ORIGINAL ARTICLE

Jinsuk J. Lee · Osama S. S. Hassan · Wenxilang Gao
Ning E. Wei · Russell J. Kohel · Xiao-Ya Chen
Paxton Payton · Sing-Hoi Sze · David M. Stelly
Z. Jeffrey Chen

Developmental and gene expression analyses of a cotton naked seed mutant

Abstract Cotton fiber development is a fundamental biological phenomenon, yet the molecular basis of fiber cell initiation is poorly understood. We examined molecular and cellular events of fiber cell development in the naked seed mutant (*N1N1*) and its isogenic line of cotton (*Gossypium hirsutum* L. cv. Texas Marker-1, TM-1). The dominant mutation not only delayed the process of fiber cell formation and elongation but also reduced the total number of fiber cells, resulting in sparsely distributed short fibers. Gene expression changes in TM-1 and *N1N1* mutant lines among four tissues were analyzed using spotted cotton oligo-gene microarrays. Using the *Arabidopsis* genes, we selected and designed ∼1,334 70-mer oligos from a subset of cotton fiber ESTs. Statistical analysis of the microarray data indicates that the number of significantly differentially expressed genes was 856 in the leaves compared to the ovules (3 days post-anthesis, DPA), 632 in the petals relative to the ovules (3 DPA), and 91 in the ovules at 0 DPA compared to 3 DPA, all in TM-1. Moreover, 117 and 30 genes were expressed significantly different in the ovules at three and 0 DPA, respectively, between TM-1 and *N1N1*. Quantitative RT-PCR analysis of 23 fiber-associated genes in seven tissues including ovules, fiber-bearing ovules, fibers, and non-fiber tissues in TM-1 and *N1N1* indicates a mode of temporal regulation of the genes involved in transcriptional and translational regulation, signal transduction, and cell differentiation during early stages of fiber development. Suppression of the fiber-associated genes in the mutant may suggest that the *N1N1* mutation disrupts temporal regulation of gene expression, leading to a defective process of fiber cell elongation and development.

Keywords Cotton (*Gossypium hirsutum* L) · Fiberless mutant · Fiber cell initiation · Microarray · Gene regulation · *Arabidopsis*

Abbreviations TM-1: Texas Marker-1 · DPA: Days post anthesis · qRT-PCR: Quantitative reverse-transcriptase polymerase chain reaction

J. J. Lee · O. S. S. Hassan · W. Gao · D. M. Stelly
Z. J. Chen (✉)
Department of Soil and Crop Sciences and Intercollegiate Program in Genetics, Texas AM University, MS 2474/Molecular Genetics, College Station, TX 77843, USA
E-mail: zjchen@tamu.edu
Tel.: +1-979-8453773
Fax: +1-979-8450456

N. E. Wei · S.-H. Sze
Departments of Computer Science
and Biochemistry and Biophysics, Texas AM University, College Station, TX 77843, USA

R. J. Kohel
Crop Germplasm Research Unit, USDA-ARS, College Station, TX 77843, USA

X.-Y. Chen
Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

P. Payton
USDA-ARS Cropping Systems Research Laboratory, Lubbock, TX 79415, USA

Present address: O. S. S. Hassan
Agricultural Genetic Engineering Research Institute (AGERI), Agricultural Research Center (ARC), Giza, Egypt

Present address: W. Gao
Department of Genetics, University of Georgia, Athens, GA 30602, USA

Present address: Z. J. Chen
Institute for Cellular and Molecular Biology, The University of Texas at Austin, 1 University station, A-4800, Austin, TX 78712-0159, USA

Introduction

Cotton represents nearly half of the world's natural fiber consumption. The morphological development of cotton

fibers has been documented (Basra and Malik 1984; Benedict et al. 1973; Graves and Stewart 1988; Mauney 1984; Ramsey and Berlin 1976; Ruan and Chourey 1998; Ruan et al. 2000; Schubert et al. 1976; Stewart 1975). Fiber initiation is a quasi-synchronous process in developing ovules during anthesis. Cotton fibers undergo four overlapping developmental stages: fiber cell initiation, elongation, cellulose biosynthesis, and maturation (Basra and Malik 1984; Tiwari and Wilkins 1995; Wilkins and Jernstedt 1999). The signals important to fiber cell differentiation must occur prior to the formation of fiber cell initials. Fiber cell initials emerge at or prior to anthesis, continue to develop up to 5 days post-anthesis (DPA), and then proceed to cell elongation and primary wall biosynthesis (5–20 DPA). The cellulose (secondary wall) biosynthesis commences at 15 DPA and reaches a peak at 25 DPA (Meinert and Delmer 1977), followed by maturation when the bolls open at ~50 DPA. Cotton fibers are derived from ovular protodermal cells (maternal tissue); only 15–25% of the cells differentiate into the commercially important and spinnable "long" fibers (Basra and Malik 1984; Kim and Triplett 2001; Wilkins and Jernstedt 1999). For the cells committed to the long-fiber development, fiber cell initiation and elongation are synchronized on each ovule, indicating that changes in gene expression during fiber differentiation and development are orchestrated through intercellular signaling and/or timing mechanisms.

The molecular process of fiber initiation is poorly understood because fiber cell initiation occurs immediately after pollination, so it is difficult to know when and how the protodermal cells are committed to fiber development. Genes controlling cell cycle, signal transduction, hormone regulation, cytoskeletal features, and formation and deposition of complex carbohydrates and cell wall proteins are implicitly involved, but their roles in fiber cell initiation are undefined. Cotton fibers are seed trichomes. Unlike leaf trichomes in *Arabidopsis* (Hulskamp 2004; Hülskamp et al. 1994; Hülskamp and Schnittger 1998; Marks 1997), cotton fibers are linear cells that never branch. Trichome formation in *Arabidopsis* leaves is mediated through positive and negative regulators such as *GLABROUS1* (*GL1*), *TRANSPARENT TESTA GLABRA* (*TTG*), *GL3*, *TRIPTYCHON* (*TRY*), and *CAPRICE* (*CPC*) (Hulskamp 2004), whereas *GLABRA2* (*GL2*) is required for cell expansion, branching, and maturation (Szymanski et al. 2000). *MIXTA*, a GL1-related gene in *Antirrhinum majus* (Noda et al. 1994), controls the development of multicellular trichomes (Payne et al. 1999). *GL1* and *GL2* encode MYB and homeodomain transcription factors, respectively. Genetic tests indicated that *GL2* acted downstream of *TTG* and *GL1* since *gl2/gl1* and *gl2/ttg* double mutants lack trichomes, whereas those with *gl2* mutations initiated trichomes normally (Hulskamp 2004; Hülskamp et al. 1994). In cotton, expression of type I genes (*GhMYB1,2*, and *3*) was detected in all tissues tested, while type II genes (*GhMYB4,5*, and *6*)

were expressed differently during development (Loguercio et al. 1999). Moreover, *GhMYB109*, a gene encoding a R2R3 MYB transcription factor, was expressed specifically in fiber initials and elongating fibers (Suo et al. 2003). Wang et al. (2004) demonstrated that overexpression of *GaMYB2* complemented *gl1* phenotype as well as induced seed trichome (though only one seed fiber) development in *Arabidopsis*, suggesting a role of *MYB*-like transcription factors in cotton fiber cell differentiation. The MYB genes appeared to be conserved among diploid species and between diploids and *G. hirsutum* L. allopolyploids (Cedroni et al. 2003). Moreover, the homoeologous MYB genes were not equally expressed, suggesting subfunctionalization of the progenitors' genes during allopolyploidization (Adams et al. 2003).

In addition to the MYB genes, Ruan and Chourey (1998) demonstrated that the expression of the sucrose synthase gene (*SS3*) was dramatically reduced in the fiberless seed (*fls*) mutant, which correlates with the defective fiber cell initials. *SS3* expression is localized to the basal areas of initiating fiber cells in the wild-type plants. Suppression of *SS3* in transgenic cotton leads to adverse effects on fiber cell initiation and elongation, and seed development (Ruan et al. 2001, 2003; Ruan and Chourey 1998;). The data suggest that sucrose synthases are essential for carbon partitioning and turgor pressure maintenance during rapid cell growth and expansion. However, fiber cell differentiation is a complex process involving many other pathways such as signal transduction and transcriptional regulation (Kim and Triplett 2001). In recent studies using subtractive PCR and membrane-based cDNA array analyses (Ji et al. 2003; Li et al. 2002), several research groups have identified over 100 genes that were highly expressed in elongating cotton fibers (10 DPA). The candidate genes are homologous to those involved in auxin, MAPK signaling pathways, lipid biosynthesis and transportation, and cell-wall biosyntheses (Ji et al. 2003; Li et al. 2002; Zhu et al. 2003).

Several "qualitative" mutants in fiber development have been reported (Endrizzi et al. 1984; Kohel 1973). The best characterized of these are the naked seed loci, *N1N1* and *n2n2*. These mutants, discovered decades ago, lack most of the lint (long and spinable) fibers. The fuzz (short) fibers develop but eventually fall off the seeds to produce black or "naked" seeds. Phenotypically, *N1N1* is slightly more extreme and consistent than *n2n2*, which is more susceptible to changes in genotypes and environmental conditions. Their similar phenotypic effects and linkage relationship in two homoeologous chromosomes (each is thought to be about 15 cM from the respective centromeres of chromosomes 12 and 26) suggest that they are homoeologous loci (Samora et al. 1994). Interestingly, *N1N1* is dominant, whereas *n2n2* is recessive. *Gossypium hirsutum* L. cv. TM-1 is an elite inbred line ($>S_{40}$) derived from the obsolete cultivar "Deltapine 14". It has been widely used in research programs, isoline development, QTL analysis, and

genetic and physical mapping. Mutant isogenic lines have been produced by backcrossing > 6 generations to TM-1.

