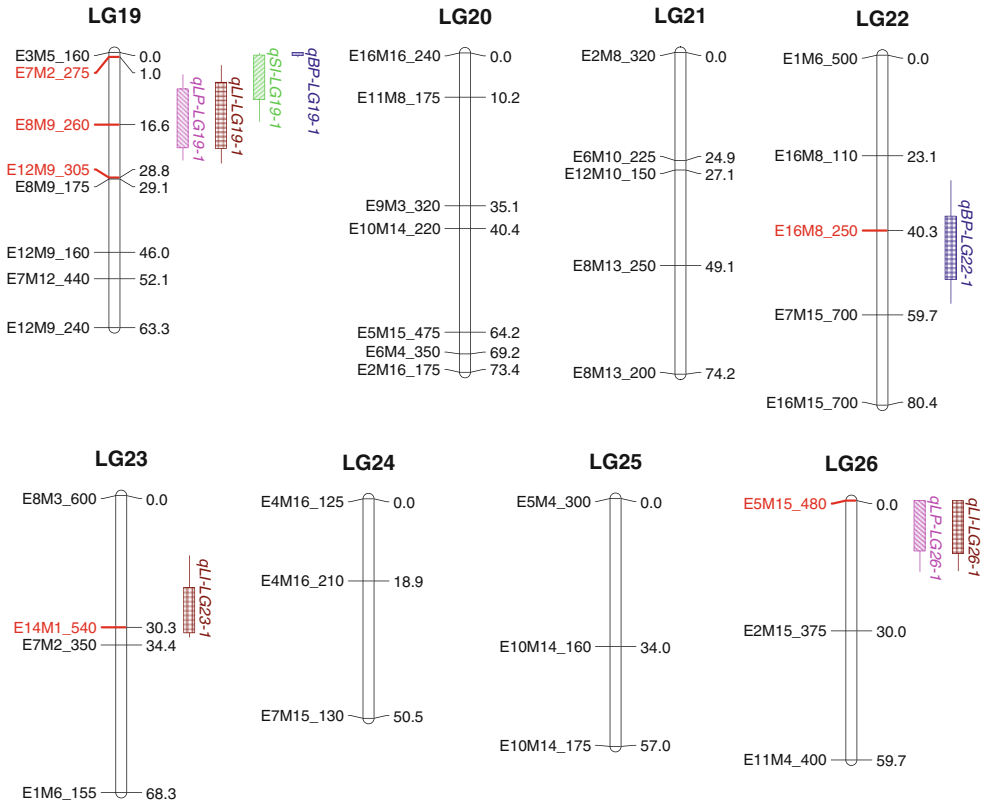
Seed index

Four QTL were detected for SI with LOD scores ranging from 2.58 to 3.05. The phenotypic variation explained by these QTL ranged from 10.35 to 22.51%. QTL *qSI-LG16-1* was detected in two environments. Alleles derived from CRI12 were responsible for the SI increase in *qSI-LG19-1*, and J8891 alleles were positive at *qSI-LG05-1*, *qSI-LG07-1,* and *qSI-LG16-1*.

Lint index

Fifteen QTL were identified for LI with LOD scores between 2.54 and 4.63. The QTL explained 8.35–36.69% of the phenotypic variation. QTL *qLI-LG02-1* and *qLI-LG06-1* were detected in two environments. Alleles derived from CRI12 were associated with increased LI for six QTL. J8891 alleles provided increased LI for the additional nine QTL. Seven QTL showed partial to

Fig. 1 continued



complete dominance, and eight QTL showed overdominance. The major QTL *qLI-LG18-1* was co-located with TDFs E8M9_260 and E12M9_305, and a second major QTL *qLI-LG08-1* was associated with E16M16_126 and E6M6_120.

Fruit branch number

Seven QTL were detected in a single environment for FBN with the LOD scores ranged from 2.56 to 5.80. And 9.79 to 36.79% of the phenotypic variation was explained by these QTL. Three QTL showed partial dominance and four showed overdominance. Increases in FBN were observed in three QTL associated with CRI12 alleles and four QTL associated with J8891 alleles. The major QTL *qFBN-LG15-1* was co-located with TDFs E1M6_250 and E10M6_600 on LG15.

Putative candidate genes revealed by homology analysis of associated TDFs

Of the total 51 TDFs sequenced, 44 with represented genes biological function (Table S3) were associated with 50 QTL for yield and yield components. Among them, nine, five, thirteen, and five TDFs representing genes were related to transcription regulation, signal transduction pathways, biosynthesis and metabolism processes, and ion, protein, or phospholipid transport, respectively. For example, three overdominant QTL *qBP-LG16-1*, *qSY-LG16-1,* and *qLY-LG16-1*, each with significant additive and dominant effects were flanked by TDF E11M15_470, which encodes an AP2 domain-containing transcription factor involved in modulating expression of genes responsive to biotic and abiotic stress diseases and genes related to the cell life-cycle, growth, and development. The overdominant QTL *qBP-LG22-1* was associated with E16M8_250, a fragment encoding a NAC domain containing protein involved in multicellular organismal development, shoot development, xylem histogenesis, and programmed cell death. The gene represented by the E3M8_250 fragment on LG07 encodes Late Embryogenesis Abundance (LEA) 14. LEA proteins, when expressed under conditions of desiccation, cold, osmotic stress, and heat are thought to be associated with abiotic stress tolerance.

Relationship between gene differential expression and yield heterosis

A total of 37 TDFs were significantly correlated with trait performance and yield heterosis (Table 2). Among them, 13 TDFs originated from J8891 and 24 from CRI12, each with correlation coefficients either in a positive or negative direction. Five TDFs were positively correlated with the

Table 1 QTL for yield and yield components detected by CIM in the IF_2 population across four environments

