Theor Appl Genet (2010) 120:1347–1366
DOI 10.1007/s00122-010-1260-6

ORIGINAL PAPER

Near-isogenic cotton germplasm lines that differ in fiber-bundle strength have temporal differences in fiber gene expression patterns as revealed by comparative high-throughput profiling

Doug J. Hinchliffe · William R. Meredith · Kathleen M. Yeater · Hee Jin Kim · Andrew W. Woodward · Z. Jeffrey Chen · Barbara A. Triplett

Received: 14 January 2009 / Accepted: 27 December 2009 / Published online: 20 January 2010

Abstract Gene expression profiles of developing cotton (*Gossypium hirsutum* L.) fibers from two near-isogenic lines (NILs) that differ in fiber-bundle strength, short-fiber content, and in fewer than two genetic loci were compared using an oligonucleotide microarray. Fiber gene expression was compared at five time points spanning fiber elongation and secondary cell wall (SCW) biosynthesis. Fiber samples were collected from field plots in a randomized, complete block design, with three spatially distinct biological replications for each NIL at each time point. Microarray hybridizations were performed in a loop experimental design that allowed comparisons of fiber gene expression profiles as a function of time between the two NILs. Overall, developmental expression patterns revealed by the microarray experiment agreed with previously reported cotton fiber gene expression patterns for specific genes. Additionally, genes expressed coordinately with the onset of SCW biosynthesis in cotton fiber correlated with gene expression patterns of other SCW-producing plant tissues. Functional classification and enrichment analysis of differentially expressed genes between the two NILs revealed that genes associated with SCW biosynthesis were significantly up-regulated in fibers of the high-fiber quality line at the transition stage of cotton fiber development. For independent corroboration of the microarray results, 15 genes were selected for quantitative reverse transcription PCR analysis of fiber gene expression. These analyses, conducted over multiple field years, confirmed the

Communicated by H. T. Nguyen.

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the United States Department of Agriculture.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1260-6) contains supplementary material, which is available to authorized users.

D. J. Hinchliffe · B. A. Triplett (✉)
USDA-ARS-SRRC, 1100 Robert E. Lee Blvd.,
New Orleans, LA 70124, USA
e-mail: barbtriplett@gmail.com

D. J. Hinchliffe
e-mail: doug.hinchliffe@ars.usda.gov

W. R. Meredith
USDA-ARS-MSA, Stoneville, MS 38776, USA
e-mail: bill.meredith@ars.usda.gov

K. M. Yeater
USDA-ARS-SPA, College Station, TX 77840, USA
e-mail: kathleen.yeater@ars.usda.gov

H. J. Kim
Department of Plant Pathology and Crop Physiology,
Louisiana State University, Baton Rouge, LA 70803, USA
e-mail: heejin.kim@ars.usda.gov

A. W. Woodward · Z. J. Chen
Section of Molecular Cell and Developmental Biology,
The University of Texas at Austin,
1 University Station A6700, Austin, TX 78712, USA
e-mail: Andrew.W.Woodward@rice.edu

Z. J. Chen
e-mail: zjchen@mail.utexas.edu

Present Address:
A. W. Woodward
Department of Biochemistry and Cell Biology,
Rice University, 6100 Main Street MS-140,
Houston, TX 77251-1892, USA

temporal difference in fiber gene expression between the two NILs. We hypothesize that the loci conferring temporal differences in fiber gene expression between the NILs are important regulatory sequences that offer the potential for more targeted manipulation of cotton fiber quality.

Introduction

The seeds of cultivated tetraploid cotton (*Gossypium hirsutum* L. and *Gossypium barbadense* L.) produce single-celled trichomes, referred to as fibers, than can attain lengths of nearly 60 mm at maturity (Kim and Triplett 2001). The value of the US cotton crop is estimated at $5.2 billion annually (National Cotton Council of America 2008). At maturity, fiber is composed almost entirely of multiple layers of thickened secondary cell wall (SCW) and is approximately 94% pure cellulose (β-1,4-linked D-glucose units). Overlapping stages of fiber development occur in all genotypes and consist of fiber initiation, fiber elongation, SCW biosynthesis or thickening, and maturation (Basra and Malik 1984). The duration and rate of each stage varies among genotypes. During the elongation stage of development, the fiber cell undergoes rapid polar expansion and is delimited by only a primary cell wall (PCW). The onset of SCW biosynthesis can vary depending on cotton species and environmental conditions, but typically occurs from 16 to 21 DPA (Basra and Malik 1984; Wilkins and Jernstedt 1999). The period of time between the beginning of SCW biosynthesis and the end of the elongation stage is referred to as the transition stage. Distinct patterns of gene expression characterize each of the developmental stages (reviewed in Smart et al. 1998; Haigler et al. 2005; Shi et al. 2006; Lee et al. 2007). During the transition stage, the expression profiles of genes related to SCW biosynthesis are up-regulated, whereas the transcript abundance of genes related to elongation and PCW biosynthesis begins to decline. The thickness of the SCW, degree of polymerization of cellulose molecules (number of D-glucose units), orientation of cellulose microfibrils (MFs), and the number of reversal points (changes in gyre of MF orientation) are all attributes that can potentially affect classification and pricing of the mature fiber (Triplett 1992).

The cotton germplasm lines MD 52ne and MD 90ne are near-isogenic lines (NILs) of cotton (*G. hirsutum* L.) both derived by backcross breeding. MD 90ne is the recurrent parent and MD 52ne is a BC_6 high-bundle strength selection (Meredith 2005a). In these studies, MD 52ne was shown to have approximately 10% higher fiber-bundle strength than the recurrent parent MD 90ne (Meredith 2005b). Cotton fiber-bundle strength is a key determining factor in the price of cotton, especially in the export market which now accounts for approximately two-thirds of the cotton grown in the United States (Jung 2008). In addition to increased fiber-bundle strength, MD 52ne has significantly lower short-fiber content (22% less) compared with MD 90ne (Meredith 2005b). Higher short-fiber content results in fewer spinnable fibers leading to reduced spinning efficiency and yarn quality and is a significant source of waste. Previous studies conducted by Meredith (2005b) on the cotton NILs MD 52ne and MD 90ne indicated that the higher fiber-bundle strength trait of MD 52ne is controlled by a small number of genes (<2) or possibly a single locus of closely linked genes (Meredith 2005b). The increased fiber-bundle strength of MD 52ne also appears to be highly heritable indicating predominately gene effects and not environment effects controlling the phenotype (Meredith 2005b).

The cotton NILs MD 52ne and MD 90ne released in 2005 (Meredith 2005a) provide an unparalleled opportunity to study the genetic components of cotton fiber strength in two near-isogenic populations. The recently developed cotton oligonucleotide gene chip for microarray analysis combines expressed sequence tag (EST) data from *Gossypium arboreum* L., *Gossypium raimondii* L., and *G. hirsutum* L. from over 50 cDNA libraries and a myriad of tissues, including initiating, elongating, and SCW-producing fibers (Udall et al. 2007). In this study, we have used high-throughput methods for analyzing fiber gene expression in this unique pair of genotypes to identify candidate genes associated with higher cotton fiber strength.

Materials and methods

Plant material

The cotton NILs MD 52ne and MD 90ne resulted from two separate backcross programs as described by Meredith (2005a). MD 90ne is an advanced generation backcross, BC_4F_{10} germplasm line derived from crossing 'Deltapine Acala 90' to a 'Deltapine 16' nectariless line designated MD 65-11ne. Half of the parentage for Deltapine Acala 90 was Deltapine 16 and the other half involved exotic and western US-type Acalas. During the backcross procedure for MD 90ne, selection was practiced only for the nectariless trait (Meyer and Meyer 1961). MD 52ne is an advanced generation BC_6F_6 germplasm line derived from crossing MD 90ne with MD 65-11ne and selecting for high-fiber strength in each segregating generation. The pedigrees of MD 90ne and MD 52ne are shown in Fig. 1. Additional pedigree information on cotton lines used to develop the progenitors of MD 52ne and MD 90ne can be found in Calhoun et al. (1994) and Bowman et al. (2006).

Field experimental design and fiber quality measurements

Fiber samples for quality evaluation were taken by hand harvest when 60% or more of the bolls were open. All

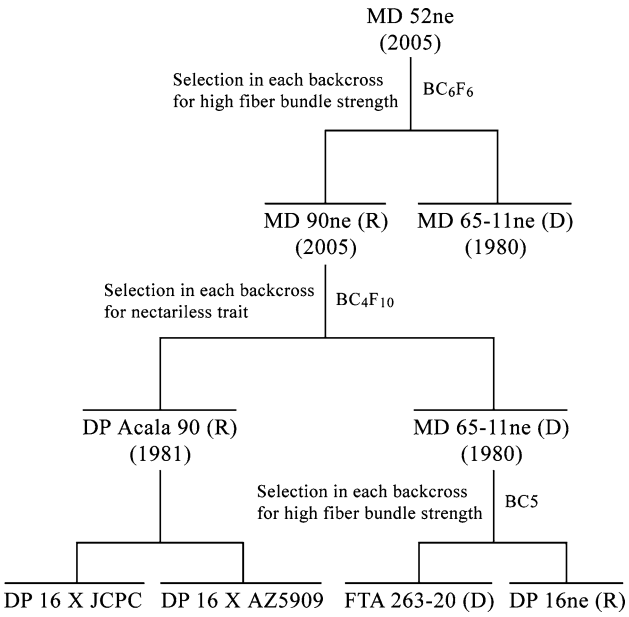


Fig. 1 Pedigrees of the cotton near-isogenic lines (*NILs*), MD 52ne and MD 90ne. Trait selections and generations are indicated for germplasm lines that were developed by backcross breeding. Germplasm line release dates are shown in *parentheses*. *R* Recurrent parent, *D* donor parent, *JCPC* John Cotton Poly Cross, *MD* Mississippi Delta, *DP* Deltapine, *FTA* ARS strain from Pee Dee Experiment Station, Florence, SC

samples were ginned on a laboratory gin. The field experimental design was a randomized complete block design. A total of 11 environments were sampled to determine fiber traits. The number of locations per year varied from one to five. Locations were near Stoneville, MS, and varied in soil type, planting dates, and other cultural practices. The number of locations was one in 2006 and 2008, four in 2005 and five in 2007. Plot size varied from one to three rows with sampling in the center row. Row dimensions were 1 m × 9 m and 1 m × 18 m. The number of replications varied from two to six. The genotypes evaluated were MD 52ne and MD 90ne and descend from the genetic study by Meredith (2005b) and subsequently released as germplasm lines (Meredith 2005a).

The data in Table 1 were from 60 observations, 30 for each germplasm line. The significant differences in fiber quality measurements of the cotton NILs were determined using analysis of variance (ANOVA). The combination of years and locations were treated as 11 random environments. The probability levels listed in Table 1 are for the one degree of freedom *F* test comparisons between the cotton NILs. No significant germplasm line by environmental interaction was detected.

Fiber tests given in Table 1 were determined by Star-Lab Inc., Knoxville, TN, USA. Bundle strength (T_1) was determined by stelometer that measures the grams-force per tex required to break a bundle of fibers, expressed as kilonewton meters per kilogram, in which fibers are held by two jaws separated by 3.2 mm. Elongation (E_1) is the percentage of elongation at the point of fiber breakage. Micronaire is a measure of fineness and maturity and was measured by the Fibronaire instrument and expressed in micronaire units. Fiber lengths were measured by the Digital Fibrograph. The 50% span length is that point on an array of fibers in which 50% of the fibers are that long or longer. Similarly, the 2.5% span length is that point on an array of fibers in which 2.5% of the fibers are that long or longer.