To gain a better understanding of gene regulation in the early stages of fiber development, we developed a cotton oligo-microarray using ~1,334 cotton genes putatively encoding chromatin proteins, transcription factors, and proteins involved in cell wall biosynthesis and signal transduction pathways. The genes were selected by BLAST analysis of *Arabidopsis* gene families. We compared gene expression profiles in various tissues including young ovules (0 and 3 DPA) and non-fiber tissues in the naked seed mutant (*N1N1*) and its isogenic wild type (*Gossypium hirsutum* L. cv. TX Maker-1 or TM-1). A subset of 23 fiber-associated genes was studied in seven fiber- and non-fiber tissues using quantitative RT-PCR analysis. The data suggest that the expression of many downstream genes encoding transcriptional and translational factors, signal transduction proteins, and cell differentiation factors is affected in the *N1N1* mutants, which may lead to a defective process in the early stages of fiber cell elongation and development.

Materials and methods

Plant materials

G. hirsutum L. cv. TM-1 and its isogenic *N1N1* mutant lines (> S6 generation) were grown in a greenhouse. Young emerging leaves (about 2 inches in diameter) were collected from seedling plants, and petals were collected at the day of anthesis (0 DPA) before the flower color changed from white to pink. Flower buds prior to anthesis (−3 DPA) were collected when the ovules were enclosed by squares of 1/3–1/2 inches in diameter. Flowers were tagged on the day of anthesis, and ovules were harvested at 0, 3, and 5 DPA. Cotton bolls (TM-1) at 10 DPA were harvested for dissecting fibers, whereas for *N1N1* the developing ovules were harvested. For each genotype, we used two biological pools, each with ten plants grown at similar stages. Leaves were harvested by pooling one leaf from each of ten individual plants. Ovules or fibers were dissected from five bolls collected in each of ten plants. The fresh tissues were frozen in liquid nitrogen and stored in a −70°C freezer or subjected to RNA extraction.

Analysis of fiber cell development by scanning electron microscopy (SEM)

SEM was performed using a modified protocol (Murai et al. 2002; Tian et al. 2003). In brief, ovules from 0 to 3 DPA were dissected from immature ovules of TM-1 and the isogenic naked seed mutant, *N1N1*. The ovules were fixed in a solution containing 3% each of formaldehyde and glutaraldehyde in 0.1 M sodium cacodylate buffer (pH = 7.4) and rinsed in 0.2 M sodium cacodylate buffer (pH = 7.4) three times. The ovules were washed in an ethanol series from 10 to 70% every 15 min. The concentration of the ethanol was increased to 100% within 18 h to dehydrate the samples for SEM analysis. The specimens were prepared by critical-point, dried with CO_2 at 1,400 and 1,800 psi, consecutively, and mounted by conductive gold paint and sputter coating. The samples were then scanned and analyzed using a JEOL (JSM-6400) SEM located in the Microscopy and Imaging Center at Texas A&M University, with an accelerating voltage of 15 kV and a working distance of 39 mm. Images were scanned and stored as TIFF files.

EST sequence analysis and oligo selection

To construct a pilot set of cotton 70-mer oligos, we selected a subset of genes based on *Arabidopsis* protein sequences (Arabidopsis Genome Initiative 2000) using TIGR Gene Indices (http://www.tigr.org/tdb/tgi/plant.shtml). Each family of the *Arabidopsis* proteins was analyzed using BLAST against the entire cotton

Table 1 *Arabidopsis* genes utilized for designing 70-mer oligos in cotton

Name of gene family	No. of *Arabidopsis* homologs	No. of cotton ESTs matched
Chromatin factors	272	276
Myb transcription factors	131	56
WRKY transcription factors	(74) 71	99
Cell wall proteins	(56) 30	26
Cell cycle proteins	61	164
Cloned fiber-cDNAs	NA	130
Map kinases (MapK)	20	137
MapKK	10	143
MapKKK	80	228
Brassinosteriods	5	5
Drought and cold inducible genes	83	808
Subtotal	2,224-875 (overlap) = 1,334	
Controls	15	152
Chloroplast genes	(130) 121	NA
Mitochondrial genes	(67) 66	NA
Total	1,334 + 15 + 121 + 66 = 1,536	

Numbers in parenthesis indicates the total number of gene families in *Arabidopsis*; NA: not applicable

Table 2 Microarray experimental design

Note each experiment consists of eight slide hybridizations or four dye-swaps. Each dye-swap is repeated once for one RNA sample (technical replications) and repeated again using another independent RNA sample (biological replications). See Fig. 3 for details about the dye-swap and experimental design (Chen et al. 2004)

Experiment	Slide No.	Dye-swap	RNA or cDNA	Cy3	Cy5
1	1–4	1–2	RNA1	TM-1 (3 DPA)	TM-1 (Leaf)
	5–8	3–4	RNA2	TM-1 (Leaf)	TM-1 (3 DPA)
2	9–12	1–2	RNA1	TM-1 (3 DPA)	TM-1 (Petal)
	13–16	3–4	RNA2	TM-1 (Petal)	TM-1 (3 DPA)
3	17–20	1–2	RNA1	TM-1 (3 DPA)	TM-1 (0 DPA)
	21–24	3–4	RNA2	TM-1 (0 DPA)	TM-1 (3 DPA)
4	25–28	1–2	RNA1	TM-1 (3 DPA)	*N1N1* (3 DPA)
	29–32	3–4	RNA2	*N1N1* (3 DPA)	TM-1 (3 DPA)
5	33–36	1–2	RNA1	TM-1 (0 DPA)	*N1N1* (0 DPA)
	36–40	3–4	RNA2	*N1N1* (0 DPA)	TM-1 (0 DPA)

EST database with a method similar to the previously published (Blanc and Wolfe 2004a, 2004b). Two proteins were considered to be similar if their amino-acid sequences shared over 38% identity or 60% similarity when the aligned sequences were ≤ 100 amino acids (a.a.) in length, or over 28% identity or 50% similarity when the aligned sequence were > 100 a.a. in length (*E*-value ≤ 0.00001). The selected ESTs were compared using BLAST against each other to eliminate duplicates that shared more than 95% sequence identity (Table 1). The remaining ESTs were used for 70-mer oligo design so that each oligo had a minimum of secondary hairpin structures (Kane et al. 2000; Rouillard et al. 2002). Sequence lengths of 70 nucleotides with similar melting temperature of $73 \pm 2°C$ were selected within 1,000 nucleotides of the 3′ end of predicted coding sequences using the software developed by ProbeSelect (Li and Stormo 2001) or Featurama (http://probepicker.source-forge.net/). We selected 1,334 oligos from cotton ESTs (http://www.tigr.org/tdb/tgi/plant.shtml), 15 controls, and 121 and 66 oligos designed from annotated *Arabidopsis* chloroplast and mitochondrial genes, respectively, using the Organelle Genome Resources at the NCBI http://www.ncbi.nlm.nih.gov/genomes/ORGAN-ELLES/organelles.html. A total of 1,536 features were spotted in duplicate on each slide.

Microarray experimental design and statistical analysis

The 70-mer oligos were synthesized at Operon (http://www.operon.com/arrays/omad.php). The oligo solution (30 μM in 3XSSC) was spotted onto SuperAmine microarray slides (SMM) using micro spotting Stealth 3 pins (SMP3) (TeleChem International, Inc. Sunnyvale, CA) in an OmniGrid Accent microarrayer GeneMachines, San Carlos, CA). Microarray design and statistical tools were developed as previously described (Lee et al. 2004; Wang et al. 2005). Messenger RNA (mRNA) was directly isolated from various tissues of TM-1 and *N1N1* using Poly(A) Pure (Ambion, Austin, TX) according to the manufacturer's recommendations (Wang et al. 2005). Cy3- and Cy5-labeled probes were generated using the CyScribe Post-Labelling kit (Amersham Biosciences Corp, Piscataway, NJ). Messenger RNA was reverse-transcribed into amino allyl-labelled cDNA. The single-strand cDNAs were coupled with CyDye-NHS ester, which binds to the modified nucleotides. We used 500 ng of mRNA in each labeling reaction using Cy3- or Cy5-dCTP (Amersham Biosciences). We used dye-swap experimental design and linear model (Chen et al. 2004; Kerr and Churchill 2001) for our microarray analysis. For one dye-swap experiment, we used two sets or four labeling reactions. After labeling, one Cy3-dCTP reaction is mixed with one Cy5-dCTP reaction to make one probe. Therefore, two "identical" probes each containing an equal amount of Cy3- and Cy5-labled cDNAs were hybridized with two slides, which constituted one dye-swap experiment. The dye-swap was repeated once as a technical replication. The large dye-swap (four slides) was repeated using another biological sample (e.g., RNAs isolated from different pools of 3-DPA ovules). Therefore, each experiment consists of four technical replications and two biological replications in a total of eight slides (Chen et al. 2004). Hybridization was performed overnight (~14 h) at 65°C. After hybridization, the slides were washed twice for 4 min each in 2X SSC, 0.2% SDS, again twice for 2 min each in 0.2XSSC, and twice for 2 min each in 0.05X SSC. After drying the slides by brief centrifugation (5 min at 850 rpm), the slides were scanned using GenePix 4000B (Axon, Foster City, CA), and the images were captured by GenePix Pro 4.1 software.

After the data were processed using natural logarithm ratios of green and red hybridization signals, a robust and locally weighted linear regression (lowess) (Cleveland 1979) was used to remove non-linear components (e.g. dye and pin effects) (Quackenbush 2002). For the duplicate spots in each feature, we used an average value for data analysis. No additional steps for data normalization and background subtraction are needed for the AVONA model (Lee et al. 2004). The data were then subjected to the analysis of variance (ANOVA) test in a linear model to estimate the significant changes in gene expression caused by the two treatments (genotypes) (Black and Doerge 2002; Lee et al. 2004). A standard *t*-test statistic was used for this comparison based on the normality assumption for the residuals. The standard false discovery rate (FDR) (Hochberg and Tamhane 1987) was applied to control multiple testing errors using a significance level $\alpha = 0.05$.