Trait[a]	QTL	Env.[b]	Flanking markers	Position (cM)	LOD	A[c]	D[d]	R^2 (%)	D/|A|[e]	GA[f]
SY	*qSY-LG01-1*	E1	E11M1_155-E11M8_150	39.6	2.64	5.0309	2.2226	6.33	0.44	PD
	qSY-LG02-1	E2	E11M16_295-E7M15_230	44.0	3.46	4.7084	−12.1686	19.91	−2.58	OD
	qSY-LG02-2	E3	E10M14_210-E11M9_380	85.1	2.94	14.1606	−12.0494	22.69	−0.85	D
	qSY-LG04-1	E3	E9M4_145-E11M8_480	23.2	3.35	8.3200	−3.3335	14.12	−0.40	PD
	qSY-LG08-1	E1	E6M6_120-E5M4_170	32.7	3.11	−6.2809	0.108	9.54	0.02	A
		E4	E6M6_120-E5M4_170	30.7	2.87	−2.7181	5.3969	7.58	1.99	OD
	qSY-LG09-1	E3	E3M6_240-E5M15_330	129.5	2.52	−2.4165	−5.7233	7.11	−2.37	OD
		E2	E8M11_400-E3M6_240	124.2	2.5	−7.9069	−7.4472	25.51	−0.94	D
	qSY-LG10-1	E3	E12M1_210-E8M13_380	58.2	4.4	1.3916	−10.5286	15.54	−7.57	OD
	qSY-LG12-1	E2	E9M8_290-E4M16_100	48.5	3.06	−2.8430	−2.7785	6.83	−0.98	D
	qSY-LG12-2	E4	E4M16_100-E3M12_300	72.8	3.61	9.0254	−10.4299	35.99	−1.16	D
		E1	E3M12_300-E12M12_400	82.5	2.85	5.8602	−12.552	34.46	−2.14	OD
	qSY-LG13-1	E3	E3M8_275-E7M2_700	68.6	2.57	−9.8747	−5.7489	23.71	−0.58	PD
	qSY-LG15-1	E2	E1M6_250-E10M6_600	20.9	3.38	−6.0926	2.9398	19.88	0.48	PD
	qSY-LG16-1	E4	E16M11_160-E8M9_500	42.2	4.83	−4.2756	13.0038	29.62	3.04	OD
		E2	E8M9_500-E12M5_300	48.0	2.6	−7.5704	8.6945	20.68	1.15	D
	qSY-LG18-1	E1	E12M10_320-E16M15_125	65.4	3.66	−9.4965	−5.2875	27.78	−0.56	PD
LY	*qLY-LG01-1*	E1	E11M1_155-E11M8_150	39.6	3.04	2.2806	0.8950	7.10	0.39	PD
		E2	E9M10_205-E11M16_310	31.9	3.04	5.8808	−6.007	21.86	−1.02	D
	qLY-LG02-1	E2	E11M16_295-E7M15_230	44.0	4.68	4.5126	−7.6475	25.68	−1.69	OD
	qLY-LG02-2	E2	E4M3_260-E9M9_530	64.5	3.59	5.8298	−5.0570	24.47	−0.87	D
	qLY-LG02-3	E2	E9M6_230-E2M15_800	101.5	2.92	4.6128	−5.4443	23.72	−1.18	D
	qLY-LG04-1	E3	E9M4_145-E11M8_480	23.2	2.5	3.3440	−0.9588	10.16	−0.29	PD
	qLY-LG08-1	E1	E6M6_120-E5M4_170	32.7	3.8	−3.0213	0.7802	12.23	0.26	PD
		E4	E6M6_120-E5M4_170	30.7	2.55	−1.2406	2.0518	6.18	1.65	OD
	qLY-LG09-1	E3	E2M3_480-E8M11_400	110.2	2.55	0.5021	−5.5911	21.71	−11.14	OD
	qLY-LG09-2	E3	E3M6_240-E5M15_330	129.5	2.51	−1.2065	2.5272	6.93	2.09	OD
	qLY-LG10-1	E3	E2M8_380-E11M16_300	5.0	2.96	−4.6204	−2.1288	22.09	−0.46	PD
	qLY-LG10-2	E3	E12M1_210-E8M13_380	58.2	3.64	−0.0881	−4.8332	15.01	−54.83	OD
	qLY-LG11-1	E4	E7M4_340-E1M9_295	3.0	2.57	1.4177	−2.9190	9.27	−2.06	OD
	qLY-LG12-1	E4	E3M12_300-E12M12_400	82.5	2.84	2.6781	−5.6753	35.05	−2.12	OD
	qLY-LG15-1	E2	E1M6_250-E10M6_600	16.9	4.62	−3.0922	1.8253	27.63	0.59	PD
	qLY-LG16-1	E4	E16M11_160-E8M9_500	42.2	4.15	−1.8023	5.2286	24.05	2.90	OD
BP	*qBP-LG01-1*	E1	E11M1_155-E11M8_150	39.6	4.12	1.5360	0.6665	10.38	0.43	PD
	qBP-LG06-1	E1	E10M10_430-E15M12_145	0.0	3.45	−1.9749	2.6173	30.30	1.33	OD
	qBP-LG09-1	E2	E2M3_480-E8M11_400	106.2	2.52	−1.2723	−1.2037	23.40	−0.95	D
	qBP-LG10-1	E4	E2M8_380-E11M16_300	6.0	3.58	−0.3095	2.3616	33.48	7.63	OD
		E3	E11M16_300-E7M12_350	12.8	4.07	−1.8090	1.5898	29.26	0.88	D
	qBP-LG10-2	E3	E12M1_210-E8M13_380	59.2	3.3	−0.4858	−2.1367	19.95	−4.40	OD
	qBP-LG12-1	E4	E9M3_440-E9M14_260	16.8	2.99	1.5348	−1.0122	17.97	−0.66	PD
	qBP-LG16-1	E4	E16M11_160-E8M9_500	42.2	5.15	−0.8764	−2.0913	28.95	−2.39	OD
	qBP-LG16-2	E4	E7M8_400-E4M16_280	72.6	3.24	−1.0484	1.6775	20.09	1.60	OD
	qBP-LG17-1	E1	E12M5_250-E7M7_315	55.0	2.62	1.1851	2.4057	16.26	2.03	OD
	qBP-LG18-1	E1	E12M10_320-E16M15_125	66.4	3.57	−2.2753	−0.9861	26.83	−0.43	PD
	qBP-LG19-1	E1	E3M5_160-E7M2_275	0.0	2.64	1.5547	−2.7793	26.38	−1.79	OD
	qBP-LG22-1	E3	E16M8_250-E7M15_700	45.3	5.01	1.4559	2.3146	40.37	1.59	OD