RNA extraction

Fiber samples for gene expression analyses were harvested from ten randomly selected, first-position bolls that had been tagged on the day of anthesis from triplicate 3-row plots grown at a single location near Stoneville, MS. In preliminary experiments conducted in 2003 and 2004, bolls were harvested from the replicate plots only at 12, 16, and 20 DPA. Subsequent field experiments conducted from 2006 to 2008 included additional developmental time points. In the 2006 and 2007 growing seasons, bolls were harvested at 8, 12, 14, 16, 20, and 24 DPA. In the 2008 growing season, bolls were harvested at 8, 10, 12, 14, 16, 18, 20, and 24 DPA. Harvested bolls were immediately chilled on ice in the field and transported to

Table 1 Average fiber traits for MD 90ne and MD 52ne across eleven combinations of locations and years

Germplasm line	Micronaire	Elongation (%)	Strength (kN m kg^{-1})	Span length (mm)	
				50%	2.5%
MD 90ne	4.63	6.69	236	14.7	29.5
MD 52ne	4.73	6.10	272	15.1	30.0
F test prob. level*	0.1083	0.0035	<0.0001	0.0167	0.0100
Environmental range					
High MD 90ne	5.15	7.88	267	16.0	30.8
High MD 52ne	5.00	7.50	308	16.1	31.3
Low MD 90ne	4.10	5.83	221	14.0	28.3
Low MD 52ne	4.13	4.67	240	14.2	29.5

* *F* test probability levels were obtained by ANOVA analysis

the lab. The seeds with fibers attached were quickly frozen in liquid nitrogen and stored at −80°C. Fibers were ginned from seeds by gently swirling the frozen samples in a mortar filled with liquid nitrogen to break fibers away from the ovules. Seeds were then removed from the mortar and the fiber tissue ground to a powder. Total RNA was extracted and an on-column DNase I digestion performed using the Sigma Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA) as per the manufacturer's protocol. RNA quantity was determined using a NanoDrop ND-1000 (NanoDrop Technologies, Wilmington, DE, USA), and RNA quality was verified using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

RNA amplification and labeling

RNA amplification and labeling was performed using the Amino Allyl MessageAmp™ II aRNA Amplification Kit (Ambion, Austin, TX USA) as per the manufacturer's protocol. In each amplification reaction, 1 µg of total RNA was used for starting material. The in vitro transcription reaction for synthesizing amplified RNA (aRNA) was performed for 12 h. Probe aRNA was indirectly labeled by dye-coupling with Cy3 or Cy5 fluorescent dyes for hybridization to the cotton oligonucleotide gene chips.

Microarray hybridization and image acquisition

For each hybridization, 2 µg of aRNA probe labeled with either Cy5 or Cy3 was used. The cotton oligonucleotide arrays utilized in this study were described in detail by Udall et al. (2007) and were printed on ArrayIt Super Amine 2 slides (TeleChem International Inc., Sunnyvale, CA, USA). The slides were prehybridized in 5× SSC, 0.1% SDS, and 0.1 mg/mL BSA (Fraction V) at 42°C for 45 min and then submerged in 0.1× SSC for 5 min three times, followed by a final submersion in ddH_2O for 30 s. The slides were dried by centrifugation at 1,000×*g* for 5 min. The probe hybridization buffer consisted of 35% formamide, 5× SSC, 0.1% SDS, and 0.1 mg/mL sheared fish testes DNA. All hybridizations were carried out at 42°C for 16 h in a Genetix 10-slide Hybridization Chamber (Genetix, New Milton, England). Following hybridization, slides were washed in 2× SSC, 0.1% SDS at 42°C for 5 min, then washed twice in 1× SSC at room temperature for 3 min, then washed twice in 0.1× SSC at room temperature for 2 min, and finally washed once in 0.05× SSC at room temperature for 30 s. After drying by centrifugation, the slides were treated with DyeSaver2 (Genisphere Inc., Hatfield, PA, USA) as per the manufacturer's protocol to prevent ozone-mediated degradation of the Cy5 dye. Slides were scanned in a GenePix 4000B microarray scanner (Molecular Devices Corporation, Sunnyvale, CA, USA) at 10 mm resolution using the GenePix Pro 6.0 software package.

Experimental design and statistical analysis

A loop hybridization experimental design was used to compare gene expression patterns in fiber cells from the NILs (MD 52ne and MD 90ne) over time (8, 12, 16, 20, and 24 DPA) (Fig. 2a). The microarray hybridizations were performed using two biological replicates per cotton germplasm line at each time point. The hybridization process also included dye-swaps of each pairing that served as a technical replicate in the analysis. Each hybridization produced a pair of red (Cy5) and green (Cy3) fluorescence intensities that serve as the inputs for all statistical analyses. For statistical analysis, the background-subtracted median fluorescence intensity of each individual DNA spot was used. The raw expression data were transformed to the $\log_2$

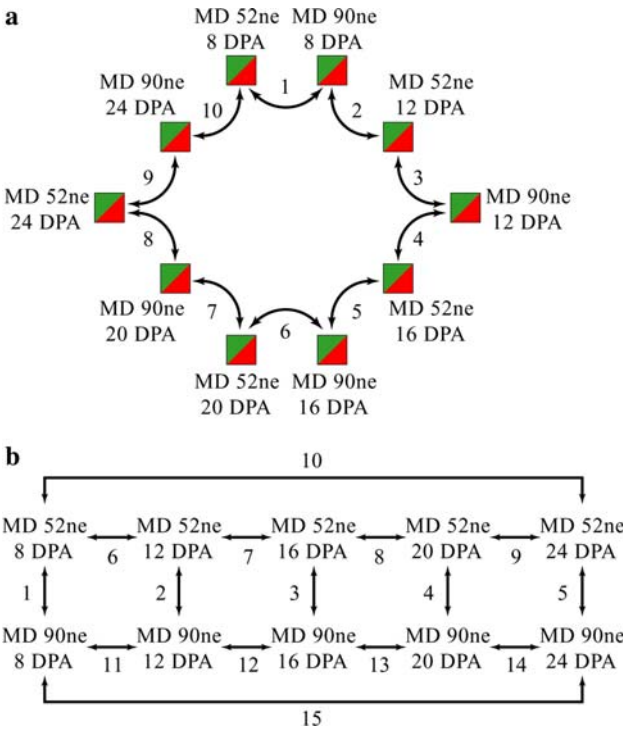


Fig. 2 Microarray hybridization experimental design and statistical analysis. **a** A loop microarray hybridization design was utilized for the experiment. Each *numbered point* in the loop represents microarray hybridization with probes generated from amplified fiber RNA of the indicated NIL and developmental stage. A total of ten microarray hybridizations (*numbered*) are indicated in the loop with each *block* representing technical replications in the form of a dye-swap. Two biological replications were performed for each microarray hybridization. **b** The microarray data derived from the loop hybridization experimental design allows comparisons (*numbered*) between the indicated developmental stages. The design also shows a more simplified view of the experiment comparisons on the cotton NILs between genotypes and over time

scale and all analyses were performed using JMP® Genomics version 3.0 (SAS Institute Inc., Cary, NC, USA).

Data distribution and ratio analysis tools were used to evaluate data quality. The data were normalized within arrays using Loess normalization and across arrays by median standardization (Dudoit et al. 2002). The normalized expression levels were further evaluated for quality using correlation and principal components analyses. The normalized expression levels were then analyzed in individual gene-specific models. The gene-specific models were fit in JMP® Genomics using mixed models analysis (Wolfinger et al. 2001). Germplasm lines, developmental stage (DPA), interaction between germplasm line and developmental stage, and dye channel were considered as fixed effects. The array effect was considered as a random effect. The estimates of means and standard errors, and difference of means, of the germplasm lines, developmental stage (DPA), and their interactions were calculated using the LSMEANS option. The multiple comparison tests were controlled for familywise error rate (FWER) with the Holm method (also known as a stepwise Bonferroni adjustment) (Holm 1979). The magnitude and significance of differences in the effects of germplasm line, developmental stage and their interaction were used to select genes for validation and further study.

The entire normalized microarray dataset output from the JMP® Genomics software (SAS Institute Inc., Cary, NC, USA) and the raw Genepix (Molecular Devices Corporation, Sunnyvale, CA, USA) data files for each microarray were submitted to the Gene Expression Omnibus (GEO) database (Barrett et al. 2007). These data are freely available under the GEO accession number GSE14388.

Functional classification and enrichment analysis of differentially expressed genes

The program Blast2GO was used to perform all bioinformatics on selected cotton tentative consensus sequences (TCs) and ESTs (Conesa et al. 2005). The end result of the analysis is the assignment of Gene Ontology (GO) terms (Ashburner et al. 2000) to selected TCs and ESTs based on BLASTx results (Gish and States 1993). Also, the Blast2GO program incorporates InterPro searches and the ability to map InterPro IDs to BLASTx-derived GO terms, merging the results into one integrated annotation result (Conesa et al. 2005). Enrichment analysis of statistically ($P < 0.006$) significant genes at different developmental time points was carried out using the integrated GOSSIP software package (Blüthgen et al. 2005) of the Blast2GO program that employs the Fisher's exact test to identify enriched GO terms in a group of genes. When comparing fiber gene transcript abundance between the two cotton NILs at the same time point (e.g., 16 DPA fibers of MD 52ne compared to 16 DPA fibers of MD 90ne), only genes considered biologically significant (twofold or greater difference in expression) were used in the enrichment analysis. The limit placed on statistical significance of the enrichment analysis was a false discovery rate (FDR)-corrected P value of 0.05 that is calculated by the Blast2GO program. In addition to the GenBank annotations generated by BLASTx searches, selected genes were used to query The *Arabidopsis* Information Resource (TAIR) database using a BLASTx search (Huala et al. 2001). TAIR sequence accessions obtained from the BLASTx searches were included in the results with the GenBank accessions.

The online program GeneVenn (Pirooznia et al. 2007) was used to construct Venn diagrams of statistically significant genes identified in the following data: genes up-regulated from 12 to 16 DPA in MD 52ne; genes up-regulated from 12 to 16 DPA in MD 90ne; genes up-regulated in MD 52ne at 16 DPA; genes up-regulated in MD 90ne at 16 DPA. A total of three Venn diagrams were generated (Fig. 3) and the Blast2GO program was used to perform enrichment analysis on all independent and overlapping Venn groups in each diagram. Only the Venn groups with statistically significant, enriched biological processes are labeled in the diagrams.

Confirmation of microarray gene expression profiles by quantitative reverse transcription PCR

Primers for quantitative reverse transcription PCR (qRT-PCR) were designed using the Primer Express 2.0 program (Applied Biosystems, Foster City, CA, USA). The reagents used for cDNA synthesis reactions were TaqMan Reverse Transcription Reagents (Roche Molecular Systems, Inc., Branchburg, NJ, USA) or the iScript™ cDNA Synthesis Kit (Bio-Rad Laboratories, Hercules, CA, USA). Control cDNA synthesis reactions to check for genomic DNA contamination during qRT-PCR consisted of the same template and components as the experimental reactions minus the reverse transcriptase enzyme. The PCR reactions were performed with ABI PowerSYBR® Green Master-Mix (Applied Biosystems, Foster City, CA, USA) or iTaq™ SYBR® Green Supermix with ROX (Bio-Rad Laboratories, Hercules, CA, USA) in an Applied Biosystems Prism 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA, USA). Thermal cycler parameters for qRT-PCR were as follows: 50°C 5 min, 95°C 2 min, 50 cycles of 95°C 15 s, 60°C 1 min. A dissociation curve was generated and used to validate that a single amplicon was present for each qRT-PCR reaction. Relative quantitation of gene transcript abundance was performed using

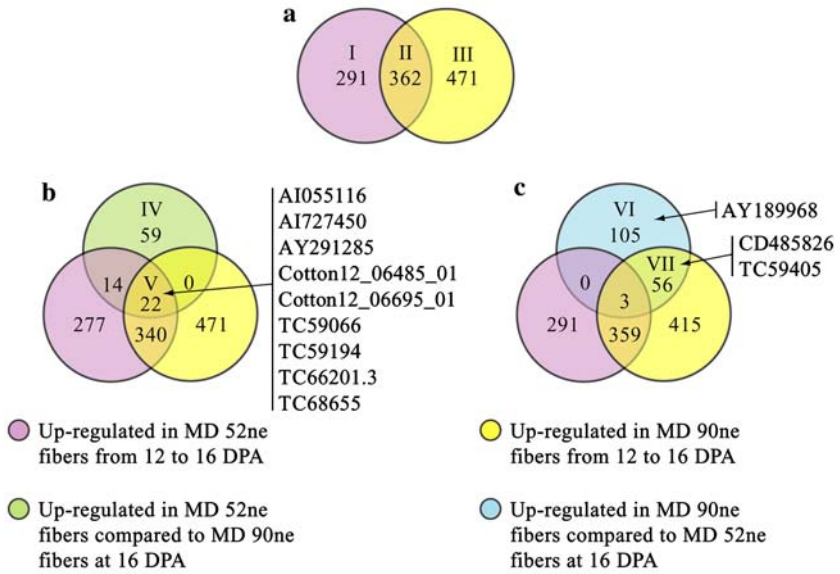


Fig. 3 Venn diagrams utilized for enrichment analysis and identification of genes. Venn group designations were assigned to gene lists that produced statistically significant enrichment analysis at a FDR-corrected $P < 0.05$. Venn groups without any group designations were not significantly enriched in any biological process. An abbreviated list of enriched GO terms associated with each Venn group is given in Table 3. **a** Genes statistically significant and up-regulated from 12 to 16 DPA in fibers of both cotton NILs, **b** genes statistically significant and up-regulated from 12 to 16 DPA in fibers of both cotton NILs, and genes statistically significant and up-regulated in fibers of MD 52ne compared to fibers of MD 90ne at 16 DPA, **c** genes statistically significant and up-regulated from 12 to 16 DPA in fibers of both cotton NILs, and genes statistically significant and up-regulated in fibers of MD 90ne compared to fibers of MD 52ne at 16 DPA

the comparative CT method as described in the ABI Prism 7700 Sequence Detection System User Bulletin #2. In an earlier study, we used *α-tubulin4* as a normalizer for qRT-PCR measurement of fiber gene expression during development (Kim and Triplett 2004a); however, significant improvements in our fiber RNA isolation procedure have resulted in much higher RNA quality and quantity. As a result, the integrity of 18S RNA varied less from sample to sample and between developmental time points than in previous experiments. Amplification efficiency and consistent CT values over fiber development time points (data not shown) led to the selection of 18S rRNA as the endogenous reference for the qRT-PCR results presented here.