Quantitative RT-PCR (qRT-PCR) analysis

Messenger RNA (mRNA) was extracted from seven different tissues including leaves, petal, ovules at −3 DPA, 0 DPA, 3 DPA, 5 DPA, and fibers at 10 DPA in TM-1 and *N1N1* mutant using Poly(A) Pure (Ambion, Austin, TX). For microarray data verification, the same mRNA used in the microarray analysis was used for qRT-PCR analysis. The cDNAs were amplified by Superscript II reverse transcriptase reaction (Invitrogen, Carlsbad, CA). For the transcript amplification, gene-specific primers (Supplementary Table 3) were designed using Primer Express version 2.0 software. The qRT-PCR reaction was carried out in a final volume of 20 μl containing 7 μl SYBR Green PCR master mix (Applied Biosystems, Foster City, CA), 1 μM forward and reverse primers, and 0.1 μM cDNA probe in a ABI7500 Real-Time PCR system (Applied Biosystems, Foster City, CA). Cotton *HISTONE3* (AF024716) was used to normalize the amount of gene-specific RT-PCR products (Wang et al. 2004). All reactions were performed in three replications using dissociation curve to control the absence of primer dimers in the reactions. The amplification data were analyzed using ABI7500 SDS software (version 1.2.2), and the fold changes were calculated using the standard in each reaction.

Results

Dominant naked seed *N1N1* mutation affects fiber cell formation and elongation in allotetraploid cotton

To examine fiber cell differentiation in the mutant at the cellular level, we used scanning electron microscopy (SEM) to observe the development of fiber cell initials in the ovular surface during early stages (0–3 DPA) of fiber development (Fig. 1). On the day of anthesis (0 DPA), ∼25–30% of the cells in TM-1 began producing fiber cell initials (buds), whereas the protodermal cells in the *N1N1* mutant remained unchanged. Fiber cell formation proceeded rapidly in TM-1 and at 1 DPA the ovular surface near the chalazhal end was covered with evenly distributed fiber cell initials (Fig. 1a, c). In *N1N1*, the fiber cell initials did not emerge until 1 DPA (or > 12–16 h post-anthesis, Fig. 1b, d) and developed at a relatively low density. We estimated the number of fiber cell initials was approximately ∼ 20–30% of that in TM-1, or ∼5–8% of the total protodermal cells. The elongation process of fiber cells was also slow and abnormal in the *N1N1* mutant (Fig. 1f, h) compared to TM-1 (Figs. 1e, g). As a result, only short fibers were formed in the *N1N1* mutant, whereas long fibers developed in TM-1. At 2 DPA, the fiber cells of TM-1 synchronously elongated with a relatively uniform length (Fig. 2c), whereas the fiber cells in the mutant developed asynchronously and were distributed unevenly over the ovular surface (Fig. 2d). From the longitude view, fiber cells initiated from the chalazas to micropylar end and covered ∼1/2–3/4 of the seed surface by the end of 2 DPA (Fig. 2c–f). After 3 DPA, TM-1 ovules were covered densely with normal elongating fibers (Fig. 1). Compared to the fiber cells in TM-1 (Fig. 2e), the development of the sparsely dispersed fiber cells in the mutant was very slow, giving rise to short defective fibers, as observed at 8 DPA (Fig. 2f). These short fibers eventually fell off the seed surface, and the seeds became "naked". The data indicate the *N1N1* mutation affects not only the number of protodermal cells for fiber differentiation but also fiber cell elongation.

Use of *Arabidopsis* genes to design cotton oligo-gene microarray

To understand the molecular basis of fiber cell differentiation and elongation, we developed spotted cotton oligo-gene microarrays using 16,695 EST Tentative Consensus (TC) assemblies (from the data posted in May 2002, http://www.tigr.org/tdb/tgi/plant.shtml), including ∼10,000 from a diploid cultivated cotton, *G. arboreum* L. and ∼7,000 from *G. hirsutum* L. In the

Table 3 The number of differentially expressed genes in fibers and non-fiber tissues and in TM-1 and *N1N1* that were detected using a common variance and/or a per-gene variance

| Experiment | Number of genes | | | | |
| | Per-gene Variance | Common Variance | Shared set [a] | | |
			Sum	Down-regulated genes [b]	Up-regulated genes [b]
1 TM-1 (3 DPA) vs. TM-1 (Leaf)	1037	909	856	412	444
2 TM-1 (3 DPA) vs. TM-1 (Petal)	880	673	632	326	306
3 TM-1 (3 DPA) vs. TM-1 (0 DPA)	449	162	91	8	83
4 TM-1 (3 DPA) vs. *N1N1* (3 DPA)	498	152	117	66	51
5 TM-1 (0 DPA) vs. *N1N1* (0 DPA)	195	65	30	20	10

[a] Shared: shared data sets of the significantly differentially expressed genes detected using both common variance and per-gene variance
[b] Up- and down-regulation referred to the genes in the former tissue in each comparison. For example, in the comparison of TM-1 (3 DPA) versus TM-1 (Leaf) the genes were down- or up-regulated in TM-1 (3 DPA)
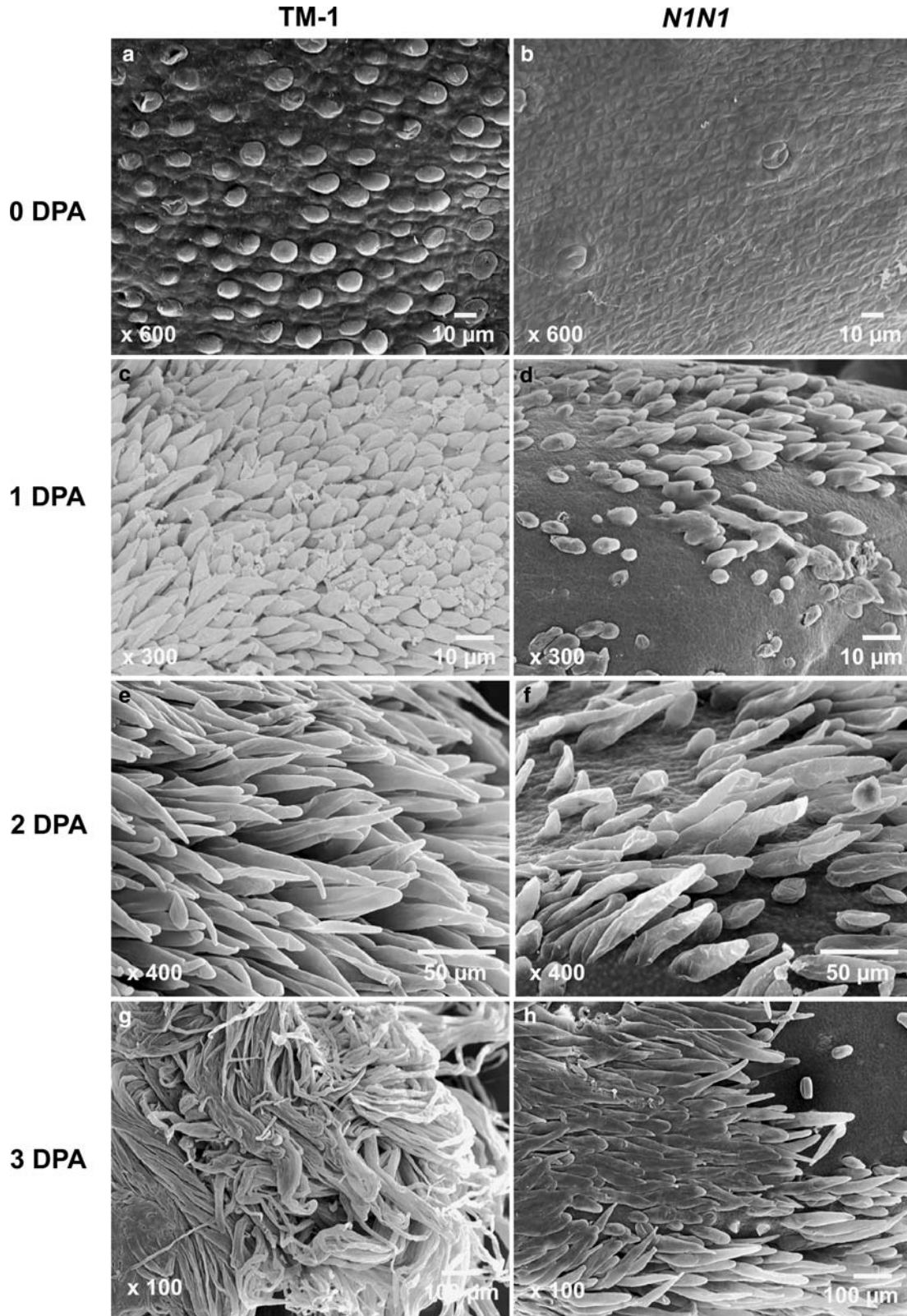
Fig. 1 Delay of fiber cell initiation in the naked seed mutant (*N1N1*) observed using scanning electron microscopy (SEM). **a–b** Ovule surface of fiber cell initials at 0 DPA in TM-1 (**a**) and the isogenic mutant line, *N1N1* (**b**); **c–d** Fiber cells at 1 DPA in TM-1 (**c**) and *N1N1* mutant (**d**); **e–f** Fiber cells in TM-1 (**e**) and *N1N1* (**f**) lines; and **g–h** Young fibers at 3 DPA in TM-1 (**g**) and *N1N1* mutant (**h**)