Table 1 continued

Trait[a]	QTL	Env.[b]	Flanking markers	Position (cM)	LOD	A[c]	D[d]	R^2 (%)	D/|A|[e]	GA[f]
BW	*qBW-LG04-1*	E1	E5M6_175-E15M12_160	79.0	2.76	0.0070	−0.1517	7.70	−21.70	OD
	qBW-LG15-1	E2	E10M6_600-E11M8_160	27.1	3.38	−0.0259	0.3387	16.26	13.08	OD
	qBW-LG15-2	E1	E4M8_340-E13M8_235	73.0	2.82	−0.1036	0.0462	7.55	0.45	PD
	qBW-LG18-1	E4	E16M15_125-E8M11_250	91.6	2.55	−0.0516	−0.5044	47.88	−9.77	OD
LP	*qLP-LG01-1*	E3	E2M5_180-E11M3_480	44.1	2.54	0.4873	0.4130	7.59	0.85	D
	qLP-LG02-1	E2	E11M16_295-E7M15_230	44.0	4.45	2.7799	−3.3860	22.36	−1.22	OD
	qLP-LG02-2	E4	E6M4_150-E4M3_260	60.8	4.25	−1.2782	0.3967	21.74	0.31	PD
	qLP-LG03-1	E3	E11M16_240-E8M14_180	38.1	3.58	0.4082	0.8320	11.51	2.04	OD
	qLP-LG04-1	E3	E12M1_200-E12M4_140	37.7	2.96	−0.0932	0.8352	9.80	8.96	OD
	qLP-LG07-1	E2	E10M14_130-E1M6_550	69.3	4.62	−0.7529	1.0897	17.23	1.45	OD
		E3	E10M14_130-E1M6_550	70.3	2.52	−0.4281	0.8093	10.19	1.89	OD
		E1	E16M8_310-E10M14_130	62.9	3.5	−0.5026	0.8384	10.78	1.67	OD
	qLP-LG09-1	E1	E12M13_210-E11M16_375	79.6	2.53	−1.0121	−0.7957	20.76	−0.79	PD
	qLP-LG11-1	E1	E12M3_670-E11M3_225	101.9	3.07	0.3856	1.5177	26.80	3.94	OD
	qLP-LG13-1	E2	E6M4_200-E12M11_225	15.8	2.72	−0.3860	0.8940	10.64	2.32	OD
	qLP-LG15-1	E4	E10M10_350-E1M6_250	8.0	4	−0.7764	0.2433	17.83	0.31	PD
		E3	E1M6_250-E10M6_600	17.1	2.77	−0.7385	0.1832	12.44	0.25	PD
	qLP-LG16-1	E3	E7M7_520-E8M11_300	100.7	3.07	0.6357	1.7869	39.80	2.81	OD
	qLP-LG17-1	E3	E11M9_160-E6M9_200	26.0	2.67	0.5545	1.7392	35.23	3.14	OD
	qLP-LG19-1	E1	E8M9_260-E12M9_305	16.6	4.93	−1.1172	0.2143	22.30	0.19	A
		E3	E7M2_275-E8M9_260	8.0	2.86	−0.8387	0.0600	13.76	0.07	A
	qLP-LG26-1	E3	E5M15_480-E2M15_375	0.0	2.89	−0.0070	−1.1319	16.75	161.03	OD
SI	*qSI-LG05-1*	E2	E12M12_350-E4M3_300	21.0	2.58	−0.3641	−0.2446	22.51	−0.67	PD
	qSI-LG07-1	E4	E16M8_550-E13M8_250	42.2	3.05	−0.2016	−0.2094	10.35	−1.04	D
	qSI-LG16-1	E2	E7M8_400-E4M16_280	75.6	2.7	−0.2039	0.4172	19.87	2.05	OD
		E3	E4M16_280-E7M7_520	84.7	3.16	−0.3751	0.2527	23.72	0.67	PD
	qSI-LG19-1	E3	E7M2_275-E8M9_260	4.0	2.67	0.0858	0.3370	11.98	3.93	OD
LI	*qLI-LG01-1*	E1	E2M3_680-E8M9_330	13.7	3.18	1.1564	−0.9367	28.91	−0.81	D
	qLI-LG01-2	E2	E2M5_180-E11M3_480	44.1	3.17	0.2393	0.0500	9.81	0.21	PD
	qLI-LG02-1	E1	E10M14_210-E11M9_380	86.1	8.28	−1.3631	1.1639	29.89	0.85	D
		E2	E10M14_210-E11M9_380	86.1	3.28	−0.3802	0.5871	20.55	1.54	OD
	qLI-LG03-1	E3	E11M16_240-E8M14_180	34.3	4.63	0.2696	0.1307	14.17	0.48	PD
	qLI-LG03-2	E1	E9M5_350-E15M12_540	91.0	3.15	1.2570	−1.3159	27.31	−1.05	D
	qLI-LG06-1	E1	E11M8_370-E14M1_265	29.4	3.57	−0.4359	0.6753	26.79	1.55	OD
		E2	E11M8_370-E14M1_265	30.4	3.09	−0.1178	0.2728	11.55	2.32	OD
	qLI-LG07-1	E2	E8M12_205-E16M8_550	32.9	2.54	−0.1409	0.1873	8.35	1.33	OD
	qLI-LG08-1	E2	E16M16_126-E6M6_120	19.4	3.43	−0.2259	0.4744	32.34	2.10	OD
	qLI-LG09-1	E1	E11M16_375-E5M4_480	82.7	3.59	−0.4168	−0.3335	24.14	−0.80	PD
	qLI-LG13-1	E4	E9M14_250-E11M7_320	1.0	2.72	0.1152	−0.1636	7.29	−1.42	OD
	qLI-LG16-1	E4	E12M10_140-E7M8_400	72.3	3.09	−0.1345	0.3754	16.94	2.79	OD
	qLI-LG18-1	E4	E16M15_125-E8M11_250	91.6	2.97	0.0183	−0.5363	36.69	−29.38	OD
	qLI-LG19-1	E1	E8M9_260-E12M9_305	16.6	3.09	−0.2652	−0.0595	9.43	−0.22	PD
	qLI-LG23-1	E1	E8M3_600-E14M1_540	30.3	3.27	−0.0758	0.4059	15.27	5.35	OD
	qLI-LG26-1	E3	E5M15_480-E2M15_375	0.0	2.62	−0.0785	−0.5077	29.53	−6.47	OD