A total of nine qRT-PCR reactions were performed for each germplasm line at each time point, representing three biological replications and three technical replications. In order to assess the statistical significance of the qRT-PCR results comparatively between the cotton NILs at each time point, a Wilcoxon signed-rank test was performed on the ΔC_t values (Yuan et al. 2006) in order to derive a 95% confidence interval and calculate the $\Delta\Delta C_t$ and fold difference in transcript abundance. A $P < 0.05$ and a greater than two-fold change in transcript abundance were considered significant. The Wilcoxon test was performed using the Analyse-it® Version 2.11 statistical analysis add-in for Microsoft Excel (Analyse-it Software Ltd, Leeds, England, UK).

Results

Fiber quality measurements

The summary of genetic differences between the two NILs over 11 environments is given in Table 1 and shows significant differences for all fiber traits except micronaire. MD 52ne produced lint with 15% greater ($P < 0.0001$) fiber strength than that of its NIL recurrent parent, MD90ne, confirming the superior fiber strength previously reported by Meredith (2005a). Both 50 and 2.5% span lengths of MD 52ne were significantly longer than those of MD 90ne; probability levels of 0.02 and 0.01, respectively. Elongation of MD 52ne was 6.1% and that for MD 90ne was significantly higher, 6.7%. The environmental effect was large for all fiber traits but no significant genotype-by-environment interaction was detected for any trait (data not shown). The environmental average bundle strength ranged from 221 to 267 kN m kg^{-1} for fibers of MD 90ne and 240 to 308 for fibers of MD 52ne kN m kg^{-1} (Table 1).

Enrichment analysis of genes differentially expressed during cotton fiber development

Preliminary studies using qRT-PCR were conducted by our laboratory to measure transcript abundance of specific *G. hirsutum* cellulose synthase genes in the NILs during two growing seasons prior to 2006. The results from both

years revealed an increased abundance of cellulose synthase 1 and cellulose synthase 2 (*GhCesA1* and *GhCesA2*) transcripts in fibers of MD 52ne compared with fibers of MD 90ne slightly before and during the transition stage of fiber development (for the results from 2003 see Supplemental Fig. 1). Since *GhCesA1* and *GhCesA2* are abundantly expressed during SCW biosynthesis in cotton fiber (Pear et al. 1996), we hypothesized that a temporal difference in fiber development existed in the cotton NILs, with fibers of the higher strength line, MD 52ne, entering the SCW biosynthesis stage earlier than fibers of MD 90ne. The focus of gene selection for corroboration of the microarray data in silico and by qRT-PCR was centered on time points related to the fiber transition and SCW biosynthesis developmental stages. In our microarray dataset, the fiber developmental time point most representative of the transition stage is 16 DPA.

Enrichment analysis was conducted on all up-regulated genes in fibers of both cotton NILs at all comparative development time points (Fig. 2b, comparisons 1, 2, 3, 4, and 5). The reference gene set used for the enrichment analyses consisted of the cotton TCs and ESTs represented on the microarray platform that were significantly expressed ($P < 0.006$) in fibers of either cotton NIL at any time point in the dataset. In this manner, our reference gene set was tailored to represent the fiber transcriptomes of the NILs during a specific time-course of fiber development and consisted of approximately 50% of the TCs and ESTs represented on the microarray. The results of the enrichment analyses, including complete lists of significantly enriched biological processes, GO identifications, FDR-corrected *P* values and the cotton microarray IDs represented in each enriched GO term, are given in Supplemental Table 1. The enrichment analysis on genes up-regulated in the 16 DPA dataset (Fig. 2b, comparison 3) revealed significant enrichment of biological processes involved in cell wall and cellulose biosynthesis exclusively in fibers of the higher fiber-bundle strength cotton germplasm line, MD 52ne. These data were of particular interest based on our preliminary studies mentioned above and the fact that the fiber transition period and entrance into the SCW biosynthesis stage typically occurs near the 16 DPA time point of fiber development in cotton grown in the Mississippi delta. Based on the up-regulated genes in fibers of MD 52ne compared with fibers of MD 90ne at 16 DPA, eleven GO terms were found to be significant at the FDR-corrected $P < 0.05$ as shown in Table 2. There were a total of ten unique sequences represented in the enriched GO terms that are related to plant cell wall biosynthesis, including two *G. hirsutum* cellulose synthase genes (*Ces*As), and orthologs of the *CesA7* gene (Taylor et al. 1999) and *COBRA-LIKE4* gene (*COBL4*) (Roudier et al. 2002) of *Arabidopsis*.

In order to further examine up-regulated genes and biological processes related to the fiber transition and SCW developmental stages, a separate enrichment analysis was conducted to include not only genes differentially expressed in fibers of the cotton NILs at 16 DPA, but also genes up-regulated in fibers of both NILs from 12 to 16 DPA (Fig. 3). The analysis revealed a significant increase in biological processes associated with cellulose and cell wall biogenesis in fibers of both NILs from 12 to 16 DPA (Fig. 3a, Venn Group II; Table 3, Venn Group II). Biological processes involved in cellulose and cell wall biosynthesis were also enriched and highly significant for 22 genes that were more abundantly expressed in 16 DPA fibers of MD 52ne compared with 16 DPA fibers of MD 90ne (Fig. 3b, Venn Group V). Table 3 lists the top significant GO terms (FDR-corrected $P < 0.05$) for each Venn group. A complete list of significantly enriched biological processes for each Venn group and the corresponding cotton microarray sequence IDs are shown in Supplemental Table 2. Further enrichment analysis was performed on genes up-regulated from 12 to 16 DPA in fibers of both NILs separately without generating Venn groups. In other words, the two test groups were Venn Group I + Venn Group II (genes up-regulated from 12 to 16 DPA in fibers of MD 52ne) and Venn Group II + Venn Group III (genes up-regulated from 12 to 16 DPA in fibers of MD 90ne). Only genes in the 12–16 DPA MD 52ne test group were significantly enriched in cellulose and cell wall biosynthetic biological processes. Even though Venn Group II in Fig. 3a indicates an overlap of biological processes involved in SCW and cellulose biosynthesis in fibers of both cotton NILs (Table 3, Venn Group II), this supplemental enrichment analysis revealed that these processes are more highly enriched in 12–16 DPA fibers of MD 52ne. Genes in the 12–16 DPA MD 90ne reference group were enriched in oxidoreductase activities that were also enriched in genes more abundantly expressed in 16 DPA fibers of MD 90ne (Fig. 3c, Venn Group VII).

In silico and qRT-PCR corroboration of microarray analysis

A total of 15 genes were selected for verification of the microarray data based on enrichment analysis of biological processes and expression profiles of genes differentially expressed during the transition and SCW biosynthesis stages of fiber development in both cotton NILs (Tables 4, 5). Prior to empirical corroboration of the comparative microarray data, the expression profiles of the selected genes from each cotton NIL were examined independently as a function of fiber development and compared with the existing literature on expression patterns of each gene. For example, *GhCesA1* and *GhCesA2* transcripts are nearly undetectable during the elongation stage of cotton fiber development, but become highly

Table 2 Enrichment analysis of gene ontologies up-regulated at 16 DPA in fibers of the germplasm line MD 52ne (significant at FDR-corrected $P < 0.05$)

GO term	Biological process	Sequence IDs of up-regulated genes
GO:0007047	Cell wall organization and biogenesis	AI055116, AI727450, AY291285, Cotton12_12955_01, TC59066, TC66201.3, TC68655, TC72660
GO:0045229	External encapsulating structure organization and biogenesis	AI055116, AI727450, AY291285, Cotton12_12955_01, TC59066, TC66201.3, TC68655, TC72660
GO:0009225	Nucleotide-sugar metabolic process	AI727450, Cotton12_07111_02, TC66201.3, TC68655, TC75914
GO:0009793	Embryonic development ending in seed dormancy	Cotton12_05329_01, Cotton12_06897_01, Cotton12_07111_02, Cotton12_14001_01, DR463436, DT566883
GO:0010382	Cell wall metabolic process	AI055116, AY291285, Cotton12_12955_01, TC68655
GO:0048316	Seed development	Cotton12_05329_01, Cotton12_06897_01, Cotton12_07111_02, Cotton12_14001_01, DR463436, DT566883
GO:0009790	Embryonic development	Cotton12_05329_01, Cotton12_06897_01, Cotton12_07111_02, Cotton12_14001_01, DR463436, DT566883
GO:0006040	Amino sugar metabolic process	AI055116, AY291285, Cotton12_25165_01
GO:0016998	Cell wall catabolic process	AI055116, AY291285, Cotton12_12955_01
GO:0009832	Cellulose and pectin-containing cell wall biogenesis	AI727450, TC59066, TC66201.3, TC68655
GO:0044247	Cellular polysaccharide catabolic process	AI055116, AY291285, TC60987

AI055116, class I chitinase (GhCTL1); AI727450, cellulose synthase (GhCesA3); AY291285, chitinase-like (GhCTL1); Cotton12_12955_01, peptidoglycan-binding LysM domain-containing protein; TC59066, COBRA-LIKE4 (AtCOBL4); TC66201.3, cellulose synthase (GhCesA2); TC68655, cellulose synthase (AtCesA7); TC72660, pectin methylesterase; Cotton12_07111_02, UDP-galactose 4-epimerase-like protein; TC75914, UDP-galactose 4-epimerase-like protein; Cotton12_05329_01, NAC domain-containing protein protein 74 (ANACO74); Cotton12_06897_01, MONOPTEROS transcription factor (AtMP); Cotton12_14001_01, ubiquitin protein ligase (ASK2); DR463436, repressor of silencing 1 (AtROS1); DT566883, flavin-containing monooxygenase/oxidoreductase (AtYUC4); Cotton12_25165_01, NADH:cytochrome-b5 reductase (ATCBR); TC60987, chloroplast beta-amylase (AtCT-BMY)

abundant beginning at the transition stage and during SCW biosynthesis (Kim and Triplett 2007; Pear et al. 1996). Based on our microarray and qRT-PCR data, the same pattern of expression was observed for *GhCesA1* and *GhCesA2* in fibers of MD 52ne and MD 90ne, supporting previous observations and providing in silico validation of our results. In addition, the *G. hirsutum* α-expansin gene (*GhExp1*) and germin-like protein gene (*GhGLP1*) were chosen as controls to quantify the expression profile of genes known to be more abundantly expressed during fiber elongation (Harmer et al. 2002; Kim and Triplett 2004b; Orford and Timmis 1998) compared with the SCW biosynthesis stage. The expression profiles of the remaining genes listed in Tables 4 and 5 are also supported in the literature. These independent, corroborating lines of evidence indicate the accuracy and usefulness of our microarray data. The genes are categorized according to the stage of fiber development that corresponds with the purported gene product function and/or highest transcript abundance (Tables 4, 5).