Fig. 2 Ovule morphology of
TM-1 (**a**) and its isogenic
*N1N1*mutant (**b**) at 0 DPA;
(**c-d**). Distribution of fiber cells
in TM-1 (**c**) and *N1N1* (d) at
2DPA; **e–f** Distribution of
fibers at 8 DPA in TM-1 (**e**) and
N1N1 (**f**). Mature seeds are
shown in the lower part of the
photos. Note that the TM-1
seed was ginned (**e**); only a
small amount of fuzz was
produced and the seed became
naked in *N1N1* (**f**)
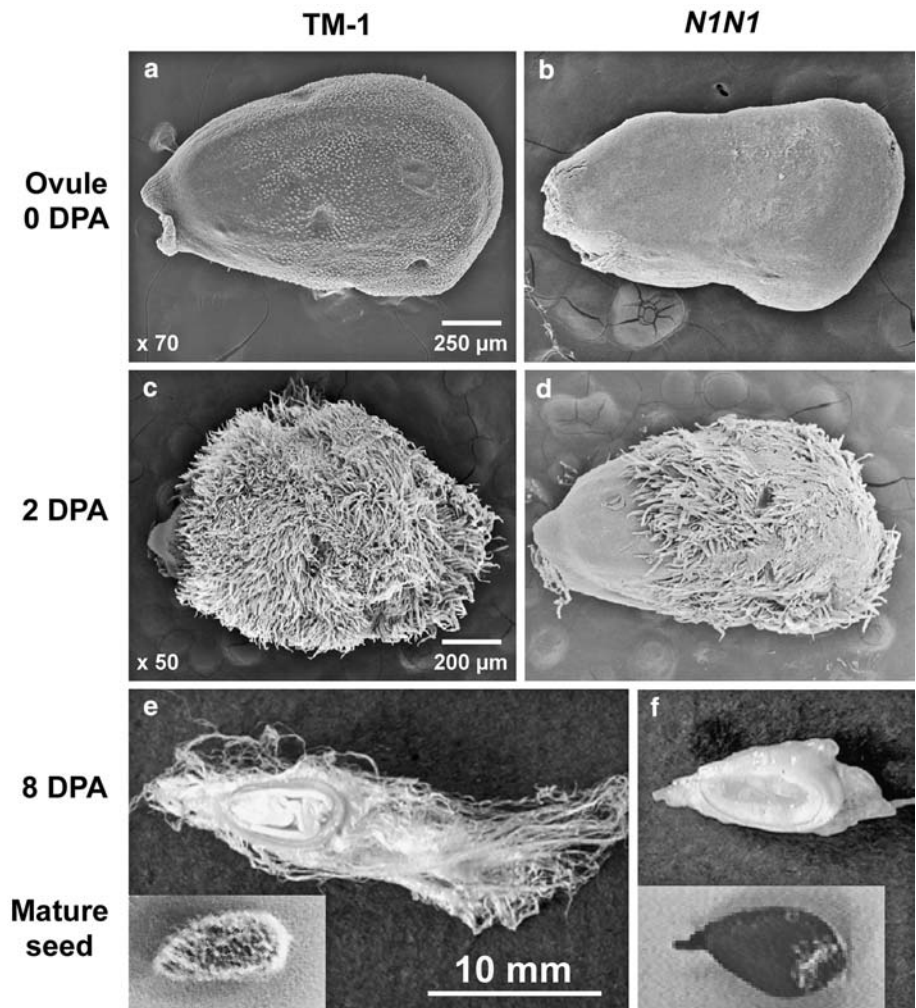
comparison of nucleotide sequences, we found the cotton ESTs had the highest percentage of sequence identity with those of *Arabidopsis* and soybean. The sequence identity between cotton, rice, and maize ESTs was not significantly lower than that between cotton and other dicotyledonous plants (data not shown). This is because the majority of cotton ESTs was derived from the 5′ end of cDNAs. The 70-mer oligos selected showed a high percentage (60–85%) of sequence identity between cotton and other species. We selected 1,334 cotton ESTs encoding putative chromatin proteins, MYB, WRKY, MAP-kinase, cell wall and cell cycle proteins (Table 1) because these genes were predicted to play important roles in fiber cell differentiation and development (Arabidopsis Genome Initiative 2000; Hulskamp 2004). For the majority of protein groups, cotton has more homologous genes than *Arabidopsis*, suggesting that allotetraploid cotton contains many redundant genes resulting from polyploidization. An exception is that, excluding the overlaps, cotton fiber ESTs contain fewer numbers of the *Arabidopsis* MYB-related genes and chromatin genes, which is reminiscent

of the estimate that cotton fiber genes represent 35% transcriptome in the cotton genome (Arpat et al. 2004). It is likely that some MYB and chromatin genes are under-represented in the current EST collections or expressed in the tissues such as immature ovules prior to fiber formation.

Microarray analysis of gene expression in fiber-bearing ovules and non-fiber tissues

We produced spotted cotton oligo-gene microarrays with a total of 1,536 elements (Supplementary Table 1) using 1,334 cotton genes, 15 controls, and 121 chloroplast genes and 66 mitochondrial genes from *Arabidopsis* (Table 1). Four dye-swap hybridizations were performed as previously described (Tian et al. 2005), resulting in a total of eight replications for each experiment (Table 2). Five experimental comparisons (Table 2) were performed to analyze gene expression changes in the developing ovules containing young fiber cells (0 DPA, 3 DPA) in TM-1 relative to its isogenic

N1N1 mutant and in the ovules relative to the leaves or petals in TM-1 (Table 2). The microarray data were subjected to lowess normalization (Cleveland 1979) and analyzed using a linear model (Black and Doerge 2002; Kerr and Churchill 2001; Tian et al. 2005). The genes that displayed statistically significant expression differences in each experiment were selected using multiple comparison tests and a false discovery rate (FDR) at the 95% confidence level (Hochberg and Tamhane 1987). The controls were not analyzed because no additional data normalization was needed in the linear model (Black and Doerge 2002; Kerr and Churchill 2001).

We compared gene expression changes in three sets of five experiments (Table 3, Fig. 3b) using dye-swaps (Fig. 3a). First, we compared gene expression divergence between fiber-bearing ovules and non-fiber tissues in TM-1. Although the ESTs were primarily derived from cotton fibers collected at 6–10 DPA, the majority of these genes were expressed in both fiber-bearing ovules and non-fiber tissues. A total of 856 (56%) genes were expressed significantly different between the fiber-bearing ovules (3 DPA) and seedling leaves, whereas 632 (41%) genes were differentially expressed between the ovules and petals. Among the 856 genes that were differentially expressed in the leaves and ovules, 444 were up-regulated in the fiber-bearing ovules and 412 genes were up-regulated in the leaves. Excluding the chloroplast genes that are often highly expressed in the leaves, ~350 genes were up-regulated in the leaves. Moreover, among the 632 genes that were differentially expressed in the ovules and petals, nearly half of them were up-regulated in the petals. The data indicate that 40–50% of the ESTs derived from cotton fibers are expressed in leaves and petals.

Second, we compared gene expression divergence between early stages of fiber development. Among the 91 genes that were expressed significantly different in the ovules at 0 and 3 DPA, eight were down-regulated and 83 were up-regulated in the fiber-bearing ovules (3 DPA), suggesting that gene activation is a mode of temporal regulation of the biological pathways during early stages of fiber development.

Third, we compared gene expression divergence between TM-1 and *N1N1* in the early stages of fiber development. In the ovules at 3 DPA, 117 genes were expressed significantly differently between TM-1 and *N1N1*. Among them, 51 were up-regulated in TM-1 and 66 were up-regulated in *N1N1*. In the ovules at 0 DPA, 20 genes were up-regulated in *N1N1*, whereas ten genes were down-regulated in *N1N1*. Fewer genes detected in the ovules at 0 DPA than at 3 DPA may have indicated a bias of current cotton EST collections that were primarily derived from the late stages (6–10 DPA) of fiber development (Arpat et al. 2004).

We further analyzed differentially regulated genes among three comparative experiments using Venn-diagrams (Fig. 4). If the genes are important to early events of fiber development, they would be up-regulated in the

ovules at 3 DPA compared to those in the leaves in TM-1 and in the ovules in *N1N1*. In the comparison of up-regulated genes between the ovules and leaves or petals (Fig. 4a), 590 genes were significantly up-regulated in both experiments. Among them, 284 and 146 genes were up-regulated in the ovules compared to the leaves and petals, respectively, while 160 genes overlapped in the two comparisons, i.e., they were up-regulated in the ovules compared to both leaves and petals. Remarkably, among 51 genes that were up-regulated in the ovules at 3 DPA in TM-1 compared to *N1N1*, 48 (94%) matched the genes that displayed significantly higher levels of expression in the ovules relative to the leaves or petals. Among them, 23 (45%) were up-regulated in the ovules relative to both leaves and petals, whereas 21 (41%) were up-regulated in the ovules compared to the leaves, and four (8%) up-regulated in the ovules relative to the petals.

Interestingly, three genes (6%), namely, two *HSP*-like genes (AI723022 and BG442970) and a stress-inducible gene (BF270275), were expressed differently only between TM-1 and *N1N1*, suggesting that *N1* mutation may be associated with stress responses that are yet to be determined.

Among 642 genes that were down-regulated in the ovules (3 DPA) compared to leaves and/or petals (Fig. 4b), 316 (49%) were up-regulated in the leaves and 230 (36%) up-regulated in the petals, while 96 genes (15%) overlapped between the two comparisons. Among 66 genes that were down-regulated in the ovules (3 DPA) of TM-1 compared to *N1N1*, only three (5%) matched to the genes that were down-regulated in the TM-1 ovules compared to both leaves and petals, while 45 and three genes were down-regulated in the TM-1 ovules compared to the leaves and petals, respectively. The large number of the down-regulated genes detected in the ovules suggests that many ESTs derived from cotton fibers (3 DPA) were not highly expressed during the development of ovules and young fiber cells. The small number of the overlapped genes that were down-regulated among three comparisons indicates that up-regulation of gene expression in the *N1N1* ovules (3 DPA) may not be associated with the regular process of fiber cell development.