Table 1 continued

Trait[a]	QTL	Env.[b]	Flanking markers	Position (cM)	LOD	A^c	D^d	R^2 (%)	D/|A|[e]	GA[f]
FBN	*qFBN-LG02-1*	E1	E2M15_160-E9M10_205	24.4	2.67	−0.8272	0.5989	28.94	0.72	PD
	qFBN-LG03-1	E3	E11M4_180-E7M15_250	1.0	3.24	−0.3634	0.6707	9.79	1.85	OD
	qFBN-LG04-1	E3	E8M11_180-E9M4_145	4.0	5.23	1.0803	−0.3721	29.07	−0.34	PD
	qFBN-LG07-1	E3	E11M16_200-E2M3_235	12.0	3.32	−0.6901	1.0146	17.94	1.47	OD
	qFBN-LG09-1	E4	E12M13_210-E11M16_375	78.6	2.56	0.4576	0.3669	21.92	0.80	PD
	qFBN-LG13-1	E3	E7M16_350-E3M8_275	54.3	2.8	−0.7396	−0.9440	20.24	−1.28	OD
	qFBN-LG15-1	E1	E1M6_250-E10M6_600	26.1	5.8	0.6628	0.8775	36.79	1.32	OD

[a] *SY* seed-cotton yield (g/plant); *LY* lint yield (g/plant); *BP* bolls/plant; *BW* boll weight (g); *LP* lint percentage (%); *SI* seed index (g); *LI* lint index (g); *FBN* fruit branch number

[b] E1, E2, E3, and E4 represent the environment Jiangpu 2004, 2005, 2008, and Linqing 2008, respectively

[c] *A* additive effects. Positive/negative signs indicate alleles from CRI12/J8891 increase trait value, respectively

[d] *D* Dominant effects. Positive/negative signs indicate the effects increasing/decreasing trait value over the population mean

[e] *D*/|*A*| dominance/|additive|

[f] *GA* gene action; *A* additive (|d/a| = 0–0.2); *PD* partial dominance (|d/a| = 0.21–0.80); *D* dominance (|d/a| = 0.81–1.20); *OD* overdominance (|d/a| > 1.2)

seedcotton and lint yields as well as yield heterosis, including E4M3_300, E1M6_250, E4M8_340, E10M6_600, and E4M16_280 encoding the CAP-binding protein, mitogen-activated protein kinase, GATA transcription factor, translation initiation factor EIF4A1, and beta-ketoacyl-CoA synthase, respectively. Six TDFs were positively correlated with both SY and LY, and included TDFs E7M15_250, E11M15_470, E12M9_305, and E16M8_250, respectively, encoding glycerol-3-phosphate acyltransferase, AP2 domain-containing transcription factor, ribonuclease II-like protein, NAC domain-containing protein and TDFs E11M16_240 and E6M4_200 encoding proteins of unknown functions. The fragment E2M3_235, encoding a proteasome subunit, was positively correlated with LY, BP, and yield heterosis. The fragment E12M5_250 was positively correlated with LY and BP and encodes glucose 1-dehydrogenase. Furthermore, seven TDFs including E4M3_260, E1M6_550, E2M5_250, E7M12_350, E12M1_210, E3M8_275, and E7M2_275 were positively correlated with heterosis performance of either yield trait. The respective TDF functions included encoding a GRAS family protein, a putative F28D10_20 protein, a phosphoglycerate kinase, a chlorophyll a-b binding protein, a magnesium transporter, a heavy metal ATPase, and a protein of unknown function.

Interestingly, E11M16_300 encoding a GAI/RGA-like protein was detected as positively correlated with yield and yield heterosis in Jiangpu during 2004, but negative correlations were identified in Linqing during 2008, indicating repression of vegetative growth may have different results on yield in different environments. E12M9_305 encoding ribonuclease II-like protein co-located with *qLP-LG19-1* and *qLI-LG19-1* was positively associated with SY, LY, and BP, but both positive and negative effects of this TDF were detected on yield heterosis, indicating the mechanisms underlying trait performance and heterosis might not be the same.

Discussion

The feasibility of QTL tagging based on a transcriptome map

Distinct tissue and cell type differentiation is highly dependent on specific patterns of gene expression and transcript accumulation (Ewing et al. 1999; Drost et al. 2010). In higher plants, housekeeping genes encoding essential metabolic enzymes or cellular components are primarily *cis* regulated (~92%) and are expressed constitutively at relatively low levels in all tissues/cell types, while tissue-specific genes that are largely *trans* regulated differ in their expression at the levels of transcription, typically with two- to ten-fold between the uninduced and induced levels (Drost et al.2010). Cotton yield and yield component traits are the cumulative results of a series of metabolic pathway, in which many interacting genes are involved. At florescence stage, cotton leaves as the main photosynthesis organs contribute most to the final production and are likely to express similar sets of genes in uninduced natural conditions; therefore functional leaves in one of the environments were sampled for RNA extraction in this study. cDNA-AFLP-based transcriptome map construction provides a viable approach to tag QTL and even the potential candidate genes controlling important qualitative and quantitative traits of interest (Liu et al. 2009; Fernández-del-Carmen et al. 2007).

Table 2 Correlation coefficients between differential expression of functional TDFs and performance of yield, yield heterosis, and other related traits