The design of qRT-PCR primers from the TC or EST represented on the cotton oligonucleotide microarray was performed following careful analysis of the target nucleotide sequence. Specifically, multiple alignments were performed on the target sequence and related sequences in the CGI 8.0 and/or CGI 9.0 databases in an attempt to design primers that would amplify only the target sequence and not homeologous sequences or other gene family members that may exhibit different expression profiles than the intended target sequence. The cotton microarray sequence IDs are listed in Table 4 with the BLASTx sequence annotations, GenBank accessions, TAIR accessions, and qRT-PCR primer sequences. Each of the selected genes, the logic behind their selection, and a comparison of transcript abundance estimated by microarray to transcript levels measured by qRT-PCR are discussed below.

Table 3 Enrichment analysis of gene ontologies up-regulated in Venn groups shown in Fig. 3 (significant at FDR-corrected $P < 0.05$)

Venn group	GO term	Biological process
I	GO:0030865	Cortical cytoskeleton organization and biogenesis
	GO:0031122	Cytoplasmic microtubule organization and biogenesis
	GO:0043622	Cortical microtubule organization and biogenesis
II	GO:0005982	Starch metabolic process
	GO:0009225	Nucleotide-sugar metabolic process
	GO:0016760	Cellulose synthase (UDP-forming) activity
	GO:0005984	Disaccharide metabolic process
	GO:0006011	UDP-glucose metabolic process
	GO:0009834	Cellulose and pectin-containing secondary cell wall biogenesis
III	GO:0004364	Glutathione transferase activity
	GO:0009696	Salicylic acid metabolic process
	GO:0010337	Regulation of salicylic acid metabolic process
	GO:0032350	Regulation of hormone metabolic process
	GO:0006749	Glutathione metabolic process
	GO:0016491	Oxidoreductase activity
IV	GO:0009793	Embryonic development ending in seed dormancy
	GO:0048316	Seed development
V	GO:0009225	Nucleotide-sugar metabolic process
	GO:0007047	Cell wall organization and biogenesis
	GO:0045229	External encapsulating structure organization and biogenesis
	GO:0009832	Cellulose and pectin-containing cell wall biogenesis
	GO:0009664	Cellulose and pectin-containing cell wall organization and biogenesis
	GO:0042546	Cell wall biogenesis
VI	GO:0000786	Nucleosome
	GO:0006334	Nucleosome assembly
	GO:0016841	Ammonia-lyase activity
	GO:0045548	Phenylalanine ammonia-lyase activity
	GO:0031497	Chromatin assembly
	GO:0065004	Protein–DNA complex assembly
VII	GO:0004022	Alcohol dehydrogenase activity
	GO:0005507	Copper ion binding

Genes related to SCW biosynthesis

Distinct sets of *CesA* genes are transcribed for PCW and SCW biosynthesis (reviewed in Taylor 2008). The high degree of sequence similarity among *CesA* genes, particularly in the conserved glycosyltransferase coding region led us to design primers to the hypervariable region (HVR) (Pear et al. 1996) of *CesA* genes selected from the microarray data (Table 4). The 16 DPA comparative microarray data (Fig. 2b, comparison 3) combined with microarray data from the 12 to 16 DPA time point comparisons in fibers of both cotton NILs (Fig. 2b, comparisons 7 and 12; Fig. 3) identified three *CesA* genes whose expression was significantly different. The *CesA*s were annotated by BLASTx searches as *G. hirsutum CesA2* (Pear et al. 1996), an ortholog of *ArabidopsisCesA7/irx3* (Taylor et al. 1999), and *G. hirsutum CesA3* (Laosinchai et al. 2000), designated in Table 4 as *GhCesA2*, *AtCesA7*, and *GhCesA3*, respectively. All three *CesA* genes followed a similar pattern of expression during fiber development from 8 to 24 DPA based on the microarray data, with higher transcript abundance (*GhCesA2* = 2.13-fold; *AtCesA7* = 3.40-fold; *GhCesA3* = 2.21-fold) in fibers of MD 52ne compared with fibers of MD 90ne at 16 DPA (Table 5). The qRT-PCR results for *GhCesA2* (TC66201.3) and *AtCesA7* (TC68655) supported the microarray data showing that both genes were more highly expressed at 16 DPA in fibers of MD 52ne (Table 5; Supplemental Fig. 2).

The microarray data for *AtCesA7* indicated significant differences in transcript abundance at 8 DPA and 12 DPA in fibers of the NILs; however, these results were not supported by qRT-PCR with primer pairs designed from the HVR-coding region of the nucleotide sequence. These results could indicate cross-hybridization on the array with

Table 4 Cotton tentative consensus sequences and ESTs selected for corroboration of microarray results

Sequence ID	Sequence description	GenBank accession	TAIR accession	Primers for qRT-PCR		References
				Forward (5′–3′)	Reverse (5′–3′)	
SCW biosynthesis stage						
TC66201.3.3	Cellulose synthase (GhCesA2)	AAB37767	AT5G44030	tcaccggttttcattgcctc	ttcaggaaggccaccatttt	Pear et al. (1996)
TC68655	Cellulose synthase (AtCesA7)	AAD32031	AT5G17420	taagcctagaagcagccgaaga	tcatgtgggacatcaataactcctt	Taylor et al. (1999)
U58283	Cellulose synthase (GhCesA1)	U58283	AT4G18780	tggactacccggtggataaggt	ctttcttgcaaagtcggctgtt	Pear et al. (1996)
TC59066	COBRA-LIKE4 (AtCOBL4)	NP_197067	AT5G15630	gcattcagcaattgtcggaa	ggcgagactcaggttctggtt	Roudier et al. (2002)
Cotton12_06695_01	Arabinogalactan protein 4 (GhAGP4)	ABR68796	AT5G60490	ttctcttctctcatgggctattga	tgcacgcacaaatcacaactt	Liu et al. (2008)
TC59194	Fasciclin-like arabinogalactan protein 3 (GhFLA3)	ABV27474	AT2G04780	cacggttgaaggcattacga	gccttgaatgcgctgtcttt	Huang et al. (2008)
Cotton12_06485_01	β-1,3-Glucanase-like protein	AAM66024	AT5G55180	tcggtcgggattaactacgg	ttccggtgacggtaagttgtt	Ruan et al. (2004)
AY291285	Chitinase-like (GhCTL1)	AAQ56598	AT3G16920	ggcagcactctggagatgga	gccgacggttggtgtttct	Zhang et al. (2004)
AI055116	Chitinase-like (GhCTL1)	AAQ56598	AT3G16920	aaacatgcatcctaactttcaggc	catcacggtgtcatcattcca	Zhang et al. (2004)
CD485826	Metallothionein-like protein	CAC39481	AT5G02380	aaaatgcactttgagggatctga	gcatccattctcaaccccaa	Jacob-Wilk et al. (2006)
TC59405	Metallothionein-like protein (GhMT1)	AAW47577	AT3G15353	gcctgtctgtcttatgtctgagtctc	tggccattccatcattgtgt	Jacob-Wilk et al. (2006)
Cotton12_15955_01	WAVE-DAMPENED2-LIKE1 (AtWDL1)	NP_850514	AT1G54460	tgcggtgatgtagtccactca	gcagcatgttttcgcctttt	Yuen et al. (2003)
Elongation stage						
AI727450	Cellulose synthase (GhCesA3)	AAD39534	AT5G05170			Laosinchai et al. (2000)
	CesA domain primers			aagaggcttggatggaattcaa	acacaccccgtacccacatag	
	HVR primers			ctcctctcaagcccaagcat	caaagcatgaagaaaacagccc	
TC59150	α-expansin 1 (GhExp1)	ABD48785	AT2G40610	cccccacgagaacactttgat	agggacgattccagctcgat	Harmer et al. (2002)
AY189968	Germin-like protein (GhGLP1)	AAF21988	AT5G20630	gcctcctccgtaccgca	tgttattgcagcttttacgaggttc	Kim and Triplett (2004b)
qRT-PCR reference	*G. hirsutum* 18S rRNA gene	U42827	AT3G41768	cgtccctgccctttgtaca	aacacttcaccggaccattca	

Table 5 Comparative microarray and qRT-PCR results for selected cotton genes

Sequence ID	8 DPA fold difference (MD 52ne–MD 90ne)		12 DPA fold difference (MD 52ne–MD 90ne)		16 DPA fold difference (MD 52ne–MD 90ne)		20 DPA fold difference (MD 52ne–MD 90ne)		24 DPA fold difference (MD 52ne–MD 90ne)	
	Microarray	qRT-PCR	Microarray	qRT-PCR	Microarray	qRT-PCR	Microarray	qRT-PCR	Microarray	qRT-PCR
SCW biosynthesis stage										
TC66201.3	–	–	–	–	2.13	2.55	–	–	–	–
TC68655	7.66	–	−2.22	–	3.40	3.50	–	–	–	–
U58283	–	–	–	–	–	3.50	–	–	–	–
TC59066	–	–	–	–	2.34	2.26	–	1.98	–	–
Cotton12_06695_01	2.69	–	−2.21	–	2.50	4.16	–	3.24	–	–
TC59194	–	–	–	2.02	2.11	2.61	–	4.52	–	–
Cotton12_06485_01	–	–	–	–	1.98	2.33	2.04	2.12	–	–
AY291285	–	–	–	–	2.02	3.47	–	3.75	–	–
AI055116	11.36	–	−2.99	–	2.01	–	–	–	–	–
CD485826	–	–	–	–	−2.66	–	2.80	–	−1.98[b]	–
TC59405	–	–	–	−2.53	−3.38	3.18	5.31	8.4	−2.83	−4.36
Cotton12_15955_01	–	–	–	–	2.03[a]	–	–	–	–	–
Elongation stage										
AI727450										
Microarray	–	–	–	–	2.21	–	–	–	–	–
CesA domain primers	–	–	–	–	–	–	–	–	–	–
HVR primers	–	–	–	–	–	–	–	–	–	–
TC59150	–	–	–	–	–	–	−7.71	−24.19	–	–
AY189968	–	–	–	–	−6.39	–	–	–	–	–

– No statistically significant data

[a] $P = 0.018$, [b] $P = 0.008$

other *CesA* genes related to PCW biosynthesis since SCW-related *CesA* transcripts are nearly undetectable during the fiber elongation stage, even using qRT-PCR. The microarray expression data for *GhCesA3* (AI727450) closely matched the expression profile of the other two *CesA* genes, *GhCesA2* and *AtCesA7*, and also fit the fiber gene expression profile of cotton *CesA*s involved in SCW biosynthesis during fiber development (Arpat et al. 2004; Gou et al. 2007; Pear et al. 1996); however, *GhCesA3* is slightly more abundantly expressed during the elongation stage of cotton fiber development and becomes down-regulated but remains constitutively expressed during the transition and SCW biosynthesis stages (Laosinchai et al. 2000).

In addition to the *CesA* genes identified for corroboration from the microarray data, our laboratory designed specific qRT-PCR primers for *G. hirsutum CesA1* for use in a previous study (Kim and Triplett 2007). We included *GhCesA1* (GenBank accession: U58283) in the qRT-PCR analysis as a positive control. As with the *GhCesA2* (TC66201.3) and *AtCesA7* (TC68655) genes, these qRT-PCR results revealed higher transcript abundance (3.5-fold) of *GhCesA1* in fibers of MD 52ne compared with fibers of MD 90ne at 16 DPA (Table 5; Supplemental Fig. 2).

TC59066 is highly identical in the predicted amino acid sequence (85% identity; 94% similarity) with the *Arabidopsis* COBRA-LIKE4 (COBL4) protein (Brady et al. 2007; Roudier et al. 2002). The microarray expression data for TC59066 mirrored the *GhCesA2* and *AtCesA7* ortholog expression profiles during fiber development, with a higher level of transcript (2.34-fold) present at 16 DPA in fibers of MD 52ne compared with fibers of MD 90ne (Table 5). These qRT-PCR results confirmed the transcript abundance of TC59066 to be 2.26-fold higher in fibers of MD 52ne compared with fibers of MD 90ne at 16 DPA and also 1.98-fold higher at 20 DPA (Table 5; Supplemental Fig. 2).