Twenty-three genes (Table 4) that were up-regulated in the ovules at 3 DPA in TM-1 from these comparisons (Fig. 4b) are the likely candidate genes important to early stages of fiber development. Indeed, many of them such as *E6* and *RDL1* are known to play roles in fiber development (John 1996; John and Crow 1992; Li et al. 2002; Wang et al. 2004). The up-regulation was also accounted for twenty other genes encoding nucleosome assembly protein, adenosylhomocysteinase, elongation factor (EF-1 alpha), calnexin homolog precursor, BURP domain-containing protein, ABC transporter, and sugar transporter, which are probably involved in the primary and secondary cell wall biosynthesis associated with rapid cell expansion and cellular elongation (see below).
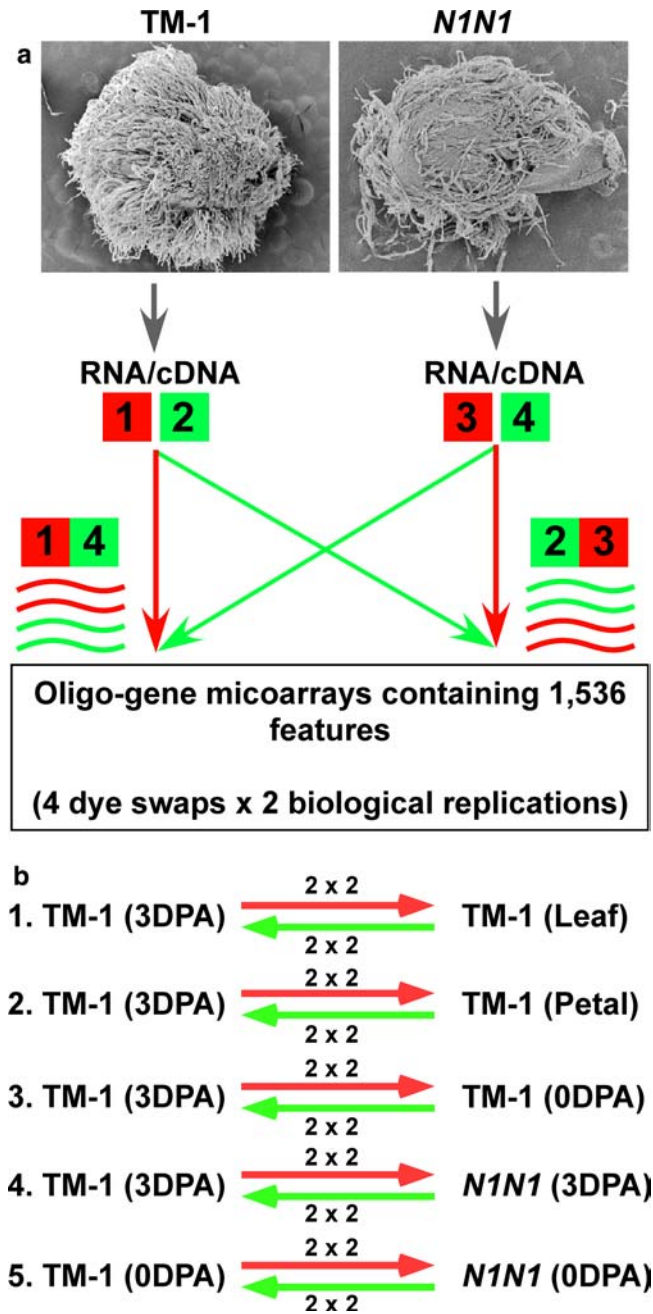
Fig. 3 Experimental design of cotton oligo-gene microarray. a The RNAs are extracted from young fibers (3 DPA) of TM-1 and N1 and reverse transcribed into cDNA. These cDNAs are labeled with florescent dye (either Cy3 or Cy5). Hybridization is performed using the probes prepared from Cy3- and Cy5- labeled cDNAs in TM-1 and N1N1 (3 DPA). Each dye-swap consisted of 2 hybridizations, which was repeated four times making 8 replications of each experiment. b Repeated dye-swap experimental design for five sets of microarray comparisons. In each comparison, the dye-swap (2 hybridizations) as shown in A was repeated once (technical replication), and the repeated dye-swap (4 hybridizations) was performed using another RNA sample (biological replication), resulting in a total of 8-slide hybridizations with 4 technical replications (2 dye-swaps) and 2 biological replications (Table 2)

Up-regulation of the candidate genes is associated with fiber development

To study developmental regulation of fiber-associated genes detected in microarray analysis, we analyzed expression patterns of 23 genes (Table 4) in seven different tissues including leaves, petals, ovules at −3, 0, 3, and 5 DPA, and fibers at 10 DPA using quantitative RT-PCR (qRT-PCR) analysis. Data for six genes representing various expression patterns are shown in Fig. 5a, and the rest of data in Table 4 and Supplementary Fig. 1. Using gene-specific primers (Supplementary Table 3), transcripts of each gene were amplified in seven tissues in TM-1 and N1N1, and the data were analyzed using relative expression ratios with standard deviations (Fig. 5a, Supplementary Table 4). The expression patterns of 21 genes detected by microarray analysis were confirmed by qRT-PCR analysis, although for some genes the relative expression ratios detected by qRT-PCR and the fold changes detected by microarrays, were not the same. The expression ratios detected in the qRT-PCR assays appeared to be much higher than those detected in the microarray analysis, suggesting that qRT-PCR is a relatively sensitive method for gene expression analysis. Alternatively, 70-mer oligos may cross-hybridize the homologous genes in the allopolyploid genomes. For two genes, SRF6 and CEP52, the expression ratios detected by qRT-PCR matched fold-changes in other comparisons but did not match the increased fold changes in the ovules at 3 DPA detected by microarray analysis (Table 4, Supplementary Table 4). One possibility is that the primers designed for the two genes may not correspond to the specific members of multiple gene families encoding the receptor protein (SRF6) and ribosomal protein (CEP52).

Our data reveal at least five types of gene expression patterns that are associated with temporal regulation of early fiber development in TM-1 and N1N1 (Fig. 5a, Table 4, Supplementary Fig. 1). First, GhPDF1 expression was highly induced in the immature ovules at −3 DPA, prior to the formation of fiber cell initials, and remained highly expressed until 5 DPA and in the fibers. Protodermal factor 1 (PDF1) is a protein involved in cell fate determination. In Arabidopsis, PDF1 is exclusively expressed in the L1 layer of vegetative and floral meristems, in organ primordia, and in protodermal cells during embryogenesis. PDF1 expression is undetectable in the epidermis of mature organs (Abe et al. 1999, 2001). Notably, GhPDF1 was highly expressed in immature ovules (−3 DPA) in the N1N1 mutant as in TM-1, whereas its expression was undetectable in the fibers at 10 DPA in N1N1. Second, genes are activated in the ovules at 0 DPA. GhMYB25 representing a small group of the large MYB transcription gene family (Arabidopsis Genome Initiative 2000). GhMYB25 expression was up-regulated in the fiber-bearing ovules but not in the non-fiber tissues, and its expression was induced in the ovules at 0 DPA, and the transcripts were accumulated at high
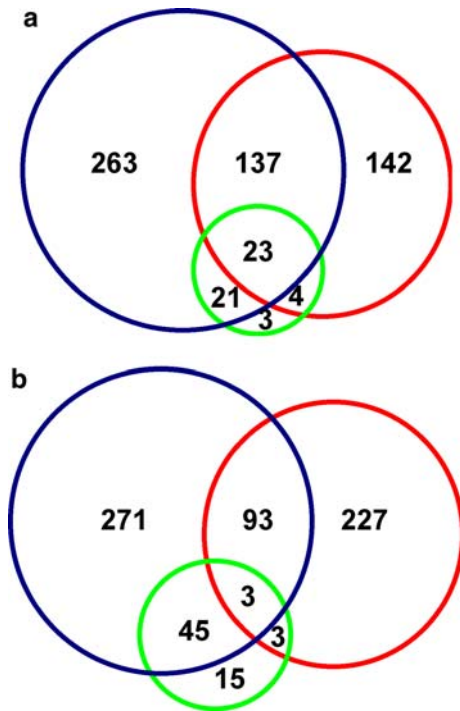
Fig. 4 Comparative analysis of differentially expressed genes detected by microarray analysis. **a** Venn diagram displays the number of up-regulated genes in the ovules (3 DPA) compared to the leaves (blue) in TM-1, in the ovules (3 DPA) relative to the petals (red) in TM-1, and in the TM-1 ovules compared to *N1N1* ovules at 3 DPA (green); **b** Venn diagram displays the number of down-regulated genes compared between the ovules (3 DPA) and leaves (blue) in TM-1, between the ovules (3 DPA) and petals (red) in TM-1, and between the TM-1 and *N1N1* ovules at 3DPA (green). The proportion shown in the diagrams may not be scaled

levels until 5 DPA. The mRNA levels declined when the fiber cells proceeded to the elongation stage (10 DPA). The accumulation of *GhMYB25* transcripts was highest in the young ovules (0 and 3 DPA), low in the fibers (10 DPA), and undetectable in the leaves, petals, and immature ovules (−3 DPA) (Fig 5). *GhMYB25* expression levels were decreased during early stages of fiber elongation (10 DPA). On the contrary, in the *N1N1* mutant lines, *GhMYB25* transcripts were accumulated in the ovules from 0 to 5 DPA and remained at a relatively high level in fibers (10 DPA). A similar activation pattern was observed for *SAC25* that were up-regulated in the ovules at 0 DPA and remained highly expressed from 3 to 10 DPA. Third, 12 of 23 genes were up-regulated starting from 3 DPA and their expression continued to rise and peaked at 10 DPA. These genes encode a variety of important proteins such as fiber E6 protein, dehydration-induced protein, ABC transporter, S-adenosylhom ocysteine hydrolase, and BURP domain-containing protein. *E6* encoding fiber protein E6 is one of the highly expressed genes in fiber tissues throughout the fiber development (John 1996; John and Crow 1992). E6 mRNA and protein are highly accumulated during stages of late primary cell wall and early secondary cell wall

biosynthesis. *E6* expression was undetectable in leaves, petals, and immature ovules (0 DPA) but up-regulated 200-fold in the ovules at 3 DPA, continued to rise 900-fold in the ovules at 5 DPA, and peaked in the fibers at 10 DPA (Fig. 5a, Supplementary Table 4), confirming the gene encoding fiber protein E6 was expressed throughout the development of fiber cells (John and Crow 1992). The fold induction of *E6* expression in the fiber-bearing ovules (3–5 DPA) and fibers (10 DPA) in *N1N1* was 1/5–1/3 of those in TM-1. Similarly, *RDL1* was up-regulated in the fiber-bearing ovules at 3 and 5 DPA and in the fibers at 10 DPA, whereas the expression levels were substantially low in *N1N1* and non-fiber tissues (Fig. 5a). *RDL1* promoter containing an L1 box and a MYB-binding motif was a fiber-specific gene (Li et al. 2002; Wang et al. 2004). It is notable that the promoter of *RDL1* contains MYB-binding elements, indicating physical interactions between MYB transcription factors and *RDL1* (Wang et al. 2004). Collectively, the data support a model (Fig. 5b) of sequential activation of many genes involved in various biological pathways, leading to the progression of fiber cell development (see below). Fourth, six genes were expressed in non-fiber tissues, but their expression levels were dramatically increased in the ovules and fibers. These genes encode proteins such as nucleosome assembly protein, elongation factor (EF-1), HSP70, receptor family, and ribosomal protein CEP52. Fifth, up-regulation was found only in one tissue. *ACT* and *RDL2* were highly expressed in TM-1 ovules at 5 DPA and *N1N1* fibers at 10 DPA, respectively. Significantly, the expression levels of 23 genes, except *RDL*, in seven tissues tested were much lower in *N1N1* than in TM-1, suggesting that the *N1N1* mutation disrupts the temporal regulation of many genes involved in various biological pathways including signals for fiber cell elongation and the number of cells committed to fiber differentiation.