TDFs	Origin	Linkage group	Traits[A]					Mid-parent Heterosis[B]	
			SY	LY	BP	BW	LP	H_{MP}-SY	H_{MP}-LY
E2M16_380	CRI12	LG01	−0.0254	−0.0404	0.1285^{*c}	−0.0149	−0.116	−0.027	−0.0624
E4M3_260	CRI12	LG02	0.0822	0.035	0.1316^{*cb}	−0.0028	−0.2395^{**de}	0.2695^{**b}	0.2678^{**b}
E11M16_295	CRI12	LG02	−0.1675	−0.1469	−0.1837^{*ea}	−0.0346	0.0115	−0.2452^{**e}	−0.2174^{*e}
7M15_250	CRI12	LG03	0.1796^{*ad}	0.1808^{*d}	0.0615	−0.1394	0.0219	0.0036	0.0054
E11M16_240	CRI12	LG03	0.1999^{*d}	0.1831^{*d}	−0.0603	0.2063^{*c}	0.0693	0.0306	0.0432
E12M1_200	J8891	LG04	−0.1718	−0.0582	0.0075	−0.1207	0.0555	−0.2090^{*a}	−0.0833
E4M3_300	J8891	LG05	0.1729^{*c}	0.1665^{*c}	0.0697	−0.0655	−0.0101	0.1891^{*c}	0.1807^{*c}
E14M1_265	CRI12	LG06	−0.1998^{**c}	-0.1931^{**c}	0.0676	−0.1324	−0.1247	−0.0420	−0.0816
E11M15_310	J8891	LG06	0.0351	0.0425	0.0203	0.1351^{*b}	−0.0182	0.0344	0.0458
E1M6_550	J8891	LG07	0.056	0.0464	0.1280^{*d}	0.0224	0.0046	0.1928^{*b}	0.2288^{*b}
E2M3_235	CRI12	LG07	0.0005	0.1243^{*a}	0.1325^{*a}	−0.0665	−0.0172	0.1914^{*a}	0.1335^{*a}
E8M12_205	CRI12	LG07	−0.0059	−0.021	0.0283	−0.0641	−0.1824^{*a}	−0.0987	−0.0987
E16M8_550	J8891	LG07	−0.114	−0.1126	−0.1131	−0.0377	0.0281	−0.2257^{*de}	0.1954^{*de}
E2M3_480	J8891	LG09	−0.1322^{*c}	−0.1518^{*c}	0.0159	0.0121	−0.1579^{*a}	0.1459	0.1028
E5M4_480	CRI12	LG09	−0.0195	−0.0186	0.1584^{*c}	−0.0068	−0.0231	−0.0300	−0.0224
E2M5_250	J8891	LG10	−0.1226^{*d}	−0.1443^{*d}	0.0458	−0.0575	−0.1327^{*d}	0.2521^{**c}	0.2321^{*c}
E7M12_350	J8891	LG10	0.1325	0.1574	0.0566	0.0014	0.0677	0.2068^{*e}	0.2459^{**eb}
E11M16_300	J8891	LG10	*0.2680*^{**ad}	*0.2624*^{**ad}	*0.2327*^{**ad}	0.0852	−0.0632	*0.3131***^{ad}	*0.3185*^{**ad}
E12M1_210	J8891	LG10	−0.2108^{*c}	−0.2552^{**c}	−0.0219	0.0623	−0.3308^{**dabce}	−0.0564	0.2616^{**cd}
E7M4_340	J8891	LG11	0.0462	0.0241	−0.0057	0.1899^{*b}	−0.0374	0.0748	0.0692
E3M8_275	J8891	LG13	0.0539	0.0289	0.1293^{*a}	−0.0512	−0.1364^{*d}	0.3036^{**ae}	0.3184^{*ae}
E6M4_200	J8891	LG13	0.1699^{*a}	0.1850^{**a}	0.1939^{**ae}	0.0008	0.0288	0.1180	0.1217
E11M7_320	J8891	LG13	0.0215	−0.0288	0.3167^{**c}	−0.0455	−0.2438^{**d}	0.0353	−0.0200
E1M6_250	J8891	LG15	0.2278^{**b}	0.2262^{**b}	0.1713^{**b}	0.0627	0.1493^{*dec}	0.2018^{*b}	0.1954^{*b}
E4M8_340	J8891	LG15	0.2135^{*e}	0.2375^{**eb}	0.0820	0.2190^{*ce}	0.1533	0.1379	0.1855^{*e}
E10M6_600	J8891	LG15	0.2128^{*c}	0.2368^{**ce}	0.0703	0.1011	0.1414	0.1440	0.2188^{*c}
E10M10_350	J8891	LG15	−0.0202	0.0059	−0.0053	0.0396	0.1864^{*d}	−0.2017^{*a}	−0.1948^{*a}
E4M16_280	J8891	LG16	0.1359^{*e}	0.1379^{*ec}	0.1318^{*c}	−0.0500	0.1419^{*c}	0.2175^{*d}	0.2321^{*d}
E11M15_470	J8891	LG16	0.2198^{*d}	0.2364^{*d}	0.1892^{*d}	0.1647	0.1649	−0.0508	−0.0840
E6M9_200	J8891	LG17	−0.0143	0.0205	−0.0410	0.1945^{*a}	0.1394	−0.1164	−0.0489
E12M5_250	J8891	LG17	−0.0523	0.1852^{*a}	0.1986^{*a}	0.0848	−0.0082	0.1439	0.1206
E7M2_275	CRI12	LG19	0.0532	−0.0177	0.1959^{*a}	−0.0736	−0.2976^{**cdeab}	0.1943^{*a}	0.0702
E8M9_260	CRI12	LG19	−0.2933^{**d}	−0.3143^{**d}	−0.2538^{**d}	0.0166	−0.2261^{**eac}	−0.2644^{**d}	−0.2897^{**d}
E12M9_305	CRI12	LG19	0.2451^{**bc}	0.2253^{**bc}	0.2636^{**b}	0.0291	−0.0556	*0.2774*^{**bc}	*0.2564*^{**bc}
E16M8_250	CRI12	LG22	0.1364	0.2341^{**c}	0.2457^{**c}	−0.0328	−0.0824	0.1042	0.1269
E14M1_540	J8891	LG23	−0.0589	−0.0494	0.0570	0.2165^{*ed}	0.1509^{*e}	−0.0862	−0.0874
E5M15_480	J8891	LG26	−0.1982^{*ec}	−0.1860^{*e}	−0.1828^{*e}	−0.0137	0.0701	−0.2441^{*eab}	−0.2379^{*e}

Correlation coefficients were calculated separately across different environments and by combining all environments

* and ** indicated significant at levels of 0.05 and 0.01 respectively

[A] Traits analyzed including two yields SY (seed cotton yield, g/plant) and LY (lint yield, g/plant) and three yield components BP (bolls/plant), BW (boll weight, g) and LP (lint percentage, %)

[B] The mid-parent heterosis of the IF_2 for SY and LY was calculated as $H_{MP} = (IF_2-(P_1 + P_2)/2) \times 100\%$, using CRI12 and J8891 as parents

[a] Jianpu 2004

[b] Jiangpu 2005

[c] Jiangpu 2008

[d] Linqing 2008

[e] Combined analysis using the means of the above four environments. If no significant correlation was detected, the value of coefficients in combined analysis was used; if significant correlation was detected in multiple environments, the value in the first environment was used. Italicized figure indicated both positive and negative correlations were detected across environments

TDFs target highly conserved coding regions, each having a unique segregation pattern, as opposed to neutral DNA markers that tend to cluster on centromere or other repetitive non-coding regions (Brugmans et al. 2002). Hovav et al. (2008) reported an overwhelming majority of genes are transcribed in cotton fibers of different developmental stages. Comparative expression profiling in different plant materials has indicated that TDFs largely represent constitutively expressed house keeping genes (Ritter et al. 2008). Therefore, transcriptome maps can be used to map QTL of important traits similar to DNA marker-based maps, irrespective of the specific stages/cell types. TDFs are transcribed coding regions of functional genes without repetitive DNA, introns and spacers; therefore, the TDFs co-located with QTL have potential for cloning the functional genes responsible for each corresponding trait and have future value in marker-assisted selection.