Two cotton TCs predicted to encode arabinogalactan proteins (AGPs), Cotton12_06695_01 and TC59194, were identified from our microarray analysis. The predicted protein sequence of Cotton12_06695_01 encodes the previously reported *G. hirsutum* AGP GhAGP4 that contains conserved fasciclin domains characteristic of fasciclin-like AGPs (FLAs) (reviewed in Johnson et al. 2003; Liu et al. 2008). The expression profile of Cotton12_06695_01 in fibers of both cotton NILs indicated low transcript levels during fiber elongation followed by a significant increase in expression beginning at the transition stage and a continued

increase in transcript abundance during the beginning of SCW biosynthesis. These data are consistent with the expression profile of *GhAGP4* in fibers of *G. hirsutum* and *G. barbadense* (Liu et al. 2008). The comparative analysis shown in Table 5 revealed statistically significant differences in Cotton12_06695_01 transcript abundance in fibers of MD 52ne compared with fibers of MD 90ne (2.7-fold higher at 8 DPA and 2.5-fold higher at 16 DPA). At 12 DPA the microarray data indicated 2.2-fold higher transcript abundance in fibers of MD 90ne compared with fibers of MD 52ne. The qRT-PCR data corroborated only the 16 DPA comparative data and revealed 4.2-fold higher transcript abundance in fibers of MD 52ne (Table 5; Supplemental Fig. 2). The qRT-PCR data also indicated 3.2-fold higher transcript abundance in fibers of MD 52ne compared with fibers of MD 90ne at 20 DPA (Table 5; Supplemental Fig. 2).

The BLASTx results for TC TC59194 indicate 100% sequence identity to a gene product of unknown function encoded by a *G. hirsutum* gene designated FB-B6 that is preferentially expressed in fibers (John 1995). The BLASTx results with the next highest identities are the *G. hirsutum* fasciclin-like arabinogalactan proteins 2 and 3 (GhFLA2 and GhFLA3) (Huang et al. 2008) that share 62% identity with the TC59194 predicted protein sequence. The *GhFLA3* gene is preferentially expressed in roots, whereas *GhFLA2* is preferentially expressed in fibers with transcript abundance gradually increasing from 2 DPA to 20 DPA (Huang et al. 2008). The microarray expression profile of TC59194 is similar to Cotton12_00695_01 in fibers of both NILs with transcript abundance increasing at the transition stage and into the SCW biosynthesis stage. Based on the comparative microarray data, transcript abundance of TC59194 is 2.1-fold higher in fibers of MD 52ne compared with fibers of MD 90ne at 16 DPA (Table 5; Supplemental Fig. 2). The qRT-PCR data corroborated this result (2.6-fold) and also revealed a higher transcript abundance at 12 and 20 DPA (2.0- and 4.5-fold, respectively) in fibers of MD 52ne compared with fibers of MD 90ne (Table 5; Supplemental Fig. 2). The expression profiles of both AGP genes were slightly delayed in fibers of MD 90ne compared with MD 52ne.

The list of genes selected for corroboration of the microarray data included a β-1,3-glucanase-like gene (Cotton12_06485_01) that is more abundantly expressed at 16 and 20 DPA (1.98 and 2.04-fold, respectively) in fibers of MD 52ne compared with fibers of MD 90ne. The predicted amino acid sequence of TC Cotton12_06485_01 shares 72% identity and 83% similarity to an *Arabidopsis* β-1,3-glucanase-like protein (GenBank accession: AAM66024). The microarray expression data were validated by qRT-PCR and are shown in Table 5 and Supplemental Fig. 2.

Two chitinase-like genes, AY291285 (*GhCTL1*) and AI055116, were identified from the enrichment analysis. AY291285 was previously shown to be more abundantly expressed in cotton fiber cells in the SCW synthesis stage of development (Zhang et al. 2004). An alignment of AY291285 and AI055116 revealed that the microarray probe sequences for both ESTs were designed from highly identical regions of the nucleotide sequences that could result in cross-hybridization on the microarray. The AI055116 microarray probe is a 70-mer oligonucleotide that is 100% identical in both cotton nucleotide sequences and the AY291285 microarray probe is a 60-mer oligonucleotide that is 95% identical to the AI055116 EST. Sequence-specific qRT-PCR primers were designed from a more divergent region of the aligned sequences in order to validate the microarray expression profiles of the *CTL* genes. The qRT-PCR data confirmed that the transcript abundance of AY291285 was higher (3.47-fold) in fibers of MD 52ne at 16 DPA compared with fibers of MD 90ne at 16 DPA (Table 5; Supplemental Fig. 2). Furthermore, the increased transcript abundance of AY291285 persisted in fibers of MD 52ne compared with fibers of MD 90ne at 20 DPA (3.75-fold). The qRT-PCR analysis of AI055116 indicated that the gene was not expressed at any time point in fibers of either germplasm line. The GenBank accession data for AI055116 revealed that the EST was generated from *G. hirsutum* boll abscission zone tissues. This finding suggests that the *GhCTL1* transcript, represented by AY291285, hybridized to both the AY291285 and AI055116 microarray probes and that AI055116 is expressed in a tissue-specific manner, but not in fibers of either cotton NIL at the time points chosen for this study.

Two cotton nucleotide sequences, EST CD485826 and TC TC59405, encoding MET proteins were selected from the distilled list of selected genes for corroboration of the microarray expression profiles. The coding sequence for TC59405 shares 91% identity to a *G. hirsutum* metallothionein (MET) gene (GenBank accession: AY857933) designated *Ghmt1*. The coding sequence for CD485826 shares 76% identity with the MET gene *Ghmet-1* reported by Jacob-Wilk et al. (2006). The microarray expression profile for both genes indicates higher transcript levels in fibers of MD 90ne at 16 DPA (CD485826: 2.66-fold and TC59405: 3.38-fold) and higher transcript levels in fibers of the MD 52ne at 20 DPA (CD485826: 2.80-fold and TC59405: 5.31-fold), and then once again higher transcript levels in fibers of MD 90ne at 24 DPA (CD485826: 1.98-fold and TC59405: 2.83-fold) (Table 5). These qRT-PCR results for TC59405 did not corroborate the microarray expression data for the 16 DPA time point; however, the microarray data for 20 and 24 DPA were corroborated. The design of primers for CD485826 proved to be more difficult; dissociation data from the qRT-PCR indicated amplification of

cDNA from more than one gene; therefore corroboration of the microarray data was not possible.

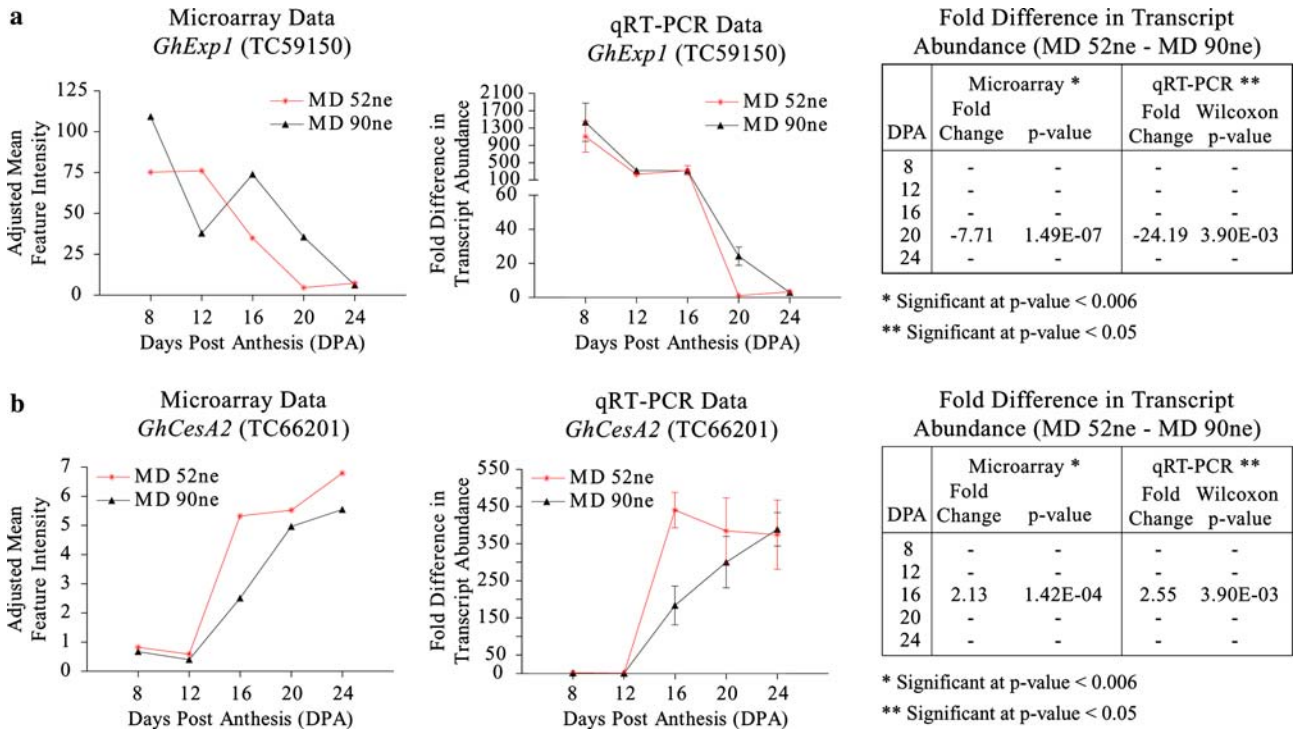
The cotton PAVE TC Cotton12_15955_01 was initially selected based on preliminary statistical testing with a Student's *t* test prior to completion of the entire microarray experiment. The *t* test results indicated that the Cotton12_15955_01 transcript was several-fold more abundant in fibers of MD 52ne compared with fibers of MD 90ne at the 16 DPA time point. In terms of functionality, Cotton12_15955_01 shares 38% identity in the predicted protein sequence with the *Arabidopsis* WAVE-DAMPENED2-LIKE1 (AtWDL1) protein (Yuen et al. 2003). In the C-terminal region of the protein, the Cotton12_15955_01 predicted protein sequence is highly identical to AtWDL1 and a paralogous sequence WAVE-DAMPENED2 (AtWVD2) The C-terminal region contains a conserved domain found in the vertebrate microtubule (MT) associated protein (MAP) Targeting Protein for Xklp2 (TPX2) (Wittmann et al. 2000). Once the entire microarray dataset was assembled and a more thorough statistical analysis conducted as described in the methodology, Cotton12_15955_01 was no longer significant based on the final criteria for statistical significance. The *P* value placed on the comparative 16 DPA microarray data for Cotton12_15955_01 was 0.018, whereas the cut-off for significance was a more stringent $P < 0.006$ (Table 5). However, since qRT-PCR primers were already designed from the nucleotide sequence, the expression profile of this potentially interesting gene was tested in both NILs. The qRT-PCR data revealed that the Cotton12_15955_01 transcript abundance was, in fact, higher (3.2-fold) in fibers of MD 52ne compared with fibers of MD 90ne at 16 DPA (Table 5). The expression over time of Cotton12_15955_01 in fibers of both cotton NILs followed a pattern similar to genes up-regulated during SCW biosynthesis (Supplemental Fig. 2).

Genes related to cotton fiber cell elongation

The EST AI727450 shares 84% identity with *GhCesA3* (GenBank accession: AAD39534) in the predicted protein sequence, which includes a portion of the conserved glycosyltransferase domain and the HVR domain. The microarray probe sequence designed for AI727450 is located in the glycosyltransferase encoding region of the EST. Two sets of qRT-PCR primers were designed for AI727450, one pair from the probe region of the EST and the other from the HVR-coding region of the EST. Both primer pairs showed sequence specificity for the target sequence based on dissociation curves generated during qRT-PCR and also revealed nearly the same expression pattern during fiber development for both cotton NILs (Supplemental Fig. 2). The pattern of AI727450 transcript abundance matches the expression profile for *GhCesA3* as reported by Laosinchai et al. (2000), with mostly constitutive expression throughout fiber development and slightly higher expression during fiber elongation. Together, the sequence similarity and expression data indicate that AI727450 is a homeolog or paralog of *GhCesA3*. Comparison of the microarray expression data of AI727450 with the corresponding qRT-PCR expression data (Supplemental Fig. 2) indicates possible cross-hybridization to the microarray probe sequence with one or more *CesA*s that are more abundantly expressed during the SCW synthesis stage of cotton fiber development.