Discussion

Cotton fiber ESTs and their expression patterns in the fibers and non-fiber tissues

The majority of cotton ESTs are derived from the fibers at 6–10 DPA in *G. arboreum* L. or *G. hirsutum* L. species. Estimates indicate that fiber transcriptome represents 35–40% of the genes in the cotton genome (Arpat et al. 2004). It is unclear about the proportion of the cotton fiber ESTs that are expressed during the development of fibers and non-fiber tissues. Microarray analysis of gene expression using fibers and non-fiber tissues suggests that approximately 40–50% current fiber ESTs are expressed in leaves and flower petals (Fig. 4, Table 3). However, the number of genes expressed in non-fiber tissues could be lower (~20%) than 40% because many non-fiber tissues (e.g., roots and stems) are not used in the study. Among 51 genes that are differentially expressed between TM-1 and *N1N1* ovules (3 DPA), 28 are expressed either in leaves

or in petals (Table 3, Fig. 4a). Moreover, five of nine randomly selected genes that display expression differences between TM-1 and *N1N1* are also expressed in non-fiber tissues (leaves and petals) in qRT-PCR analysis (data not shown). The high percentage of ESTs that are expressed in non-fiber tissues indicates that many genes involved in general biological pathways, such as metabolism, energy production, and the biosynthesis of primary and secondary cell walls, are expressed throughout the development of vegetative tissues and reproductive organs. However, during fiber cell development their expression may be dramatically induced in response to rapid cell expansion and growth.

Consistent with the above notion, in a previous study using 12,227 fiber ESTs from *G. arboreum* L. Arpat et al. (2004) found that only 81 genes were up-regulated and 2,553 "expansion-associated" genes were down-regulated during the developmental switch from primary to secondary cell wall biosynthesis. This low percentage of up-regulated genes detected in the stages of fiber cell expansion and elongation is reminiscent of ~100 genes detected in the early stages of fiber development (Fig. 4, Table 3) regardless of the small set of genes used in this study. The data obtained from two experiments using different stages of fiber development are difficult to be compared to detect a common set of genes, if any, that is

Table 4 List of 23 genes that were up-regulated in the fiber-bearing ovules (3DPA) in TM-1

GenBank Accession	TIGR CGI [a]	Symbol	Description	TM-1 (3DPA)/ *N1N1* (3DPA) [b]	TM-1 (3DPA)/ TM-1 (Petal) [b]	TM-1 (3DPA)/ TM-1 (Leaf) [b]	Expression Patterns in qRT-PCR analysis [c]
BM356398	TC27365	E6	Fiber protein E6 (clone CKE6-4A)	1.74 (2.07 ± 0.23)	51.24 (28.20 ± 3.51)	47.05 (168.66 ± 11.42)	III
AY072823	TC37712	FDH	Fiddlehead-like protein	1.94 (1.84 ± 0.40)	2.38 (2.71 ± 0.51)	2.69 (9.26 ± 1.45)	IV
AY072821	TC37267	RDL1	Dehydration-induced protein RD22-like protein	1.82 (9.28 ± 1.98)	1.63 (3.49 ± 0.16)	5.35(229.32 ± 22.65)	III
AW587505	TC27812	WBC1	ABC transporter	2.16 (5.37 ± 0.20)	1.71 (1.10 ± 0.10)	2.96 (8.83 ± 1.42)	III
AW587516	TC27322	SAHH	S-adenosylhomocysteine hydrolase	2.12 (9.91 ± 0.63)	44.07 (48.09 ± 8.47)	12.23 (12.91 ± 1.20)	III
AY094465	TC32989	STP	Putative sugar transport protein (*Arabidopsis thaliana*)	1.82 (3.68 ± 0.24)	2.33 (2.66 ± 0.10)	3.59 (21.14 ± 1.40)	III
U12858	TC39106	NAP	Nucleosome assembly protein I-like protein	1.70 (2.05 ± 0.04)	3.91 (4.55 ± 0.85)	3.23 (12.15 ± 0.17)	IV
BM356400	TC27288	SPP	Serine protease-like protein (*Arabidopsis thaliana*)	1.72 (3.98 ± 0.43)	4.90 (107.88 ± 8.94)	3.81(10.51 ± 1.20)	III
AF336283	TC38383	MYB25	GhMYB25	1.86 (3.40 ± 0.37)	3.59 (465.23 ± 28.25)	3.02 (2307.50 ± 189.94)	II
AI726131		RDL2	Dehydration-responsive protein RD22	2.32 (7.52 ± 0.71)	2.76 (2.43 ± 0.24)	13.94 (244.08 ± 41.08)	V
AY086643	TC37978	SAC25	Putative pod-specific dehydrogenase SAC25	2.71 (12.13 ± 0.47)	3.16 (1637.62 ± 63.51)	2.63(5.07 ± 0.22)	II
AW561923		CNX	Calnexin (*Glycine max*)	1.69 (4.31 ± 0.63)	1.87(31.37 ± 1.55)	5.26 (144.83 ± 14.73)	III
AW561930	TC27269	EF1A	Elongation factor-1 alpha	1.65 (2.64 ± 0.48)	4.66(2.08 ± 0.43)	4.61(5.18 ± 1.54)	IV
AW587470	TC39685	HSP70	Heat shock protein 70 (*Petunia x hybrida*)	1.80(3.45 ± 0.58)	2.29(15.44 ± 1.61)	4.62 (7.63 ± 0.74)	IV
BG442892		CEP52	Ubiquitin/ribosomal protein CEP52 (*Trypanosoma brucei*)	1.58(5.81 ± 0.30)	1.88(0.37 ± 0.04)	2.18(1.90 ± 0.27)	IV
BG447373	TC32637	FAD3	Omega-3 fatty acid desaturase, endoplasmic reticulum	1.70(5.05 ± 0.44)	2.54 (93.50 ± 6.60)	1.88 (45.36 ± 4.47)	III
AY518291	TC28115	SRF6	Strubbelig receptor family 6 (*Arabidopsis thaliana*)	1.62(1.72 ± 0.17)	3.09 (0.52 ± 0.04)	3.21 (1.70 ± 0.06)	IV
AF141375	TC27374	PDF1	Protodermal factor 1 (*Arabidopsis thaliana*)	1.70(2.64 ± 0.21)	67.03 (636.96 ± 31.32)	11.64 (24.68 ± 2.11)	I
AY343972	TC37094	BDC1	BURP domain-containing protein	2.44(5.24 ± 0.86)	3.56(2.23 ± 0.22)	15.04(189.80 ± 19.56)	III
AC004473	TC27702	BDC2	BURP domain-containing protein	1.79(19.67 ± 2.53)	3.28(1161.27 ± 154.42)	3.89(107.21 ± 4.83)	III
L12258	TC38154	NOD26	Nodulin-26 (*Glycine max*)	1.97(17.67 ± 0.43)	2.31 (94.21 ± 4.59)	5.89 (66.18 ± 4.04)	III
AF059484	TC37642	ACT	Actin	1.70(5.26 ± 0.16)	8.43(3.16 ± 0.39)	7.05 (83.75 ± 4.46)	V
AC026875	TC27277	EF1-like	Translation elongation factor	1.58 (3.87 ± 0.53)	5.65 (7.88 ± 1.19)	4.87 (10.66 ± 1.01)	III

[a] Cotton gene index at TIGR (http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species = cotton)
[b] The upper and lower numbers in each column indicate the fold changes of gene expression in each comparison detected by microarray and qRT-PCR analyses, respectively
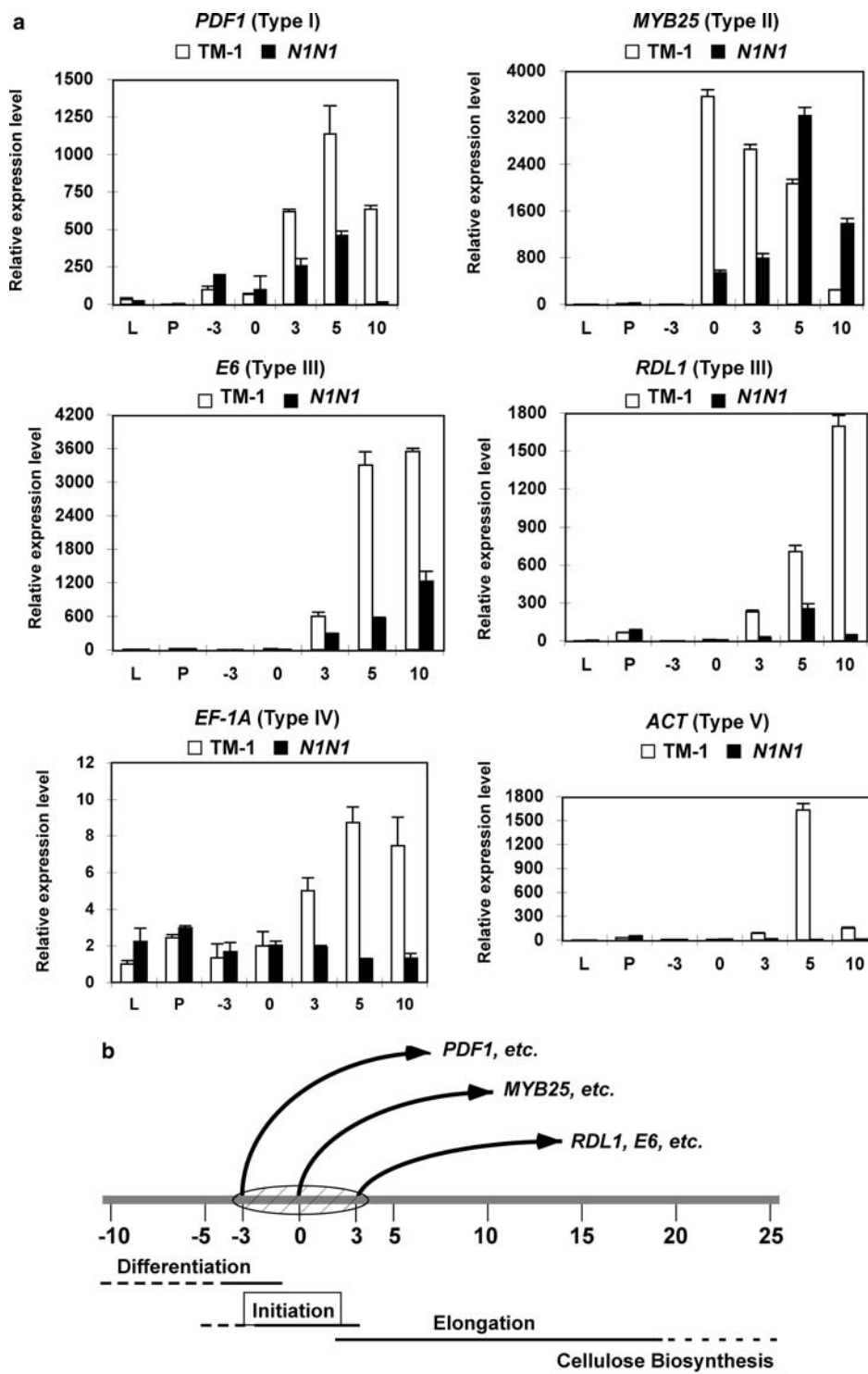[c] See the text for the description of five gene expression patterns detected in qRT-PCR analysis