The absence/presence of cDNA-AFLP fragment polymorphisms in a segregating population is predominately the result of genomic sequence polymorphisms in the coding region. This is likely due to SNPs or indels in the transcribed sequences (Brugmans et al. 2002). Some polymorphisms are probably cDNA specific, generated by splicing-site changes after transcription. In the present research, each polymorphic band was read separately; therefore, co-dominant allelic fragments generated from the same primer combination were initially mapped at the same or very close position, and only one fragment remained in the map to represent the allelic site. Band intensity polymorphisms representing quantitative differences in gene expression were not included due to the difficulty in discriminating by direct visualization. However, intensity polymorphisms that demonstrated clear-cut distinctions between parents and IF_2 were genotyped into absence/presence classes and used in map construction. The percentage of mapped TDFs (302/573, 52.7%) was lower than reported in Brugmans et al. (2002) (65.9%) and Li et al. (2003) (72.6%) , primarily caused by the inefficiency of dominant markers and the higher number of distorted markers adopted. The LGs contained averagely 12 markers. The LG24-26 each consisting of three markers and the QTL detected on them might be not reliable, therefore other markers are needed to fill the gaps in the future.

Since DNA markers were used to construct a genetic linkage map for the original RIL mapping population for QTL mapping in the RIL (Wang et al. 2006) and in the IF_2 (Wang et al. 2007), integration of the current expression map with the prior mapping work using SSR markers will help reveal the likely chromosomes of the 26 linkage groups constructed from the TDF markers. We tried, and unfortunately these two kinds of maps are hard to integrate since cDNA-AFLP transcriptome can be identified as cis and trans traits. As we know, cis trait is one that genetically maps to the physical location of the gene encoding its mRNA, suggesting that variation at the locus is responsible for the heritable changes in gene expression. A trans trait maps to a region distinct from its physical location and implies the location of a potential regulator acting in trans. In species such as cotton in which no whole genome sequence is available, we presently cannot distinguish cis and trans traits, resulting in no reliable integration of two kinds of maps.

Complexity of elite gene exploitation for trait improvement in cotton breeding

Yield and yield components are highly sensitive to environmental variation. Consequently, replicated field experiments across multiple environments are required to detect stable QTL for these traits. In this study, XZM2-derived IF_2 were generated for QTL mapping of yield and yield components at four different environmental sites. A total of 71 QTL were detected, with 13 in at least two environments. Significant clusters of QTL for highly related traits were found in several linkage groups. Nearly all QTL for seed cotton yield (SY) and lint yield (LY) were mapped at the same positions, and yield QTL were co-located with at least one QTL for yield components. These results suggested that the genetic potential of cotton varieties cannot be predicted simply based on phenotype. Results further indicated the inherent complexity of elite gene exploitation for trait improvement in hybrid breeding. It is notable that 38.8% of the total QTLs associated with improvement of yield and yield heterosis were detected from the phenotypically inferior parent CRI12. Yield heterosis exhibited in XZM2 should be the result of elite gene combinations from both parents; therefore, the low-yield parent CRI12 clearly contributes to heterosis. Similar results have been reported in both cotton and other crops; and alleles from phenotypically inferior parents have been associated with trait performance improvement for many detected QTL (Chee et al. 2005; Huang et al. 2003; Xie et al. 2008, Liu et al. 2009). Tanksley and McCouch (1997) pointed out that the phenotype of a plant is only a modest predictor of its genetic potential. For example, a high-yielding line often maintains positive alleles at most loci associated with yield, and typically an inferior parent contributes a superior allele from one or more loci. For this reason, QTL and other marker-based methodologies are suitable strategies for improving breeding efficiency. The present study has profound implications that although as the low-yield parent, CRI12 as good combining ability variety clearly contributes to heterosis in XZM2.

Mining and utilization of candidate genes for yield related traits

A lot of effort has been placed on tagging QTL for important fiber quality and agronomic traits in cotton (Ulloa et al. 2005; He et al. 2007; Guo et al. 2006; Wu et al. 2009), and cotton genome sequencing is currently underway (Yu 2010). However, little is known regarding the genes controlling these complicated quantitative traits. In the current study, potential candidate genes were analyzed by homology searches for 44 TDFs co-located with QTL for yield and yield components. According to Paterson and Smith (1999), the average DNA content of 1 cM in genetic maps is estimated to be 400–600 kb. The 95% confidence interval for QTL detected in this study was predominantly 10 cM; therefore many genes could be within one QTL interval, making it difficult to identify all the functional genes. However, neighboring genes tend to have related function and co-expressed (Michalak 2008; Williams and Bowles 2004; Zhan et al. 2006), and co-located TDFs are modest indicators of true functional genes. However, the genotypes that did not generate cDNA-AFLP bands can also provide beneficial allele for QTL. In such instance, the other allele relative to the gene represented by the TDF, which escaped detection by cDNA-AFLP techniques due to lack of restriction site, might be the potential candidate gene. In this study, correlation analysis between TDF expression pattern and yield performance detected some TDFs positively associated with yield and yield heterosis; indicating the proteins encoded by these putative candidate genes were potentially involved in the biological processes that contribute to the accumulation of cotton yield. Some TDFs exhibited expression negatively correlated with trait performance, consistent with a recessive or negative overdominant corresponding QTL. In particular, TDF E6M9_320 co-located with three recessive QTL for SY, LY, and LP on LG02, which were not related to an increase in yield but might be responsible for improving fiber quality by encoding putative cellulose synthase.