As previously stated, the *G. hirsutum* α-expansin 1 (*GhExp1*) gene was initially chosen as a control candidate gene for qRT-PCR to further validate the microarray results. The *GhExp1* gene (GenBank accession: AF512539) is represented on the cotton oligonucleotide microarray by TC59150. The 16 DPA comparative microarray data indicated no difference in *GhExp1* expression (Fig. 2b, comparison 3; Table 5) in fibers of the NILs, a result that was verified by qRT-PCR (Table 5; Supplemental Fig. 2); however, the comparative data from the 20 DPA time point (Fig. 2b, comparison 4) revealed a 7.71-fold higher transcript abundance in MD 90ne compared with the level of *GhExp1* transcripts in MD 52ne (Table 5). Verification of the microarray data by qRT-PCR revealed the difference in *GhExp1* transcript levels to be much more pronounced, with a 24.2-fold higher abundance in fibers of MD 90ne (Table 5; Supplemental Fig. 2). Another α-expansin EST that was not used for validation of the microarray data (BG442643) was also up-regulated at 20 DPA in fibers of MD 90ne (5.6-fold) compared with fibers of MD 52ne at 20 DPA. Based on BLASTx results, the predicted protein sequence of BG442643 shares 82% identity with *G. hirsutum α-expansin 4* (Harmer et al. 2002). RT-PCR data presented by Harmer et al. (2002) indicated that *GhExp1* and *GhExp4* share similar expression profiles in developing cotton fibers, but *GhExp4* is expressed at much lower levels than *GhExp1*. Since a number of genes related to fiber transition and SCW biosynthesis stages were already analyzed in detail, we focused on genes related to fiber elongation to further support the germplasm line-specific differential expression data for α-expansin 1 (*Ghexp1*) (Fig. 4; Table 5).

The cotton nucleotide sequence AY189968 encodes the *G. hirsutum* germin-like protein1 (GhGLP1). The *GhGLP1* transcript was previously shown to be fiber-specific and more abundant in *G. hirsutum* elongating fibers and then expressed at a much lower basal level during the SCW biosynthesis stage (Kim and Triplett 2004b). The microarray expression profile of *GhGLP1* was validated in silico in fibers of both cotton NILs, with a transcript being more abundant during fiber elongation and less abundant during the transition stage and SCW biosynthesis. In addition, the



DPA	Microarray * Fold Change	Microarray * p-value	qRT-PCR ** Fold Change	qRT-PCR ** Wilcoxon p-value
8	-	-	-	-
12	-	-	-	-
16	-	-	-	-
20	-7.71	1.49E-07	-24.19	3.90E-03
24	-	-	-	-

* Significant at p-value < 0.006

** Significant at p-value < 0.05

DPA	Microarray * Fold Change	Microarray * p-value	qRT-PCR ** Fold Change	qRT-PCR ** Wilcoxon p-value
8	-	-	-	-
12	-	-	-	-
16	2.13	1.42E-04	2.55	3.90E-03
20	-	-	-	-
24	-	-	-	-

* Significant at p-value < 0.006

** Significant at p-value < 0.05

Fig. 4 Microarray and qRT-PCR expression profiles of two representative genes. **a**, **b** The *left panels* present the microarray fiber gene expression profiles, back-transformed from the $\log_2$ scale, of both cotton NILs over time. The *middle panels* are the qRT-PCR data used to validate the microarray fiber gene expression profiles that were tested for statistical significance using a Wilcoxon signed-rank test. The *right panels* indicate only the statistically significant and biologically significant (>2-fold difference) data for each time point comparison. A positive fold change value indicates higher transcript abundance in fiber of MD 52ne, and a negative fold change value indicates higher transcript abundance in fibers of MD 90ne. **a***G. hirsutum α-expansin 1* (*GhExp1*) cotton gene chip ID TC59150, **b***G. hirsutum cellulose synthase 2* (*GhCesA2*) cotton gene chip ID TC66201.3

microarray results indicated the *GhGLP1* transcript was 6.4-fold higher in fibers of MD 90ne compared with fibers of MD 52ne at 16 DPA (Table 5). Gene-specific qRT-PCR primers for *GhGLP1* confirmed the temporal pattern of *GhGLP1* expression in fibers of both cotton NILs (Supplemental Fig. 2); however, no significant difference in *GhGLP1* transcript was evident in fibers of MD 52ne compared with fibers of MD 90ne.

Expression profiles of selected genes from multiple growing seasons

In addition to corroborating the microarray data, several genes were selected for qRT-PCR analyses in subsequent growing seasons with additional time points as described in the methodology. These genes were *GhCesA2*, *AtCesA7*, *AtCOBL4*, *GhExp1*, and *AtWDL1*. Preliminary data from the first growing season of this study were included for the genes that were analyzed in that year, *GhCesA1*, *GhCesA2*, and *GhExp1*, with the time points limited to 12, 16, and 20 DPA. In the case of genes that continue to up-regulate starting at the transition stage and into the SCW biosynthesis stage of cotton fiber development, the pattern of MD 52ne fibers entering the transition stage prior to fibers MD 90ne was evident for all years except the 2007 season (Supplemental Fig. 1). In 2003, the fibers of MD 52ne appeared to be well into the transition stage and entering the SCW biosynthesis stage at 12 DPA as indicated by the large fold increase in transcript abundance of *GhCesA1* and *GhCesA2* in fibers of MD 52ne compared with fibers of MD 90ne. In the 2007 season, the transition stage was much later in the fibers of both cotton NILs compared to the other seasons, with the largest increases in transcripts of SCW biosynthesis-related genes occurring after 18 DPA. Although the observed differences in the expression levels of *GhExp1* between fibers of MD 52ne and MD 90ne from the microarray data were corroborated by qRT-PCR (Table 5; Supplemental Fig. 2), there were no significant differences in *GhExp1* expression levels in prior or subsequent years of the study (Supplemental Fig. 1).

Discussion

The microarray and qRT-PCR analysis of selected genes suggests that there is a difference in the temporal shift in biological processes leading to the SCW stage of cotton fiber development between the cotton NILs. This finding

was first evidenced by monitoring the expression profiles of genes (*GhCesA1* and *GhCesA2*) that are indicators of the SCW biosynthesis stage in fibers of the cotton NILs. These observations served as the basis for our hypothesis that the fibers of MD 52ne were entering the transition and SCW biosynthesis stages of development earlier than the fibers of MD 90ne. These preliminary data were supported in subsequent years of the study by comparative global gene expression analysis and qRT-PCR of selected genes (Supplemental Fig. 1). The environmental effect on the timing of the transition stage and the beginning of SCW biosynthesis during cotton fiber development were evident in each year of the study, particularly in the 2003 and 2007 seasons (Supplemental Fig. 1). The effect of temperature on cotton fiber development is well-documented and indicates that cooler temperatures, especially at night, cause a decrease in the rate of fiber elongation and an increase in the duration of the elongation stage of fiber development (Gipson and Joham 1969; Gipson and Ray 1969; Thaker et al. 1989). These observations have also been confirmed in vitro using cotton ovule cultures (Haigler et al. 1990. We hypothesize that lower night temperatures during boll/fiber development in the 2007 season delayed the transition period. Had we sampled beyond 24 DPA in 2007, the expression profiles for selected genes in different years may have been more similar. Significant environmental effects on the relative duration of fiber developmental stages also precluded direct comparisons of microarray results between years.

In looking at the microarray expression profiles of genes related to fiber cell elongation and SCW biosynthesis, the interval between collection time points of immature fiber samples (4 days) was not sufficient to reveal a clear distinction between the NILs throughout fiber development. For example, the expression of genes related to SCW biosynthesis is significantly up-regulated from 12 to 16 DPA in fibers of both NILs. Likewise, the expression of genes related to fiber cell elongation, such as expansins, show similar expression profiles over time in fibers of both NILs. The comparative aspect of this study allowed a more thorough analysis of the transcript abundance of selected genes, and allowed us to make inferences into the biological processes that differed in fibers of MD 52ne and MD 90ne over time. Group V of the Venn diagram in Fig. 3b represents genes that are up-regulated in fibers of MD 52ne from 12 to 16 DPA and up-regulated in fibers of MD 52ne compared with fibers of MD 90ne at 16 DPA. These genes are enriched in biological processes related to SCW biosynthesis (Table 3; Supplemental Table 2), suggesting the fibers of MD 52ne have entered the SCW stage of development. Conversely, Group VII of the Venn diagram in Fig. 3c represents genes that are up-regulated in fibers of MD 90ne from 12 to 16 DPA and up-regulated in fibers of MD 90ne compared with fibers of MD 52ne at 16 DPA. The genes in this Venn group are significantly enriched in biological processes related to oxidoreductase activity (Supplemental Table 2). Since the Venn Group V in Fig. 3b also includes genes involved in cellulose and cell wall biosynthesis up-regulated in fibers of MD 90ne from 12 to 16 DPA, it is likely that the transition to SCW biosynthesis is slightly delayed in MD 90ne fibers.

The microarray dataset revealed a large number of aquaporins were more abundantly expressed at 20 DPA in fibers of MD 90ne compared to fibers of MD 52ne (GEO accession GSE14388), supporting the notion that fibers of MD 90ne are delayed in entering the transition stage and may also have a more prolonged elongation stage. The rapid polar expansion of cotton fiber cells occurring from approximately 5 to 20 DPA requires an increase in fiber intracellular turgor pressure. Dhindsa et al. (1975) demonstrated in vitro and in vivo that the increased turgor pressure within the fiber cell coincides with an increase in the accumulation of potassium (K^+) and malate, presumably acting as osmoregulators to decrease the osmotic potential and cause water to move into the fiber cell from the maternal tissue. This observation corresponds with the up-regulation of the β-1,3-glucanase (Cotton12_06485_01) gene in fibers of MD 52ne compared with fibers of MD 90ne at 16 and 20 DPA (Table 5; Supplemental Fig. 2). Ruan et al. (2001) presented evidence that the elongation stage of cotton fiber development coincides with the closure of plasmodesmata in the membrane interface connecting the fiber cell to the maternal tissue. This plasmodesmatal closure coincides with callose deposition at the base of the fiber, presumably sealing the membrane interface between the fiber and the maternal tissue, thus preventing an influx of solutes and maintaining tugor pressure within the fiber cell (Ruan et al. 2004). Conversely, the degradation of callose at the fiber base coincided with plasmodesmatal opening and up-regulation of the *GhGluc1* that encodes a callose-degrading β-1,3-glucanase enzyme (Ruan et al. 2004). Based on the literature, the observed up-regulation of β-1,3-glucanase expression levels in fibers of MD 52ne compared to fibers of MD 90ne at 16 and 20 DPA presents convincing evidence that fibers of MD 52ne have entered the transition stage leading into SCW biosynthesis. This finding supports our hypothesis of a temporal difference in fiber development between the NILs.

The apparent temporal difference in fiber development does not seem to extend to other aspects of boll development since there is no difference in the maturation rate of bolls in the NILs for the years covered by this study or in years prior to the germplasm line release (Meredith 2005a). The micronaire of mature fibers from both NILs showed no statistical difference, indicating no difference in fiber fineness (Table 1). Micronaire is a fiber quality measurement that is critical in determining the market value of cotton.

It is measured by compressing a constant mass of cotton fibers to a fixed volume and then measuring the air permeability of the fibers. This measures the surface area of the cotton fibers and gives an indication of the individual fiber thickness, which is interpreted as fineness. Micronaire is also considered as an indirect measure of fiber maturity; however, micronaire measures only fiber surface area. Initial studies conducted on the NILs revealed a slightly higher fiber maturity when measured by more direct means using the Advance Fiber Information System (AFIS); however, this was not evident in subsequent years of the study (Meredith 2005b). We also attempted to assess fiber maturity by means of cross-section analysis of fiber cells and direct measurements of fiber perimeter and SCW thickness, but no statistical difference was evident in fibers of MD 52ne compared with fibers of MD 90ne (data not shown).