Fig. 5 A sequential activation of the fiber-associated genes during early stages of fiber development. **a** Gene expression was analyzed in seven tissues including fibers and non-fibers using quantitative RT-PCR (qRT-PCR) analysis. *L*: leaves; *P*: petals; 0, 3, and 5: ovules at 0, 3, and 5 DPA, respectively; 10: fibers at 10 DPA. The expression patterns of six genes shown in the figure represented five gene-activation patterns (Type I–V) observed from early stages of fiber cell development. The expression patterns of other genes can be found in Table 4 and Supplementary Fig. 1; **b** A simplified model for temporal activation of gene expression during early stages of fiber cell development. The genes were up-regulated sequentially in the ovules at −3, 0, and 3 DPA during stages of fiber cell expansion and elongation. Down-regulation of the fiber-associated genes in the *N1N1* mutant led to a defective process of fiber cell development (see text for details). The stage from −3 to 3 DPA (hatched oval) marked fiber cell initiation or formation

regulated consistently during different stages of fiber development.

Developmental regulation of fiber-associated genes in TM-1 and N1N1 mutant

Fiber cell expansion and elongation occur continuously through a diffuse-growth mechanism (Tiwari and Wilkins 1995). It is notable that the timing of up-regulation of 23 fiber-associated genes coincides with temporal control of fiber cell development, ranging from −3 DPA (1 gene), 0 DPA (2 genes), 3 DPA (12 genes), to 5 or 10 DPA (2 genes) (Fig. 5a, Supplementary Fig. 1). Six of 23 genes were up-regulated in the ovules at 3 DPA and in the fibers at 10 DPA but also expressed in the non-fiber tissues such as leaves and petals. Thus, ovular development at 3 DPA is a critical step for rapid cell expansion and cellular growth. The data support a model of temporal activation of regulatory networks during early stages of fiber cell development (Fig. 5b). External and internal signals for fiber cell differentiation may be transmitted to fiber cell primordia, leading to the activation of "patterning" genes (Hulskamp 2004) including GhPDF1. Arabidopsis homolog of GhPDF1 encoding a putative extracellular proline-rich protein is exclusively expressed in the L1 layer of shoot apices and the protoderm of organ primordia (Abe et al. 1999, 2001). Molecular events of gene activation in fiber cell primordia may be coupled with sequential activation of transcription factors and proteins such as MYB transcription factors and RDL1 proteins that are important to fiber or trichome cell differentiation. There is evidence that cotton and Arabidopsis use similar transcription factors for regulating trichome development. Notably, GaMYB2 and RDL1 promoter interact physically, and overexpressing GaMYB2 complements the gl1 mutant phenotype as well as induces the development of seed trichomes in Arabidopsis (Wang et al. 2004). The data suggest a critical role of MYB transcription factors in fiber cell differentiation. Interestingly, RDL1 and MYB25, a homolog of AtMYB17 and 106 (Arabidopsis Genome Initiative 2000) that are in a separate branch but closely related to GL1, were up-regulated in the fiber-developing ovules at 0 and 3 DPA, respectively, suggesting that multiple components of MYB transcription factors and other proteins are involved in fiber cell differentiation. Up-regulation of many other genes such as WBC1, FDH, EF1A, and NOD26 (Table 4, Supplementary Table 4) is likely involved in late stages of fiber cell differentiation. For example, GhWBC1 encoding an ATP-binding transporter of the WBC subfamily is expressed at low levels in the ligon-lintless mutant that is defective in fiber cell elongation (Zhu et al. 2003).

N1N1 mutation not only delays the onset of fiber cell initiation by 12–24 h but also reduces the number of undifferentiated protodermal cells that develop fiber cell initials. We identified 117 genes and 30 genes that displayed expression changes in the N1N1 mutant at 3 DPA and 0 DPA, respectively (Table 2). When per-gene variance is used, 498 and 195 genes were expressed significantly different between the mutant and wild type in these two stages. Moreover, equal number of genes were up- or down-regulated in the ovules at 3 DPA, whereas 20 genes were up-regulated and 10 down-regulated in the ovules at 0 DPA in N1N1. Relatively equal number of genes that were up- or down-regulated in N1N1 (Table 3) suggest that the N1N1 mutation has both positive and negative effects on gene regulation in the early stages of fiber development, leading to fewer fibers and a defective process of fiber cell elongation.

Twenty-three genes with a few exceptions were expressed exclusively in fiber-related tissues including ovules (−3 and 0 DPA), fiber-bearing ovules (3 and 5 DPA), and fibers (10 DPA) and down-regulated in the N1N1 mutant, which are the likely candidate genes involved in the process of fiber cell formation. All 23 fiber-associated fiber genes were up-regulated in the ovules and fibers compared to the non-fiber tissues in TM-1. Compared to TM-1, the expression levels of these genes, except for RDL2 and GhMYB25, were dramatically reduced in the corresponding tissues tested in N1N1 (Fig. 5a, Supplementary Table 4), suggesting that temporal regulation of gene activation is disrupted in the N1N1 mutant.

We note that for a technical difficulty in dissecting fibers from N1N1 mutants, we used fiber-bearing ovules at 10 DPA in the mutant compared to fibers at 10 DPA in the wild type, which may obscure the detection of fiber-related genes. For example, RDL2 was expressed only in the fibers at 10 DPA in N1N1 but not in TM-1, and GhMYB25 was expressed at higher levels in the fiber-bearing ovules at 5 DPA and fibers at 10 DPA in N1N1 than in TM-1. The expression differences detected at 10 DPA may be caused by different tissues used in the study (fibers in TM-1 and fiber-bearing ovules in N1N1). Alternatively, GhMYB25 and RDL2 expression may be affected by negative regulators or affected indirectly by interacting with other protein factors induced by the N1N1 mutation, which is reminiscent of the equal number of the genes that were up- or down-regulated in the N1N1 mutant (Table 3).

FDH encodes an enzyme involved in the synthesis of long-chain lipids found in the cuticle (Lolle et al. 1992; Pruitt et al. 2000), and mutations in FDH suppress epidermal cell interactions in Arabidopsis and exhibit a deleterious effect on trichome development (Yephremov et al. 1999). The FDH-like gene is highly expressed in developing fibers (Li et al. 2002) and is repressed in the N1N1 mutant, suggesting down-regulation of the FDH-like gene in the N1N1 background might be associated with abnormal development of fiber cell initials. Down-regulation of the genes (Fig. 5a, Supplementary Fig. 1) involved in cell differentiation (e.g., PDF1), transcriptional regulation (e.g., MYB25), signal transduction (e.g., SPP, BDC1, and BDC2), transport facilitation (e.g., STP and WBC1), cell wall

biosynthesis (e.g., *FDH*, *RDL1*, and *NOD26*), and translational regulation (e.g., *EF1A* and *EF1-like*) are likely to be associated with a series of defective processes of fiber cell formation and elongation in the *N1N1* mutant. *ACT* transcripts were highly accumulated in TM-1 ovules at 5 DPA, suggesting a role of actin cytoskeleton during fiber development (Li et al. 2005). Notably, S-adenosylhomocysteine hydrolase is a key enzyme involved in the intracellular methylation reactions (Fojtova et al. 1998; Tanaka et al. 1997). Suppression of *SAHH* may lead to pleiotropic effects on polarized growth of developing fiber cells in the *N1N1* mutant. Our data suggest that *N1* mutation affects many downstream genes involved in various biological pathways including cell differentiation, transcriptional and translational regulation, and signal transduction that are essential for the determination of the number of fiber cells and fiber cell elongation during early stages of fiber development.

In summary, our data indicated: (1) the dominant mutation (*N1N1*) delayed fiber cell formation and reduced the number of fiber cell initials; (2) *N1N1* had both negative and positive effects on gene regulation associated with fiber development; (3) 20–50% fiber ESTs were expressed in both fibers and non-fiber tissues; (4) many genes played a role in the temporal regulation of gene expression during early stages of fiber development; and (5) the fiber-associated genes were concertedly down-regulated in the *N1N1* mutant.