The ultimate goal of this study was to identify genes responsible for yield and yield components and potentially cotton heterosis. The results of associated TDFs provided insights into the final cloning of candidate genes, which might be crucial in the ultimate performance of quantitative traits. In the future, related clones will be identified by screening the cotton BAC library, and integrated into the cotton physical map in our lab, facilitating the map-based cloning of functional genes for yield and yield components. Furthermore, QTL-associated TDFs can be converted into user-friendly CAPs markers for marker-assisted breeding.

Acknowledgments This work was financially supported in part from the 973 (2011CB109303) and 863 (2009AA101104). We thank Dr. RJ Kohel in USDA-ARS for a critical review of the manuscript and Prof. Yuqiang Li in Hunan Cotton Research Institute for providing breeder-owned parental seeds of XZM2.

References

Abdalla AM, Reddy OUK, El-Zik KM, Pepper AE (2001) Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theor Appl Genet 102:222–229

Altschul SF, Madden TL, SchaVer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

An CF, Saha S, Jenkins JN, ScheZer BE, Wilkins TA, Stelly DM (2007) Transcriptome profiling, sequence characterization, and SNP based chromosomal assignment of the *EXPANSIN* genes in cotton. Mol Genet Genomics 278:539–553

An CF, Saha S, Jenkins JN, Ma DP, ScheZer BE, Kohel RJ, John ZY, Stelly DM (2008) Cotton (*Gossypium* spp.) R2R3-MYB transcription factors SNP identification, phylogenomic characterization, chromosome localization, and linkage mapping. Theor Appl Genet 116:1015–1026

Bachem C, van der Hoeven RS, de Bruijn SM, Vreugdenhil D, Zabeau M, Visser RG (1996) Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant J 9:745–753

Bachem C, Oomen R, Visser R (1998) Transcript imaging with cDNA-AFLP: a step by step protocol. Plant Mol Biol Rep 16:157–173

Basten CJ, Weir BS, Zeng ZB (2001) QTL Cartographer, Version 1.15. Department of Statistics. North Carolina State University, Raleigh

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2008) GenBank. Nucleic Acids Res 36(Database issue):D25–D30

Breyne P, Zabeau M (2001) Genome-wide expression of plant cell cycle modulated genes. Curr Opin Plant Biol 4:136–142

Breyne P, Dreesen R, Cannoot B, Rombaut D, Vandepoele K, Rombauts S, Vandeerhaeghen R, Inzé D, Zabeau M (2003) Quantitative cDNA-AFLP analysis for genome-wide expression studies. Mol Gen Genomics 269:173–179

Brugmans B, Fernández del Carmen A, Bachem CW, van Os H, van Eck HJ, Visser RG (2002) A novel method for the construction of genome wide transcriptome maps. Plant J 31(2):211–222

Chee PW, Rong J, Williams-Coplin D, Schulze SR, Paterson AH (2004) EST derived PCR-based markers for functional gene homologues in cotton. Genome 47:449–462

Chee P, Draye X, Jiang CX, Decanini L, Delmonte T, Bredhauer R, Smith CW, Paterson AH (2005) Molecular dissection of interspecific variation between *Gossypium hirsutum* and *Gossypium barbadense* (cotton) by a backcross-self approach: I. Fiber elongation. Theor Appl Genet 111:757–763

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Drost DR, Benedict CI, Arthur Berg A, Evandro Novaes E, Novaes CR, Yu QB, Dervinis C, Maia JM, Yap J, Miles B, Kirst M (2010) Diversification in the genetic architecture of gene expression and transcriptional networks in organ differentiation of Populus. PNAS 107(18):8492–8497

Ewing RM, Kahla AB, Poirot O, Lopez F, Audic S, Claverie JM (1999) Large scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. Genome Res 9:950–959

Fan SY, Le JG, Cheng GJ, Wu CJ (2008) Construction of Chinese Cabbage-pak-choi transcriptome map with cDNA-AFLP techniques. Sci Agri Sinica 41(6):1735–1741

Fernández-del-Carmen A, Celis-Gamboa C, Visser RG, Bachem C (2007) Targeted transcript mapping for agronomic traits in potato. J Exp Bot 58:2761–2774

Frelichowski JE, Palmer M, Main D, Tomkins JP, Cantrell RG, Stelly DM, Yu J, Kohel RJ, Ulloa M (2006) Genetic mapping of microsatellites derived from BAC-end sequences of *Gossypium hirsutum* Acala 'Maxxa'. Mol Genet Genom 275(5):479–491

Guo WZ, Ma GJ, Zhu YC, Yi CX, Zhang TZ (2006) Molecular tagging and mapping of quantitative trait loci for lint percentage and morphological marker genes in Upland cotton. J Integ Plant Biol 48(3):320–326

Guo WZ, Cai CP, Wang CB, Han ZG, Song XL, Wang K, Niu XW, Wang C, Lu KY, Shi B, Zhang TZ (2007) A microsatellite-based, generich linkage map reveals genome structure, function and evolution in *Gossypium*. Genetics 176(1):527–541

Han ZG, Guo WZ, Song XL, Zhang TZ (2004) Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. Mol Gen Genet 272:308–327

Han Z, Wang C, Song X, Guo WZ, Gou J, Li C, Chen X, Zhang TZ (2006) Characteristics, development and mapping of *Gossypium hirsutun* derived EST-SSR in allotetraploid cotton. Theor Appl Genet 112:430–439

He DH, Lin ZX, Zhang XL, Nie YC, Guo XP, Zhang YX, Li W (2007) QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of *Gossypium hirsutum* × *Gossypium barbadense*. Euphytica 153:181–197

Hovav R, Udall JA, Hovav E, Rapp R, Flagel L, Wendel JF (2008) A majority of cotton genes are expressed in single-celled fiber. Planta 227:319–329

Hua JP, Xing YZ, Xu CG, Sun XL, Yu SB, Zhang QF (2002) Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. Genetics 162:1885–1895

Hua JP, Xing YZ, Wu WR, Xu CG, Sun XL, Zhang QF (2003) Single-locus heterotic effects and dominance-by-dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 100(5):2574–2579

Huang XQ, Cöster H, Ganal MW, Röder MS (2003) Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). Theor Appl Genet 106:1379–1389

Jiang JX, Zhang TZ (2003) Extraction of total RNA in cotton tissues with CTAB-acidic phenolic method. Cotton Sci 15(3):166–167 (in Chinese)

Kendall M (1948) Rank Correlation Methods. Charles Griffin & Company Limited, London