A more mature fiber with a thicker SCW does not explain the increased fiber-bundle strength of MD 52ne since fewer fibers would compose the beard used in the stelometer measurement. Fibers of *G. barbadense* cottons are typically higher in fiber-bundle strength than fibers of *G. hirsutum* cottons, a fact that is partially explained by *G. barbadense* having finer fibers and consequently more fibers by weight. An extreme example of more mature fibers with a thicker SCW that have lower fiber-bundle strength are fibers of the diploid cotton species *G. arboreum*. The mature fibers of *G. arboreum* have a SCW nearly twice the thickness of the mature fibers of *G. hirsutum* and *G. barbadense*, but fiber-bundle strength is 44–63% lower and single fiber strength is 9–13% lower in *G. arboreum* fibers (Benedict et al. 1999). These results suggest that no correlation exists between the amount of cellulose deposited in the SCW and fiber strength, meaning that other factors such as cellulose MF orientation and/or degree of polymerization have a greater effect on the trait.

Previous studies of late-stage expression profiling in cotton fiber development have focused mainly on a single cotton germplasm line or cultivar, with comparative analysis between developmental time points and/or tissue types (Arpat et al. 2004; Gou et al. 2007; Hovav et al. 2008; Li et al. 2002; Shi et al. 2006; Tu et al. 2007; Wu et al. 2007). The microarray data compiled in our study present an analysis of genes expressed over a broad range of developmental time points and comparatively between fibers of two cotton NILs. This experimental approach revealed interesting similarities in our list of selected genes (Table 4) with those related to SCW biosynthesis in heterologous systems. In three separate studies that identified genes highly correlated with cellulose and SCW biosynthesis in *Arabidopsis* vascular tissues (Brown et al. 2005; Ko et al. 2006; Persson et al. 2005), there was significant overlap with genes selected from our microarray analysis that were more highly expressed in fibers of MD 52ne compared with fibers of MD 90ne at the 16 DPA transition stage of fiber development (Tables 4, 5). The BLASTx results from the TAIR database were used to compare the results of our analysis with previous studies (Table 4). The following sequences selected from our study are found in gene lists that correlate with SCW biosynthesis in *Arabidopsis* vascular tissues (Brown et al. 2005; Ko et al. 2006; Persson et al. 2005): the *ArabidopsisCesA7* ortholog encoded by TC68655; the *ArabidopsisCOBL4* ortholog encoded by TC59066; the *G. hirsutum* arabinogalactan protein 4 *GhAGP4* encoded by Cotton12_06695_01, which is highly similar to *Arabidopsis* fasciclin-like arabinogalactan proteins *FLA11* and *FLA12*; and the *G. hirsutum* cellulose synthase *CesA2* that is orthologous to *ArabidopsisCesA4*. The *G. hirsutum* chitinase-like1 protein GhCTL1 encoded by AY291285 is also found in the SCW biosynthesis correlated gene lists generated in two of the *Arabidopsis* studies (Brown et al. 2005; Persson et al. 2005). The overlap of genes identified in this study with SCW-related *Arabidopsis* genes supports the notion of utilizing *Arabidopsis* as a heterologous system for functional analysis of cotton genes implicated in fiber SCW biosynthesis.

Genes that were selected from the microarray data that were not directly correlated with SCW biosynthesis could encode proteins that function to improve the strength of cotton fibers. Two of these genes are of interest, *Arabidopsis* homologs named *AtWVD2* and *AtWDL1*. Overexpression of *AtWVD2* or *AtWDL1* under the control of the cauliflower mosaic virus (CaMV) 35S promoter in elongating *Arabidopsis* root cells resulted in a change in root skewing from rightward to leftward, shorter roots with increased diameter, and a higher density of root hairs (Yuen et al. 2003). The *AtWVD2* overexpression phenotype is accompanied by alteration of cortical MT orientation from transverse to longitudinal in epidermal cells of the root elongation zone (Yuen et al. 2003). AtWDL1 and AtWVD2 were also shown to localize to cortical MTs in multiple tissue types of *Arabidopsis* and AtWVD2 was shown to bind to and bundle polymerized MTs in vitro (Perrin et al. 2007). In developing cotton fiber cells, layers of cellulose MFs are deposited beginning at the transition stage and during the SCW biosynthesis stage. The orientation of the cellulose MFs becomes more longitudinal with each successive layer and this change is mirrored by cortical MT orientation in cotton ovules grown in vitro (Seagull 1986; Yatsu and Jacks 1981). The cortical MTs of SCW-synthesizing fiber cells also change gyre concurrent with the reversal points of the cellulose MFs (Seagull 1986; Yatsu and Jacks 1981). This observation suggests that an association may exist between cortical MTs and cellulose MFs and possibly that cortical MTs direct cellulose MF deposition. Since cellulose MF orientation is correlated with fiber strength, the cotton *AtWDL1* ortholog is of particular interest

as a potential determinate of higher fiber strength. Although there is no direct evidence that cortical MTs direct cellulose MF orientation in cotton fiber cells, real-time fluorescent imaging results presented by Paredez et al. (2006) suggests that there is interaction between cortical MTs and cellulose synthase 6 (AtCesA6) in cells of etiolated *Arabidopsis* hypocotyls. More recent evidence using live cell imaging of intact *Arabidopsis* roots supports a similar role for cortical MTs and SCW cellulose synthase complexes (Wightman and Turner 2008).

The selected genes that encode MET proteins merit further investigation based on their suggested role in regulating the oxidative state and stability of GhCesA1 (Jacob-Wilk et al. 2006), and their reported linkage to a known cotton fiber-bundle strength QTL (Park et al. 2005). The CD485826 coding sequence was very similar to the published *Ghmet-1* sequence, but also shares significant homology with two ESTs used to develop four EST-SSR (simple sequence repeat) markers and one EST-CSR (complex sequence repeat) marker that mapped within close proximity (~3 to 4 cM) to a fiber-bundle strength QTL on chromosome 15 of the D-subgenome of tetraploid cotton (Park et al. 2005). The EST-SSR markers are designated MUCS141, MUCS152, MUCS410 and MUCS422 and the EST-CSR marker is designated MUSS598. The ESTs from which they were designed, CON_006_03864 and CON_018_04414, can be found in the Cotton Marker Database (Blenda et al. 2006) by querying the SSR database. A more recent genetic linkage map of tetraploid cotton, with greater resolution and highly enriched in markers representing expressed genes, revealed two more *met* genes in addition to the ones reported by Park et al. 2005, with GenBank accessions DR044136 and CO124976 (Guo et al. 2007). The four mapped *met* genes all encode Type II METs and share significant nucleotide and protein sequence homology (CON_006_03864 and CON_018_04414 are 100% identical) with the cotton microarray *met* EST CD485826 and *Ghmet-1*. The *met* genes map to a 4.6 cM region of chromosome 15; however, the study did not include QTL analysis, so the presence of a stable fiber-bundle strength QTL near the *met* gene cluster on chromosome 15 was not confirmed in the mapping population of Guo et al. (2007).

The results of this study suggest that variations in the mechanism controlling the temporal pattern of fiber development can impact mature fiber-bundle strength and other fiber quality parameters. It is unclear if one or more of the genes identified in this study, through individual function or acting synergistically, are ultimately responsible for the increased fiber-bundle strength observed in fibers of MD 52ne compared with fibers of MD 90ne. In *Arabidopsis*, insertion mutations in SCW cellulose synthase genes are known to cause structural weaknesses in vascular bundles presumably due to cellulose deficiency in the SCW of vascular tissues (Brown et al. 2005; Persson et al. 2005; Taylor et al. 1999). Also, an insertion mutation in the *Atcobl4* gene results in irregularly formed xylem that is deficient in cellulose (irregular xylem 6 or *irx6*) (Brown et al. 2005). Similar mutant phenotypes resulting in reduced mechanical strength of stem tissues are found in the rice *brittle culm2* mutant (*Osbc1*) (Li et al. 2003) and maize *brittle stalk-2* mutant (*Zmbk2*) (Ching et al. 2006), and are the result of mutations in *COBL4* orthologs. The effect of knockout mutations or overexpression of orthologous genes in cotton fibers is unknown, but their relation to mechanical strength in other plant systems makes them likely candidates for functional analysis in cotton fiber.

The genetic studies conducted by Meredith (2005b) indicated that 1.23 ± 0.16 genes control the increased fiber-bundle strength in MD 52ne compared with MD 90ne. This result indicates that a single gene or a genetic locus may result in the higher strength phenotype. The genes identified in this study are coordinately regulated in a temporal and tissue-specific manner and, therefore, likely controlled by a common, global regulator such as a transcription factor. At present, the identity of such a transcription factor in cotton is unknown; however, the involvement of NAC domain-containing transcription factors in SCW formation of *Arabidopsis* xylem cells is well-documented (Mitsuda et al. 2005, 2007; Yamaguchi et al. 2008). Kubo et al. (2005) presented convincing evidence that overexpression of the NAC domain-containing transcription factor *VASCULAR-RELATED NAC-DOMAIN7* (*VND7*) triggers SCW differentiation in multiple tissue types of *Arabidopsis*. Similar results were observed in tobacco BY-2 cells with *VND7* expressed under the control of an inducible promoter (Yamaguchi et al. 2008).

The prospect of a single gene controlling the temporal pattern of fiber developmental that results in higher fiber-bundle strength is appealing since the phenotype may be more readily adapted using a transgenic approach. Once identified, a single gene could also be easily tracked in a conventional breeding program using marker-assisted selection (MAS) with gene-specific PCR primers that would forego the regulatory complications that exist with transgenic crop plants. Breeding selection for a single gene or a genetic locus can also be efficiently carried out using a smaller population than what is required for a more quantitative trait. A molecular marker(s) within or in close proximity to a tightly linked genetic locus containing genes responsible for the higher fiber-bundle strength phenotype would serve the same purpose as a gene-specific marker and allow the phenotype to be tracked in a breeding program.

Acknowledgments The authors thank the staff of the United States Department of Agriculture-Agricultural Research Service, Crop Genetics and Production Research Unit for their technical expertise in conducting the field work. We also thank Jeff Cary, Jay Shockey, Josh Udall, Jonathan Wendel, Sam Yang, and two anonymous reviewers for their helpful comments during the preparation of this manuscript. This research was funded by United States Department of Agriculture-Agricultural Research Service project 6435-21000-015-00 and by a grant from the National Science Foundation Plant Genome Research Project—Award Number 0624077.

References

Arpat AB, Waugh M, Sullivan JP, Gonzales M, Frisch D, Main D, Wood T, Leslie A, Wing RA, Wilkins TA (2004) Functional genomics of cell elongation in developing cotton fibers. Plant Mol Biol 54:911–929

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–29

Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Edgar R (2007) NCBI GEO: mining tens of millions of expression profiles—database and tools update. Nucleic Acids Res 35(Database issue):D760–D765

Basra AS, Malik CP (1984) Development of the cotton fiber. Int Rev Cytol 89:65–113

Benedict CR, Kohel RJ, Lewis HL (1999) Cotton fiber quality. In: Smith CW, Cothren JT (eds) Cotton: origin, history, technology, and production. Wiley, New York, pp 269–288

Blenda A, Scheffler J, Scheffler B, Palmer M, Lacape JM, Yu JZ, Jesudurai C, Jung S, Muthukumar S, Yellambalase P, Ficklin S, Staton M, Eshelman R, Ulloa M, Saha S, Burr B, Liu S, Zhang T, Fang D, Pepper A, Kumpatla S, Jacobs J, Tomkins J, Cantrell R, Main D (2006) CMD: a cotton microsatellite database resource for *Gossypium* genomics. BMC Genomics 7:132

Blüthgen N, Brand K, Cajavec B, Swat M, Herzel H, Beule D (2005) Biological profiling of gene groups utilizing Gene Ontology. Genome Inform 16:106–115

Bowman DT, Gutierrez OA, Percy RG, Calhoun DS, May OL (2006) Pedigrees of upland and pima cotton cultivars released between 1970 and 2005. Mississippi Agricultural and Forestry Experiment Station Bulletin 1155

Brady SM, Song S, Dhugga KS, Rafalski JA, Benfey PN (2007) Combining expression and comparative evolutionary analysis. The COBRA gene family. Plant Physiol 143:172–187

Brown DM, Zeef LA, Ellis J, Goodacre R, Turner SR (2005) Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics. Plant Cell 17:2281–2295

Calhoun DS, Bowman DT, May OL (1994) Pedigrees of upland and pima cotton cultivars released between 1970 and 1990. Mississippi Agricultural and Forestry Experiment Station Bulletin 1017