References

Abe M, Takahashi T, Komeda Y (1999) Cloning and characterization of an L1 layer-specific gene in Arabidopsis thaliana. Plant Cell Physiol 40:571–580

Abe M, Takahashi T, Komeda Y (2001) Identification of a cis-regulatory element for L1 layer-specific gene expression, which is targeted by an L1-specific homeodomain protein. Plant J 26:487–494

Adams KL, Cronn R, Percifield R, Wendel JF (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci USA 100:4649–4654

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

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

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

Benedict CR, Smith RH, Kohel RJ (1973) Incorporation of ^{14}C-photosynthate into developing cotton balls, *Gossypium hirsutum* L. Crop Sci 13:88–91

Black MA, Doerge RW (2002) Calculation of the minimum number of replicate spots required for detection of significant gene expression fold change in microarray experiment. Bioinformatics 18:1609–1616

Blanc G, Wolfe KH (2004b) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16:1667–1678

Blanc G, Wolfe KH (2004a) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell 16:1679–1691

Cedroni ML, Cronn RC, Adams KL, Wilkins TA, Wendel JF (2003) Evolution and expression of MYB genes in diploid and polyploid cotton. Plant Mol Biol 51:313–325

Chen ZJ, Wang JL, Tian L, Lee HS, Wang JYJ, Chen M, Lee JSJ, Josefsson C, Madlung A, Watson B, Lippman Z, Vaughn M, Pires JC, Colot V, Doerge RW, Martienssen RA, Comai L, Osborn TC (2004) The development of an Arabidopsis model system for genome-wide analysis of polyploidy effects. Bio J Linn Soc 82:689–700

Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. J Amer Stat Assoc 74:829–836

Endrizzi JE, Turcotte EL, Kohel RJ (1984) Qualitative genetics, cytology and cytogenetics. In: Kohel RJ, Lewis DF (eds) Agronomy: Cotton. American Society of Agronomy Inc, Madison WI, pp 59–80

Fojtova M, Kovarik A, Votruba I, Holy A (1998) Evaluation of the impact of S-adenosylhomocysteine metabolic pools on cytosine methylation of the tobacco genome. Eur J Biochem 252:347–352

Graves DA, Stewart JM (1988) Chronology of the differentiation of cotton (*Gossypium hirsutum* L) fiber cells. Planta 175:254–258

Hochberg Y, Tamhane AC (1987) Multiple comparison procedures. Wiley, New York, 450 p

Hulskamp M (2004) Plant trichomes: a model for cell differentiation. Nat Rev Mol Cell Biol 5:471–480

Hülskamp M, Schnittger AT (1998) Spatial regulation of trichome formation in *Arabidopsis thaliana*. Semin Cell Dev Biol 9:213–220

Hülskamp M, Misera S, Jurgens G (1994) Genetic dissection of trichome cell development in *Arabidopsis*. Cell 76:555–566

Ji SJ, Lu YC, Feng JX, Wei G, Li J, Shi YH, Fu Q, Liu D, Luo JC, Zhu YX (2003) Isolation and analyses of gene preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. Nucleic Acids Res 31:2534–2543

John ME (1996) Structural characterization of genes corresponding to cotton fiber mRNA, E6: reduced E6 protein in transgenic plants by antisense gene. Plant Mol Biol 30:297–306

John ME, Crow LJ (1992) Gene expression in cotton (*Gossypium hirsutum* L) fiber: cloning of the mRNAs. Proc Natl Acad Sci U S A 89:5769–5773

Kane MD, Jatkoe TA, Stumpf CR, Lu J, Thomas JD, Madore SJ (2000) Assessment of the sensitivity and specificity of oligonucleotide (50 mer) microarrays. Nucleic Acids Res 28:4552–4557

Kerr MK, Churchill GA (2001) Experimental design for gene expression microarrays. Biostatistics 2:183–201

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

Kohel RJ (1973) Genetic nomenclature in cotton. J Hered 64:291–295

Lee HS, Wang JL, Tian L, Jiang HM, Black MA, Madlung A, Watson B, Lukens L, Pires JC, Wang JJ, Comai L, Osborn TC, Doerge RW, Chen ZJ (2004) Sensitivity of 70-mer oligonucleotides and cDNAs for microarray analysis of gene expression in Arabidopsis and its related species. Plant Biotechnol J 2:45–57

Li F, Stormo GD (2001) Selection of optimal DNA oligos for gene expression arrays. Bioinformatics 17:1067–1076

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 XB, Fan XP, Wang XL, Cai L, Yang WC (2005) The cotton ACTIN1 gene is functionally expressed in fibers and participates in fiber elongation. Plant Cell 17:859–875

Loguercio LL, Zhang JQ, Wilkins TA (1999) Differential regulation of six novel MYB-domian genes defines two distinct expression patterns in allotetraploid cotton (*Gossypium hirsutum* L). Mol Gen Genet 261:660–671

Lolle SJ, Cheung AY, Sussex IM (1992) Fiddlehead: an *Arabidopsis* mutant constitutively expressing an organ fusion program that involves interactions between epidermal cells. Dev Biol 152:383–392

Marks MD (1997) Molecular genetic analysis of trichome development in *Arabidopsis*. Ann Rev Plant Physiol Plant Mol Biol 48:137–163

Mauney JR (1984) Anatomy and morphology of cultivated cottons. In: Kohel RJ, Lewis DF (eds) Agronomy: cotton. pp 59–80

Meinert MC, Delmer DP (1977) Changes in biochemical composition of cell-wall of cotton fiber during development. Plant Physiol 59:1088–1097

Murai K, Takumi S, Koga H, Ogihara Y (2002) Pistillody, homoeotic transformation of stamens into pistil-like structures, caused by nuclear-cytoplasm interaction in wheat. Plant J 29:169–181

Noda K, Glover BJ, Linstead P, Martin C (1994) Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor. Nature 369:661–664

Payne T, Clement J, Arnold D, Lloyd A (1999) Heterologous myb genes distinct from GL1 enhance trichome production when overexpressed in *Nicotiana tabacum*. Development 126:671–682

Pruitt RE, Vielle-Calzada JP, Ploense SE, Grossniklaus U, Lolle SJ (2000) FIDDLEHEAD, a gene required to suppress epidermal cell interactions in *Arabidopsis*, encodes a putative lipid biosynthetic enzyme. Proc Natl Acad Sci U S A 97:1311–1316

Quackenbush J (2002) Microarray data normalization and transformation. Nat Genet 32:496–501

Ramsey JC, Berlin JD (1976) Ultrastructure of early stages of cotton fiber differentiation. Bot Gaz 137:11–19

Rouillard JM, Herbert CJ, Zuker M (2002) OligoArray: genome-wide oligonucleotide design for microarrays. Bioinformatics 18:486–487

Ruan YL, Chourey PS (1998) A fiberless seed mutation in cotton is associated with lack of fiber cell initiation in ovule epidermis and alterations in sucrose synthase expression and carbon partitioning in developing seeds. Plant Physiol 118:399–406

Ruan YL, Llewellyn DJ, Furbank RT (2000) Pathway and control of sucrose import into initiating cotton fibers. Aust J Plant Physiol 27:795–800

Ruan Y, Llewellyn D, Furbank R (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, Llewellyn DJ, Furbank RT (2003) Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. Plant Cell 15:952–64

Samora PJ, Stelly DM, Kohel RJ (1994) Localization and mapping of the Le1 and Gl2 of cotton (*Gossypium hirsutum* L). J Hered 85: 152–157

Schubert AM, Benedict CR, Gates CE, Kohel RJ (1976) Growth and development of the lint fibers of Pima S-4 cotton. Crop Sci 16:539–543

Stewart JM (1975) Fiber initiation on the cotton ovule (*G hirsutum*). Amer J Bot 62:723–730

Suo J, Liang X, Pu L, Zhang Y, Xue Y (2003) Identification of GhMYB109 encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (*Gossypium hirsutum* L). Biochim Biophys Acta 1630:25–34

Szymanski DB, Lloyd AM, Marks MD (2000) Progress in the molecular genetic analysis of trichome initiation and morphogenesis in *Arabidopsis*. Trends Plant Sci 5:214–219

Tanaka H, Masuta C, Uehara K, Kataoka J, Koiwai A, Noma M (1997) Morphological changes and hypomethylation of DNA in transgenic tobacco expressing antisense RNA of the S-adenosyl-L-homocysteine hydrolase gene. Plant Mol Biol 35:981–986

Tian L, Wang J, Fong MP, Chen M, Cao H, Gelvin SB, Chen ZJ (2003) Genetic control of developmental changes induced by disruption of *Arabidopsis* histone deacetylase 1 (*AtHD1*) expression. Genetics 165:399–409

Tian L, Fong MP, Wang JJ, Wei NE, Jiang H, Doerge RW, Chen ZJ (2005) Reversible histone acetylation and deacetylation mediate genome-wide, promoter-dependent and locus-specific changes in gene expression during plant development. Genetics 169:337–345

Tiwari SC, Wilkins TA (1995) Cotton (*Gossypium hirsutum*) seed trichomes expand via diffuse growing mechanism. Can J Bot 73:746–757

Wang S, Wang JW, Yu N, Li CH, Luo B, Gou JY, Wang LJ, Chen XY (2004) Control of plant trichome development by a cotton fiber MYB gene. Plant Cell 16:2323–2334

Wang J, Lee JJ, Tian L, Lee HS, Chen M, Rao S, Wei EN, Doerge RW, Comai L, Chen ZJ (2005) Methods for genome-wide analysis of gene expression changes in polyploids. Methods Enzymol 395:570–596

Wilkins TA, Jernstedt JA (1999) Chapter 9. Molecular genetics of developing cotton fibers. In: Basra AM (ed) Cotton fibers. Hawthorne Press, New York, pp 231–267

Yephremov A, Wisman E, Huijser P, Huijser C, Wellesen K, Saedler H (1999) Characterization of the FIDDLEHEAD gene of *Arabidopsis* reveals a link between adhesion response and cell differentiation in the epidermis. Plant Cell 11:2187–2201

Zhu YQ, Xu KX, Luo B, Wang JW, Chen XY (2003) An ATP-binding cassette transporter GhWBC1 from elongating cotton fibers. Plant Physiol 133: 580–588