Kohel RJ, Yu J, Park Y-H, Lazo GR (2001) Molecular mapping and characterization of traits controlling fiber quality in cotton. Euphytica 121:163–172

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

Li, YQ, Zeng ZY, Jing L, Yang FQ, Wan YL (1997) Selection and application of Xiangzamian 2. Crop Research (4):27–29 (in Chinese)

Li G, Gao M, Yang B, Quiros CF (2003) Gene for gene alignment between the Brassica and Arabidopsis genomes by direct transcriptome mapping. Theor Appl Genet 107:168–180

Lin Z, He D, Zhang X, Nie Y, Guo X, Feng C, Stewart JMcD (2005) Linkage map construction and mapping QTL for cotton fiber quality using SRAP, SSR and RAPD. Plant Breed 124:180–187

Liu HW, Wang XF, Pan YX, Shi RF, Zhang GY, Ma ZY (2009) Mining cotton fiber strength candidate genes based on transcriptome mapping. Chin Sci Bull 54:4651–4657

McCouch SR, Cho YG, Yano PE, Blinstrub M, Morishima H, Kinoshita T (1997) Report on QTL nomenclature. Rice Genet Newslett 14:11–13

Mei M, Syed NH, Gao W, Thaxton PM, Smith CW, Stelly DM, Chen ZJ (2004) Genetic mapping and QTL analysis of fiber related traits in cotton (*Gossypium*). Theor Appl Genet 108:280–291

Michalak P (2008) Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. Genomics 91: 243–248

Pan YX, MA J, Zhang GY, Han GY, Wang XF, Ma ZY (2007) cDNA-AFLP profiling for the fiber development stage of secondary cell wall synthesis and transcriptome mapping in cotton. Chin Sci Bull 52:2358–2364

Park Y-H, Alabady MS, Sickler B, Wilkins TA, Yu J, Stelly DM, Kohel RJ, El-Shihy OM, Cantrell RG, Ulloa M (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line (RIL) cotton population. Mol Genet Genom 274:428–441

Paterson AH, Smith RH (1999) Future horizons: biotechnology for cotton improvement. In: Smith WC (ed) Cotton: origin, history, technology and production. Wiley, USA, pp 415–432

Ranik M, Creux NM, Myburg AA (2006) Within-tree transcriptome profiling in wood-forming tissues of a fast-growing Eucalyptus tree. Tree Physiol 26:365–375

Reinisch AJ, Dong JM, Brubaker CL, Wendel JF, Paterson AH (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* × *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829–847

Ren LH, Guo WZ, Zhang TZ (2002) Identification of quantitative trait loci (QTLs) affecting yield and fiber properties in chromosome 16 in cotton using substitution line. Acta Botanica Sinica 44(7):815–820

Ritter E, Ruiz de Galarreta JI, van Eck HJ, Sánchez I (2008) Construction of a potato transcriptome map based on the cDNA–AFLP technique. Theor Appl Genet 116:1003–1013

Rong J-K, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee PW, Delmonte TA, Ding XL, Garza JJ, Marler BS, Park C-H, Pierce GJ, Rainey KM, Rastogi VK, Schulze SR, Trolinder NL, Wendel JF, Wilkins TA, Williams-Coplin TD, Wing RA, Wright RJ, Zhao X, Zhu L, Paterson AH (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166:389–417

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, New York

Samuelian S, Kleine M, Ruyter-Spira CP, Klein-Lankhorst RM, Jung C (2004) Cloning and functional analyses of a gene from sugar beet up-regulated upon cyst nematode infection. Plant Mol Biol 54:147–156

Shappley ZW, Jenkins JN, Meredith WR, McCarty JC Jr (1998) An RFLP linkage map of Upland cotton, *Gossypium hirsutum* L. Theor Appl Genet 97:756–761

Shen XL, Guo WZ, Lu QX, Zhu XF, Yuan YL, Zhang TZ (2007) Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in Upland cotton. Euphytica 155: 371–380

Stuber CW, Edwards MD, Wendel JF (1987) Molecular marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. Crop Sci 27:639–648

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277(22): 1063–1066

Ulloa M, Saha S, Jenkin N, Meredith WR, McCarty JC, Stelly DM (2005) Chromosomal assignment of RFLP linkage groups

harboring important QTL on an intraspecific cotton (*Gossypium hirsutum* L.) joinmap. J Hered 96:132–144

Van Ooijen JW, Voorrips RE (2001) JoinMap® version 3.0: software for the calculation of genetic linkage maps. Plant Res Int, Wageningen

Voorrips RE (2006) MapChart 2.2: software for the graphical presentation of linkage maps and QTLs. Plant Res Int, Wageningen

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Wang BH, Guo WZ, Zhu XF, Wu YT, Huang NT, Zhang TZ (2006) QTL mapping of fiber quality in an elite hybrid derived-RIL population of Upland cotton. Euphytica 152:367–378

Wang BH, Wu YT, Guo WZ, Zhu XF, Huang NT, Zhang TZ (2007) QTL analysis and epistasis effects dissection of fiber qualities in an elite cotton hybrid grown in second-generation. Crop Sci 47:1384–1392

Williams EJ, Bowles DJ (2004) Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*. Genome Res 14:1060–1067

Wu J, Gutierrez OA, Jenkins JN, McCarty JC, Zhu J (2009) Quantitative analysis and QTL mapping for agronomic and fiber traits in an RI population of Upland cotton. Euphytica 165: 231–245

Xie X, Jin F, Song MH, Suh JP, Hwang HG, Kim YG, McCouch SR, Ahn SN (2008) Fine mapping of a yield-enhancing QTL cluster associated with transgressive variation in an *Oryza sativa* × *O. rufipogon* cross. Theor Appl Genet 116:613–622

Yu SX (2010) Progress on Upland cotton sequencing. In: Proceed Plant & Animal Genomes XVIII Conference, Town & Country Convention Center, San Diego, CA. January 9–13, 2010

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

Zhan S, Horrocks J, Lukens LN (2006) Islands of co-expressed neighbouring genes in *Arabidopsis thaliana* suggest higher-order chromosome domains. Plant J 45:347–357

Zhang J, Guo WZ, Zhang TZ (2002) Molecular linkage map of allotetraploid cotton (*Gossypium hirsutum* L × *Gossypium barbadense* L.) with a haploid population. Theor Appl Genet 105:1166–1174