Ching A, Dhugga KS, Appenzeller L, Meeley R, Bourett TM, Howard RJ, Rafalski A (2006) *Brittle stalk 2* encodes a putative glycosylphosphatidylinositol-anchored protein that affects mechanical strength of maize tissues by altering the composition and structure of secondary cell walls. Planta 224:1174–1184

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676

Dhindsa RJ, Beasley CA, Ting IP (1975) Osmoregulation in cotton fiber. Plant Physiol 56:394–398

Dudoit S, Yang YH, Callow MJ, Speed TP (2002) Statistical methods for identifying genes with differential expression in replicated cDNA microarray experiments. Stat Sin 12:111–139

Gipson JR, Joham HE (1969) Influence of night temperature on growth and development of cotton (*Gossypium hirsutum* L.) III. Fiber elongation. Crop Sci 9:127–129

Gipson JR, Ray LL (1969) Fiber elongation rates in five varieties of cotton (*Gossypium hirsutum* L.) as influenced by night temperature. Crop Sci 9:339–341

Gish W, States DJ (1993) Identification of protein coding regions by database similarity search. Nat Genet 3:266–272

Gou JY, Wang LJ, Chen SP, Hu WL, Chen XY (2007) Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. Cell Res 17:422–434

Guo W, Cai C, Wang C, Han Z, Song X, Wang K, Niu X, Wang C, Lu K, Shi B, Zhang T (2007) A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in Gossypium. Genetics 176:527–541

Haigler CH, Rao NR, Roberts EM, Huang JY, Upchurch DR, Trolinder NL (1990) Cultured ovules as models for cotton fiber development under low temperatures. Plant Physiol 95:89–96

Haigler CH, Zhang D, Wilkerson CG (2005) Biotechnological improvement of cotton fiber maturity. Physiol Plant 124:285–294

Harmer SE, Orford SJ, Timmis JN (2002) Characterisation of six α-expansin genes in *Gossypium hirsutum* (upland cotton). Mol Genet Genomics 268:1–9

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6:65–70

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

Huala E, Dickerman A, Garcia-Hernandez M, Weems D, Reiser L, LaFond F, Hanley D, Kiphart D, Zhuang J, Huang W, Mueller L, Bhattacharyya D, Bhaya D, Sobral B, Beavis B, Somerville C, Rhee SY (2001) The *Arabidopsis* Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res 29:102–105

Huang GQ, Xu WL, Gong SY, Li B, Wang XL, Xu D, Li XB (2008) Characterization of 19 novel cotton FLA genes and their expression profiling in fiber development and in response to phytohormones and salt stress. Physiol Plant 134:348–359

Jacob-Wilk D, Kurek I, Hogan P, Delmer DP (2006) The cotton fiber zinc-binding domain of cellulose synthase A1 from *Gossypium hirsutum* displays rapid turnover in vitro and in vivo. Proc Natl Acad Sci USA 103:12191–12196

John ME (1995) Characterization of a cotton (*Gossypium hirsutum* L.) fiber mRNA (Fb-B6). Plant Physiol 107:1477–1478

Johnson KL, Jones BJ, Bacic A, Schultz CJ (2003) The fasciclin-like arabinogalactan proteins of *Arabidopsis*. A multigene family of putative cell adhesion molecules. Plant Physiol 133:1911–1925

Jung S (2008) Positioning cotton in the market for quality: an application of needs assessment for west Texas. In: Proceedings of the 2008 Beltwide Cotton Conferences. National Cotton Council of America, Memphis, TN, USA, pp 369–381

Kim HJ, Triplett BA (2001) Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. Plant Physiol 127:1361–1366

Kim HJ, Triplett BA (2004a) Characterization of GhRac1 GTPase expressed in developing cotton (*Gossypium hirsutum* L.) fibers. Biochim Biophys Acta 1679:214–221

Kim HJ, Triplett BA (2004b) Cotton fiber germin-like protein. I. Molecular cloning and gene expression. Planta 218:516–524

Kim HJ, Triplett BA (2007) Cellulose synthase catalytic subunit (*Cesa*) genes associated with primary or secondary wall biosynthesis in developing cotton fibers (*Gossypium hirsutum*). In: Proceedings of the world cotton research conference-4. International Cotton Advisory Committee, Washington DC, USA

Ko JH, Beers EP, Han KH (2006) Global comparative transcriptome analysis identifies gene network regulating secondary xylem development in *Arabidopsis thaliana*. Mol Genet Genomics 276:517–531

Kubo M, Udagawa M, Nishikubo N, Horiguchi G, Yamaguchi M, Ito J, Mimura T, Fukuda H, Demura T (2005) Transcription switches for protoxylem and metaxylem vessel formation. Genes Dev 19:1855–1860

Laosinchai W, Cui X, Brown RM Jr (2000) A full length cDNA of cotton cellulose synthase has high homology with the *ArabidopsisRSW1* gene and the cotton *CelA1* gene. Plant Physiol 122:291

Lee JJ, Woodward AW, Chen ZJ (2007) Gene expression changes and early events in cotton fibre development. Ann Bot 100:1391–1401

Li CH, Zhu YQ, Meng YL, Wang JW, Xu KX, Zhang TZ, Chen XY (2002) Isolation of genes preferentially expressed in cotton fibers by cDNA filter arrays and RT-PCR. Plant Sci 163:1113–1120

Li Y, Qian Q, Zhou Y, Yan M, Suna L, Zhanga M, Fua Z, Wanga Y, Hanc B, Panga X, Chena M, Li J (2003) *BRITTLE CULM1*, which encodes a COBRA-like protein, affects the mechanical properties of rice plants. Plant Cell 15:2020–2031

Liu D, Tu L, Wang L, Li Y, Zhu L, Zhang X (2008) Characterization and expression of plasma and tonoplast membrane aquaporins in elongating cotton fibers. Plant Cell Rep 27:1385–1394

Meredith WR (2005a) Registration of MD 52ne high fiber quality cotton germplasm and recurrent parent MD 90ne. Crop Sci 45:806–807

Meredith WR (2005b) Minimum number of genes controlling cotton fiber strength in a backcross population. Crop Sci 45:1114–1119

Meyer JR, Meyer VG (1961) Origin and inheritance of nectariless cotton. Crop Sci 1:167–169

Mitsuda N, Sekic M, Shinozaki K, Ohme-Takagi M (2005) The NAC transcription factors NST1 and NST2 of *Arabidopsis* regulate secondary wall thickenings and are required for anther dehiscence. Plant Cell 17:2993–3006

Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, Shinozaki K, Ohme-Takagi M (2007) NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of *Arabidopsis*. Plant Cell 19:270–280

National Cotton Council of America (2008) United States Cotton Production. http://www.cotton.org. Accessed 15 Sept 2008

Orford SJ, Timmis JN (1998) Specific expression of an expansin gene during elongation of cotton fibers. Biochim Biophys Acta 1398:342–346

Paredez AR, Somerville CR, Ehrhardt DW (2006) Visualization of cellulose synthase demonstrates functional association with microtubules. Science 312:1491–1495

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

Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM (1996) Higher plants contain homologs of the bacterial *celA* genes encoding the catalytic subunit of cellulose synthase. Proc Natl Acad Sci USA 93:12637–12642

Perrin RM, Wang Y, Yuen CY, Will J, Masson PH (2007) WVD2 is a novel microtubule-associated protein in *Arabidopsis thaliana*. Plant J 49:961–971

Persson S, Wei H, Milne J, Page GP, Somerville CR (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. Proc Natl Acad Sci USA 102:8633–8638

Pirooznia M, Nagarajan V, Deng Y (2007) GeneVenn—a web application for comparing gene lists using Venn diagrams. Bioinformation 1:420–422

Roudier F, Schindelman G, DeSalle R, Benfey PN (2002) The COBRA family of putative GPI-anchored proteins in *Arabidopsis*. A new fellowship in expansion. Plant Physiol 130:538–548

Ruan YL, Llewellyn DJ, Furbank RT (2001) The control of single-celled cotton fiber elongation by developmentally reversible gating of plasmodesmata and coordinated expression of sucrose and K+ transporters and expansin. Plant Cell 13:47–60

Ruan YL, Xu SM, White R, Furbank RT (2004) Genotypic and developmental evidence for the role of plasmodesmatal regulation in cotton fiber elongation mediated by callose turnover. Plant Physiol 136:4104–4113

Seagull RW (1986) Changes in microtubule organization and wall microfibril orientation during in vitro cotton fiber development: an immunofluorescent study. Can J Bot 64:1373–1381

Shi YH, Zhu SW, Mao XZ, Feng JX, Qin YM, Zhang L, Cheng J, Wei LP, Wang ZY, Zhu YX (2006) Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. Plant Cell 18:651–664

Smart LB, Vojdani F, Maeshima M, Wilkins TA (1998) Genes involved in osmoregulation during turgor-driven cell expansion of developing cotton fibers are differentially regulated. Plant Physiol 116:1539–1549

Taylor NG (2008) Cellulose biosynthesis and deposition in higher plants. New Phytol 178:239–252

Taylor NG, Scheible WR, Cutler S, Somerville CR, Turner SR (1999) The irregular xylem3 locus of *Arabidopsis* encodes a cellulose synthase required for secondary cell wall synthesis. Plant Cell 11:769–780

Thaker VS, Saroop S, Vaishnav PP, Singh YD (1989) Genotypic variations and influence of diurnal temperature on cotton fibre development. Field Crops Res 22:129–141

Triplett BA (1992) Strategies for improving cotton fiber quality. In: Cotton fiber cellulose: structure, function, and utilization. National Cotton Council, Memphis, TN, USA, pp 107–114

Tu LL, Zhang XL, Liang SG, Liu DQ, Zhu LF, Zeng FC, Nie YC, Guo XP, Deng FL, Tan JF, Xu L (2007) Genes expression analyses of sea-island cotton (*Gossypium barbadense* L.) during fiber development. Plant Cell Rep 26:1309–1320

Udall JA, Flagel LE, Cheung F, Woodward AW, Hovav R, Rapp RA, Swanson JM, Lee JJ, Gingle AR, Nettleton D, Town CD, Chen ZJ, Wendel JF (2007) Spotted cotton oligonucleotide microarrays for gene expression analysis. BMC Genomics 8:81

Wightman R, Turner SR (2008) The roles of the cytoskeleton during cellulose deposition at the secondary cell wall. Plant J 54:794–805

Wilkins TA, Jernstedt JA (1999) Molecular genetics of developing cotton fibers. In: Basra AS (ed) Cotton fibers: developmental biology, quality improvement, and textile processing. The Hawthorn Press, NY, pp 231–269

Wittmann T, Wilm M, Karsenti E, Vernos I (2000) TPX2, a novel *Xenopus* MAP involved in spindle pole organization. J Cell Biol 149:1405–1418

Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Afshari C, Paules RS (2001) Assessing gene significance from cDNA microarray expression data via mixed models. J Comput Biol 8:625–637

Wu Z, Soliman KM, Bolton JJ, Saha S, Jenkins JN (2007) Identification of differentially expressed genes associated with cotton fiber development in a chromosomal substitution line (CS-B22sh). Funct Integr Genomics 8:165–174

Yamaguchi M, Kubo M, Fukuda H, Demura T (2008) VASCULAR-RELATED NAC-DOMAIN7 is involved in the differentiation of all types of xylem vessels in *Arabidopsis* roots and shoots. Plant J 55:652–664

Yatsu LY, Jacks TJ (1981) An ultrastructural study of the relationship between microtubules and microfibrils in cotton (*Gossypium hirsutum* L.) cell wall reversals. Am J Bot 68:771–777

Yuan JS, Reed A, Chen F, Stewart CN Jr (2006) Statistical analysis of real-time PCR data. BMC Bioinformatics 7:85

Yuen CY, Pearlman RS, Silo-Suh L, Hilson P, Carroll KL, Masson PH (2003) WVD2 and WDL1 modulate helical organ growth and anisotropic cell expansion in *Arabidopsis*. Plant Physiol 131:493–506

Zhang D, Hrmova M, Wan CH, Wu C, Balzen J, Cai W, Wang J, Densmore LD, Fincher GB, Zhang H, Haigler CH (2004) Members of a new group of chitinase-like genes are expressed preferentially in cotton cells with secondary walls. Plant Mol Biol 54:353–372